

Paketsubstitution in Audiosignalen bei paketorientierter Audioübertragung

Von der Fakultät IV-Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
- Dr.-Ing.-
genehmigte Dissertation

von
Dipl.-Ing. Lilia Lajmi

Promotionsausschuss:

Vorsitzender: Prof. Dr.-Ing. Post
Berichter: Prof. Dr.-Ing. Peter Noll
Berichter: Prof. Dr.-Ing Peter Vary

Tag der wissenschaftlichen Aussprache: 29.04.2003

Berlin 2003
D 83

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Nachrichtentechnik und Theoretische Elektrotechnik, Fachgebiet Fernmelde-technik der Technischen Universität Berlin.

Dem Direktor des Instituts, Herrn Professor Dr.-Ing. P. Noll, danke ich für die Anregung zum Thema der Arbeit. Ihm gilt mein besonderer Dank für die Betreuung der Arbeit und für die Übernahme des Hauptreferats.

Herrn Professor Dr.-Ing. P. Vary möchte ich für das entgegengebrachte Interesse an der Arbeit und die Übernahme des Korreferates danken.

Ebenso gilt mein Dank Herrn Professor Dr.-Ing. Post für die Übernahme des Vorsitzes der Prüfungskommission.

Herrn Professor Dr.-Ing. T. Sikora und all meinen Institutskollegen, die durch Diskussionen und Anregungen zum Gelingen der Arbeit beigetragen haben, danke ich herzlich.

Mein ganz besonderer Dank gilt allen Kollegen der Firma Panasonic AVC Networks für das gute Arbeitsklima und der Nachsicht während der Prüfungsvorbereitung. Ich danke Herrn Dr.-Ing. Ingo Chmielewski für sein Verständnis und seine wertvollen Ratschläge, sowie meinem Kollegen Henning Heiber.

Mein ganz besonderer Dank gilt natürlich meinem Mann Adel und meinem Sohn Laith für ihre Nachsicht, Unterstützung, Erheiterung und Geduld während der gesamten Arbeit.

Braunschweig, im April 2003.

Lilia Lajmi

Kurzfassung

In paketvermittelnden Netzen - wie insbesondere dem Internet - werden zunehmend auch Audiodaten übertragen. Typische Anwendungen sind z.B. *Real-Time-Streaming* oder *Voice-over-IP*. Diese basieren i.A. auf dem unzuverlässigem Datagram-Service UDP, um Verzögerungen durch erneute Übertragung von verlorenen Paketen (wie bei TCP) zu vermeiden. Die UDP- und RTP-Internet-Protokolle bieten keine Möglichkeit, eine fehlerfreie Übertragung von Paketen für Echtzeitsdienste wie Audio oder Sprache zu garantieren. Es entstehen Paketverluste, welche die Dienstqualität beeinträchtigen und einen Qualitätsverlust der Audiodaten verursachen. Es existieren verschiedene Verfahren, um diesen Qualitätsverlust möglichst klein zu halten.

Ziel dieser Arbeit war es, die bei paketerorientierter Übertragung von Audiosignalen durch Paketverluste hervorgerufenen Störeffekte mit geeigneten Paketsubstitutionsverfahren zu reduzieren oder gar zu beheben. Ziel war es also einen besseren subjektiven Höreindruck des empfängerseitig ausgegebenen Audiosignals zu erreichen.

Im Verlauf dieser Arbeit wurde eine Reihe von Verfahren zur Paketsubstitution in Audiosignalen untersucht. Für die Verdeckung einer Lücke, die auf einen Paketverlust während der Übertragung zurückzuführen ist, ist eine weitgehende Parametrisierung, Analyse und darauf beruhende Inter- bzw. Extrapolation des Audiosignals erforderlich. Die Signalbehandlung erfolgte sowohl im Zeit- als auch im Frequenzbereich, ausgehend von den aus der Audiocodierung und -verarbeitung bekannten Algorithmen. Beispiele hierfür sind die lineare Prädiktion, die Stereoprädiktion und die sinusoidale Modellierung von Audiosignalen.

Die Signalbehandlung im Zeitbereich basiert auf abstrakten, mathematisch motivierten Zielvorstellungen, welche das Audiosignal im Zeitbereich betrachten und eine bestmögliche Annäherung an den Signalverlauf des Idealsignals anstreben. Für die Rekonstruktion von Paketverlusten in Audiosignalen ist die damit erreichbare subjektive Audioqualität notwendigerweise begrenzt, aufgrund der Diskrepanz zwischen dem zugrundeliegenden mathematischen Konzept und dem völlig andersartigen Funktionsschema des menschlichen Hörsinns.

Eine entscheidende Verbesserung in der wahrnehmungsbezogenen Anpassung gelang durch den Übergang in der verwendeten Signalrepräsentation von der Zeitbereichs- in eine mathematisch äquivalente Spektralbereichsdarstellung. Durch den Übergang auf dieses Konzept konnte gegenüber der Signalbehandlung im Zeitbereich eine wesentliche Steigerung der subjektiven Audioqualität erreicht werden.

INHALTSVERZEICHNIS

<u>1</u>	<u>EINLEITUNG.....</u>	<u>8</u>
<u>2</u>	<u>ANALYSE VON AUDIOSIGNALEN</u>	<u>13</u>
2.1	AUDIOSIGNALE IM STEREO-FORMAT	13
2.2	EINTEILUNG VON AUDIOSIGNALEN.....	17
2.3	ANALYSE IM ZEITBEREICH.....	19
2.3.1	NULLDURCHGANGSRATE.....	19
2.3.2	AUTOKORRELATION.....	20
2.3.3	GRUNDFREQUENZ.....	22
2.3.4	HÜLLKURVEN-DETEKTION.....	23
2.4	ANALYSE IM FREQUENZBEREICH	25
2.4.1	FOURIERANALYSE.....	25
2.4.2	DISKRETE COSINUSTRANSFORMATION.....	28
2.5	DIE SINUSOIDALE MODELLIERUNG VON AUDIOSIGNALEN	28
2.5.1	GRUNDLAGEN DES SINUSOIDALEN MODELLS	29
2.5.2	SINUSOIDALE ANALYSE.....	30
2.5.3	SINUSOIDALE SYNTHESE.....	31
<u>3</u>	<u>AUDIOÜBERTRAGUNG ÜBER PAKETVERMITTELNDE NETZE.....</u>	<u>34</u>
3.1	ASYNCHRONOUS TRANSFER MODE (ATM).....	35
3.1.1	AUFBAU EINER ATM-ZELLE	36
3.1.2	ATM-REFERENZMODELL.....	37
3.2	INTERNET-PROTOKOLL.....	38
3.3	REAL-TIME AUDIO-PROTOKOLLE.....	40
3.3.1	REAL-TIME TRANSPORT PROTOCOL	40
3.3.2	REAL-TIME STREAMING PROTOCOL	41
3.3.3	RESOURCE RESERVATION PROTOCOL.....	41
3.4	VORTEILE DER PAKETVERMITTLUNG	42
3.5	PROBLEME DER PAKETVERMITTLUNG.....	42
3.5.1	VARIABLE LAUFZEITEN.....	43
3.5.2	NETZÜBERLASTUNG.....	43
3.5.3	KANALFEHLER.....	43
3.6	VERLUST-CHARAKTERISTIK	44
<u>4</u>	<u>SUBSTITUTIONSSIGNALE FÜR PAKETVERLUSTE.....</u>	<u>45</u>
4.1	SENDERSEITIGE VERFAHREN.....	45

4.1.1	RETRANSMISSION.....	46
4.1.2	INTERLEAVING.....	47
4.1.3	FORWARD ERROR CORRECTION.....	48
4.2	EMPFÄNGERSEITIGE VERFAHREN.....	49
4.3	SUBSTITUTION AUF SIGNALEBENE.....	53
4.3.1	NULLSUBSTITUTION.....	54
4.3.2	PAKETWIEDERHOLUNG.....	55
4.3.3	SUBSTITUTION MIT RAUSCHEN.....	57
4.3.4	ODD-EVEN-SAMPLE INTERPOLATION.....	57
4.3.5	LINKS/RECHTS-ERSATZ.....	58
4.3.6	VERFAHREN DER MAXIMALEN KREUZKORRELATION.....	60
4.3.7	PACKET MERGING.....	68
4.4	SUBSTITUTION MITTELS LINEARER PRÄDIKTION.....	68
4.4.1	SYNTHESE EINES SUBSTITUTIONSSIGNALS AUS DER INNOVATION.....	70
4.4.1.1	Einfache Verfahren.....	71
4.4.1.2	Pitchsynchrone Paketwiederholung.....	71
4.4.1.3	Links/Rechts-Ersatz der Innovation.....	75
4.4.1.4	Gewichtete Interpolation.....	76
4.5	STEREOPRÄDIKTION ZUR SIGNALSUBSTITUTION BEI PAKETVERLUSTEN	88
4.5.1	SYNTHESE EINES SUBSTITUTIONSSIGNALS AUS INNOVATION.....	88
4.5.2	BESTIMMUNG DER OPTIMALEN FILTERKOEFFIZIENTEN EINES STEREOPRÄDIKTORS.....	90
4.5.3	INSTABILITÄTSPROBLEM DES SLPC-SYNTHESEFILTERS.....	94
4.5.3.1	Stabilitätskontrolle mit PARCOR-Koeffizienten.....	95
4.5.3.2	Stabilitätskontrolle mit Polstellen-Untersuchung.....	97
4.5.3.3	Behandlung der Instabilität des Synthesefilters.....	101
4.6	SUBSTITUTION IM FREQUENZBEREICH.....	111
4.6.1	MOTIVATION FÜR DIE REKONSTRUKTION VON PAKETVERLUSTEN IM FREQUENZBEREICH... ..	111
4.6.2	SPEKTRALES MORPHING.....	111
4.6.3	FREQUENZ-TRACKING.....	119
5	<u>VERGLEICH DER UNTERSUCHTEN SUBSTITUTIONSVERFAHREN.....</u>	133
5.1	OBJEKTIVE BEWERTUNG.....	133
5.1.1	SIGNAL-TO-NOISE RATIO.....	134
5.1.2	SEGMENT-SNR.....	135
5.1.3	FREQUENZGEWICHTETER SEGMENT-SNR.....	136
5.1.4	OBJECTIVE DIFFERENCE GRADE.....	136
5.2	SUBJEKTIVE BEWERTUNG MIT MOS-MESSUNG.....	137
5.3	UNTERSUCHUNGSBEDINGUNGEN.....	138
5.4	VERWENDETES AUDIOMATERIAL.....	140
5.5	SIMULATION AUF SIGNALEBENE	145
5.6	SIMULATION AUF DER EBENE DER LPC-INNOVATION UND MIT STEREOPRÄDIKTION	153
5.7	SIMULATION AUF FREQUENZBEREICHSEBENE	158
6	<u>ZUSAMMENFASSUNG UND AUSBLICK</u>	164

7 LITERATURVERZEICHNIS	168
-------------------------------------	------------

ANHANG A.....	178
----------------------	------------

A.1 ABKÜRZUNGSVERZEICHNIS.....	178
---------------------------------------	------------

A.2 VERZEICHNIS DER VERWENDETEN FORMELZEICHEN.....	181
---	------------

ANHANG B.....	183
----------------------	------------

B.1 POLSTELLENVERSCHIEBUNG.....	183
--	------------

B.2 QUADRATISCHE INTERPOLATION ZUR GENAUEN FREQUENZBESTIMMUNG.....	185
---	------------

B.3 METHODE DER KLEINSTEN FEHLER-QUADRATE FÜR DIE ERMITTLUNG DER AMPLITUDE UND PHASE EINER SINUSKOMPONENTE.....	187
--	------------

ANHANG C: SIMULATIONSWERTE.....	189
--	------------

C.1.1 SUBSTITUTION DES PCM-SIGNALS.....	189
--	------------

C.1.2 SUBSTITUTION DER LPC-INNOVATION.....	192
---	------------

C.1.3 SUBSTITUTION MIT DER STEREOPRÄDIKTION.....	193
---	------------

C.1.4 SUBSTITUTION IM FREQUENZBEREICH.....	194
---	------------

1 Einleitung

Die wachsende Verwendung von multimedialen Daten, wie z.B. Audio- oder Videodaten auf paketvermittelnden Netzen wie dem Internet zur Echtzeitkommunikation und -übertragung bringt einen Wandel im typischen Gebrauch von Computern mit sich.

Bei der digitalen Audio- und Videoübertragung via Internet wurden anfangs die Audiosignale und Videosequenzen grundsätzlich nur als Dateien bereitgestellt und mussten oft in einem zeitaufwendigen *Download* heruntergeladen werden, ehe sie dann von der Festplatte abgespielt werden konnten. Eine echte Neuerung brachte die Einführung von *Streaming-Verfahren*. Das Abspielen von Audio- und Videodaten erfolgt dabei direkt aus dem Internet ohne vorheriges Herunterladen. Diese Prozedur stellt relativ strenge Echtzeit-Anforderungen an das Übertragungsmedium, um diese Daten in guter Qualität zu übertragen und zu empfangen.

Bei der Übertragung von digitalen Audiodaten über paketvermittelnde Netze werden Audiosignale in kleine Datenblöcke unterteilt. Diese Blöcke, auch Pakete genannt, enthalten codierte Audiosignal-Segmente. Sie werden unabhängig voneinander übertragen. Beim Empfänger kann es dazu kommen, dass die gesendeten Audiopakete in falscher Reihenfolge, verspätet oder überhaupt nicht am Empfänger eintreffen. Diese Pakete werden als verloren angesehen und können daher nicht ausgegeben werden. Es entstehen Lücken im decodierten Audiosignal (s. Bild 1.1). Wird ein Audio- bzw. Sprachsignal übertragen, fällt dieses für eine gewisse Zeit aus, was die Klangqualität in hohem Maß beeinträchtigen und so für den Hörer sehr unangenehm sein kann. Eine Ursache dieser Paketverluste kann z.B. eine Überlastung im Netz sein, die zu Pufferüberläufen in den Vermittlungsknoten führt.

Ziel dieser Arbeit ist es daher, die bei paketorientierter Übertragung von Audiosignalen durch Paketverluste hervorgerufenen Störeffekte mit geeigneten Paketsubstitutionsverfahren zu reduzieren oder gar zu beheben. Ziel ist es einen besseren subjektiven Höreindruck des empfängerseitig ausgegebenen Audiosignals zu erreichen.

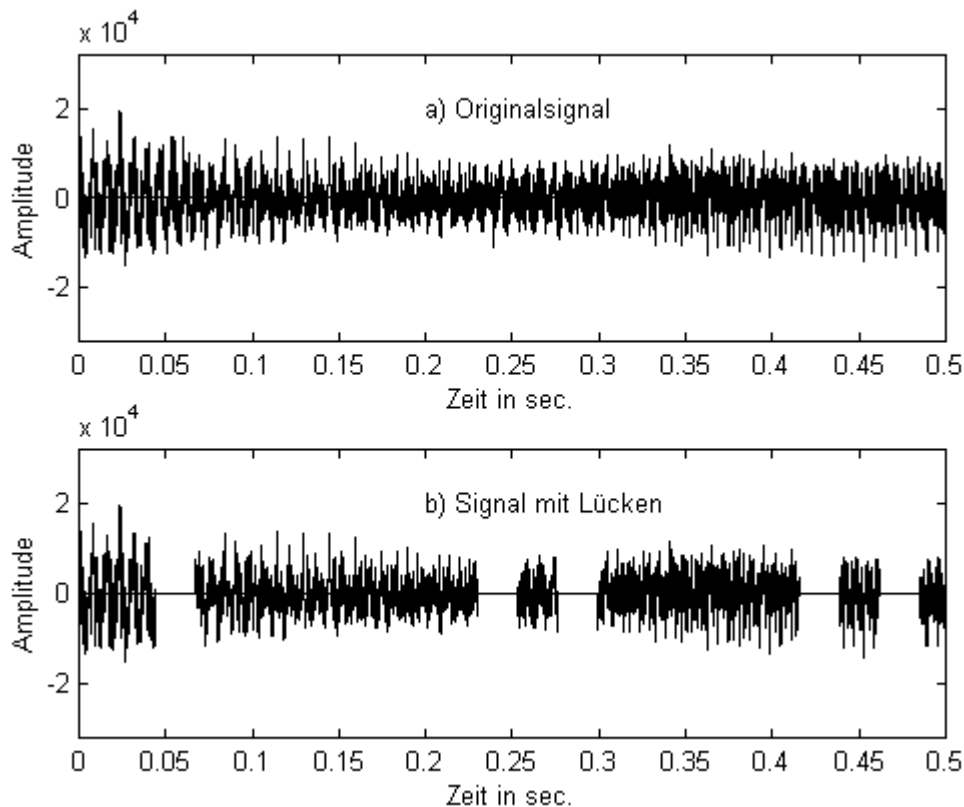


Bild 1.1: a) Original-Audiosignal der Dauer 0,5 Sekunden; b) Audiosignal mit Lücken entstanden durch Paketverluste. Die Breite der Lücke ist 1024 Abtastwerte (22,22 ms).

Stand der Forschung

In der Literatur wurde eine Reihe von Substitutionsverfahren vorgestellt, die auf Sprach- bzw. Audiosignale angewendet werden. Die Rekonstruktion der durch Paketverluste entstandenen Lücken im auszugebenden Signal erfolgt empfängerseitig. Man unterscheidet decodiererexterne und decodiererinterne Verfahren:

- **Decodiererexterne** Verfahren werden als zusätzliches Modul in die Nachrichtenübertragungskette eingefügt. Diese Verfahren können unmittelbar vor (Modul (1) in Bild 1.2) oder nach (Modul (2)) dem Quellendecodierer eingesetzt werden. Im ersten Fall wird bei einem Paketverlust ein kompletter Satz von Parametern und Koeffizienten, die ein komprimiertes Audiosignal beschreiben, rekonstruiert, indem beispielsweise die Parameter aus dem letzten korrekt empfangenen Paket wiederholt werden. Im zweiten Fall erfolgt die Paketsubstitution nach dem Quellendecodierer. Es werden in diesem Fall PCM-Audiosignale (PCM *Pulse Code Modulation*) rekonstruiert.

- **Decodiererinterne** Verfahren nehmen Änderungen des Decodierers selbst vor (in Modul (3) integriert). Es werden in diesem Fall komprimierte Audiosignale rekonstruiert. In den meisten Audio- und Sprachdecodierern sind *Concealment*-Module integriert, die bei einem Paketverlust versuchen, die fehlenden codierten Parameter durch z.B. Interpolation oder Wiederholung der vor dem Verlust vorhandenen Parameter, zu schätzen.

Die Untersuchungen zur Rekonstruktion von Lücken in Audiosignalen beschränken sich in dieser Arbeit auf die decodiererexterne Verfahren, die unmittelbar nach dem Decodierer eingesetzt werden. Die untersuchten bzw. vorgeschlagenen Substitutionsverfahren sind also unabhängig vom Codierverfahren (Bild 1.2).

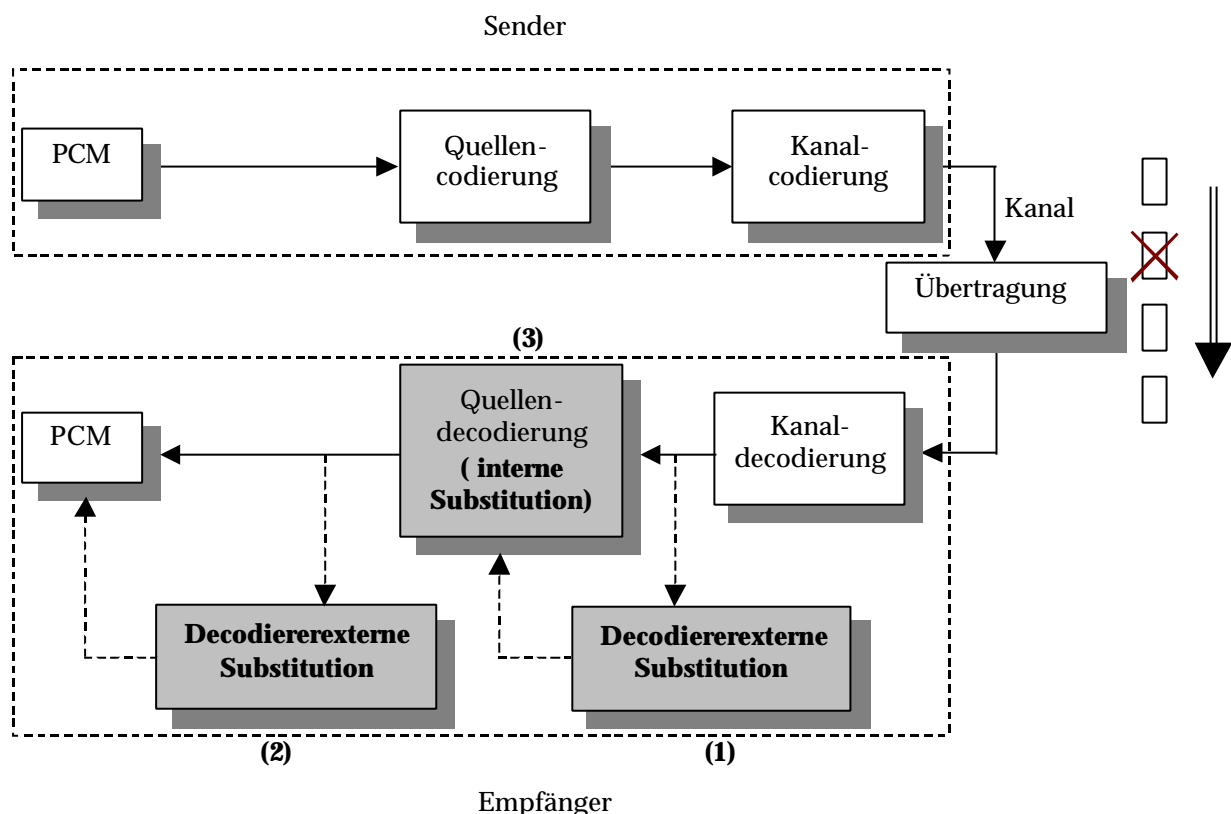


Bild 1.2: Empfängerseitige Paketsubstitution in decodiererexternen (Module 1 und 3) oder decodiererinternen (Modul 2) Substitutionsmodulen.

Diese Arbeit beschäftigt sich mit der Rekonstruktion der Lücken in qualitativ hochwertigen Audiosignalen, insbesondere Musik. Im Bereich der Sprachkommunikation ist Fehlerverschleierung kein Problem mehr. Aufgrund eines vorhandenen Spracherzeugungsmodells erwiesen sich einige Verfahren [Clü98a] [Ros01] [Goo86] als sehr wirkungsvoll und werden in der Praxis eingesetzt. Paketverluste in Audiosignalen lassen sich dagegen aufgrund der komplizierten Natur von Audiosignalen nur schwer rekonstruieren.

Gliederung der Arbeit

Der Kern dieser Arbeit liegt in der Rekonstruktion von Lücken in Audiosignalen, was notwendigerweise die Analyse des Audiosignals in der Umgebung der Lücke erfordert. Im Fall einer mehrkanaligen Übertragung kann auch das Signal eines benachbarten Kanals zur Schätzung der Lücke im Audiosignal herangezogen werden.

Das dieser Einleitung nachfolgende **Kapitel 2** ist den theoretischen Grundlagen der Analysemöglichkeiten für Audiosignale, sowohl im Zeit- als auch im Frequenzbereich, gewidmet. Es werden dort die Grundlagen behandelt, die für das Verständnis dieser Arbeit benötigt werden.

In **Kapitel 3** wird das Prinzip der Übertragung von Audiodaten über paketvermittelnde Netze betrachtet. Es werden sowohl die Vorteile als auch die Probleme, die durch eine paketorientierte Übertragung von digitalisierten Daten hervorgerufen werden, vorgestellt. Im Hinblick auf die Echtzeitübertragung von Audiodaten werden die dabei verwendeten Internet-Protokolle und die Probleme des Übertragungsmediums bzw. der verwendeten Protokolle, welche zu Paketverlusten während der Übertragung von Daten führen, vorgestellt. Diese Probleme werden insbesondere für Echtzeitanwendungen dargestellt.

Sollte es sich bei den zu übertragenden Daten um datenkomprimierte Audiodaten handeln, so entstehen aufgrund der aufgetretenen Paketverluste Lücken im auszugehenden Audiosignal, welche die Audioqualität in hohem Maß beeinträchtigen können. Für die Verbesserung der Audioqualität ist hier eine Rekonstruktion bzw. Verschleierung solcher Lücken unentbehrlich.

In **Kapitel 4** werden mögliche Lösungen für das Problem von Paketverlusten in Audiosignalen erörtert. Es werden unterschiedliche Paketsubstitutionsverfahren vorgestellt und untersucht.

Die Untersuchung erfolgt in vier Bereichen:

- Substitution der Lücke auf Signalebene,
- Substitution auf der Ebene des Differenzsignals nach einer Dekorrelation des Signals mittels linearer Prädiktion,
- Substitution für Stereosignale auf der Ebene des Differenzsignals nach einer Dekorrelation des Stereosignals mit der Stereoprädiktion,
- Substitution der Lücke im Frequenzbereich nach einer Transformation des Audiosignals in den Frequenzbereich mit der diskreten Cosinus-Transformation (DCT) oder mit der diskreten Fourier Transformation (DFT).

Kapitel 5 gibt einen Überblick über einige Bewertungsmerkmale, die zur Bewertung der nach der Substitution von Paketverlusten empfundenen Qualität des Audiosignals verwendet werden können. Es werden objektive und subjektive Bewertungskriterien vorgestellt. Objektive Bewertungsmaße werden mathematisch erfasst, sie haben in der Regel eine gerin-

ge Aussagekraft hinsichtlich der subjektiven Audioqualität. Es werden folgende objektive Bewertungsmaße verwendet: SNR , SNR_{seg} , $\text{SNR}_{f_0\text{-seg}}$ und ODG (*Objective Difference Grade*), letzterer wird mit den im ITU-R Standard empfohlenen Messverfahren PEAQ ermittelt. In Kapitel 6 folgt eine Bewertung der in Kapitel 4 vorgestellten Substitutionsverfahren durch objektive und subjektive Vergleiche der Audioqualität.

Kapitel 6 liefert eine Zusammenfassung der erzielten Ergebnisse sowie einen Ausblick auf mögliche Weiterentwicklungen.

2 Analyse von Audiosignalen

Audiosignale können verschiedene Signaleigenschaften und Charakteristiken aufweisen. Zu der Kategorie von Audiosignalen gehören beispielsweise Sprachsignale, Musik oder auch Geräusche, diese können sehr unterschiedliche Merkmale besitzen.

Audiosignale werden im allgemeinen in digitaler Form als Folge von quantisierten Abtastwerten gespeichert. Die häufigste Darstellungsform ist PCM (**P**ulse **C**ode **M**odulation) mit einer Auflösung von 16 Bit pro Abtastwert.

Typische Werte für Abtastfrequenzen von Audiosignalen liegen zwischen 8 und 96 kHz. Beispielsweise wird die Abtastfrequenz 44,1 kHz bei Audio-CDs und die Abtastfrequenz 8 kHz beim Telefonstandard verwendet.

2.1 Audiosignale im Stereo-Format

Zur Auswahl einer geeigneten Abtastfrequenz eines Audiosignals kommt die Wahl der Anzahl der Kanäle hinzu. Die Wandlung von einem analogen in ein digitales Signal kann sowohl in Mono (ein Kanal), als auch in Stereo (zwei Kanäle) erfolgen. Bei einer Abtastung von Stereosignalen werden zwei unabhängige Kanäle abgetastet. Man erhält damit die doppelte Datenmenge. Neuere Formate erlauben auch eine Wandlung mit mehr als 2 Kanälen, um einen *Surround-Sound* zu erreichen.

Der Begriff „Stereosignal“ wird in dieser Arbeit für zweikanalige Audiosignale verwendet. Die PCM-Abtastwerte einer Stereodatei sind abwechselnd angeordnet: Jeweils die Abtastwerte des linken und des rechten Kanals stehen der Reihe nach in der Stereodatei. Die Information über der Kanalzahl, Auflösung und Abtastfrequenz des Audiosignals werden bei den meisten Audioformaten in einem *Header* abgelegt (s. Bild 2.1). Als Beispiel ist das *Waveform-Format* (Wav) zu nennen, das eines der weitverbreitetsten Audiodatei-Formate darstellt. Wav-Dateien haben einen *Header* der Länge 44 Byte. Die Abtastwerte der Audiodatei liegen dabei im PCM-Format vor.

Header	L1	R1	L2	R2	L3	R3
--------	----	----	----	----	----	----	-----	-----

Bild 2.1: Abtastwerte-Anordnung einer Stereodatei im Waveformat.

Hochwertige digitale Audiosignale werden meist stereophon übertragen. Eine Ähnlichkeit von linkem und rechtem Kanal bei Stereosignalen ist oft, aber nicht immer gegeben.

Bild 2.2 zeigt ein Stereosignal, mit einer starken Ähnlichkeit der Signalverläufe beider Kanäle (a) Kanal links, b) Kanal rechts, c) Differenzsignal zwischen beiden Kanälen). Es handelt sich hier um einen Pop-Musik Titel. Beide Signalverläufe sind sehr ähnlich. Meist besteht bei solchen Signalen nur eine minimale Phasendifferenz.

Eine Korrelationsuntersuchung des linken Kanals mit sich selbst (Autokorrelationsfunktion aus Bild 2.3) und zwischen beiden Kanälen (Kreuzkorrelationsfunktion aus Bild 2.3), zeigt die hohe Ähnlichkeit der Kanäle, die durch die eindeutige Spitze der Kreuzkorrelationsfunktion bei der Verschiebung $\tau = 0$ beschrieben wird.

Bei Pop- und Rock-Musik sind die Kanäle meist sehr ähnlich. Transienten und Attacken sind oft auf beiden Kanälen vorhanden. Stark synthetische Pop-Musik enthält Effekte, die schnelle Frequenz- und Amplitudenänderungen erzeugen.

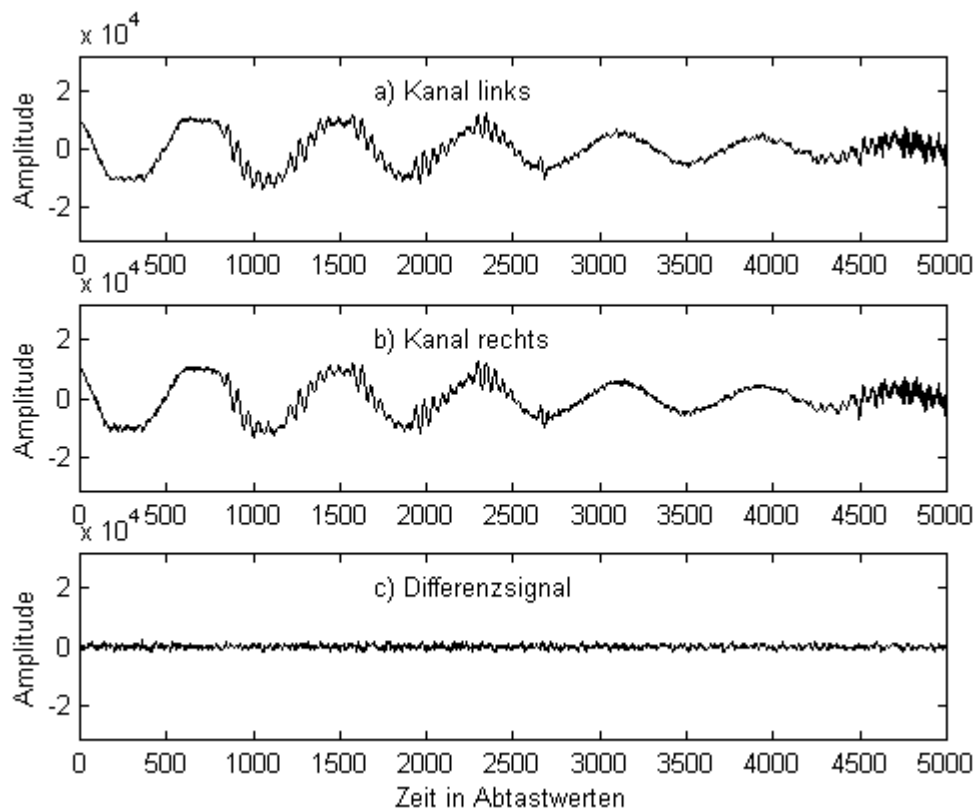


Bild 2.2: Ein Ausschnitt über 5000 Abtastwerte (113 ms) aus dem Stereosignal eines Pop-Musik-Titels mit starker Ähnlichkeit zwischen beiden Kanälen. a) Kanal links; b) Kanal rechts; c) Differenzsignal zwischen beiden Kanälen.

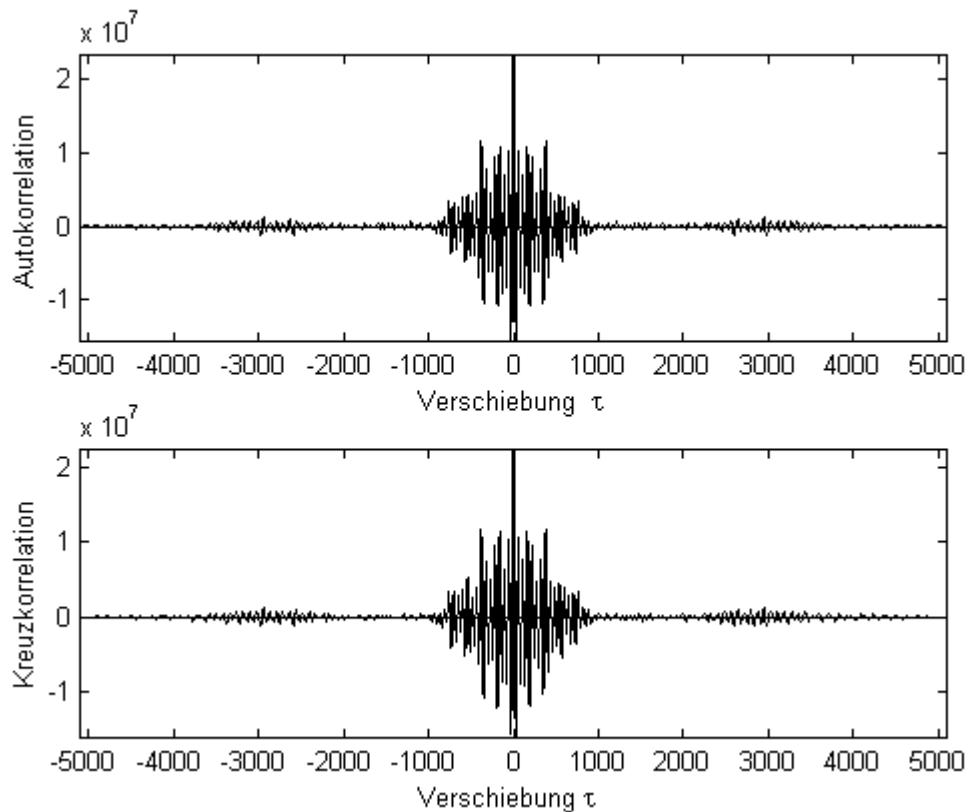


Bild 2.3: Autokorrelations- und Kreuzkorrelationsfunktion des Stereosignals aus Bild 2.2. Die Autokorrelationsfunktion beschreibt die Korrelation des linken Kanals mit sich selbst und die Kreuzkorrelationsfunktion die Korrelation zwischen beiden Kanälen. Die Korrelationsuntersuchung erfolgte für einen Ausschnitt der Länge 5000 Abtastwerte (113 ms).

Bild 2.4 zeigt ein Stereosignal, das kaum Ähnlichkeit des Signalverlaufs von beiden Kanälen aufweist. Es handelt sich hierbei um einen Orchester-Musik-Titel.

Bei klassischer und Orchester-Musik unterscheiden sich die Kanäle stark von einander. Die Instrumente werden sehr unterschiedlich auf die Kanäle verteilt. Eine Abhängigkeit zwischen dem linken und dem rechten Kanal ist in solchen Fällen nicht unbedingt vorhanden bzw. zufällig. Insgesamt finden sich mehr Intervalle, die sich als quasiperiodisch bezeichnen lassen. Das ist damit erklärbar, dass in bestimmten Bereichen Instrumente dominieren, deren Einzelcharakteristik stark harmonisch geprägt ist.

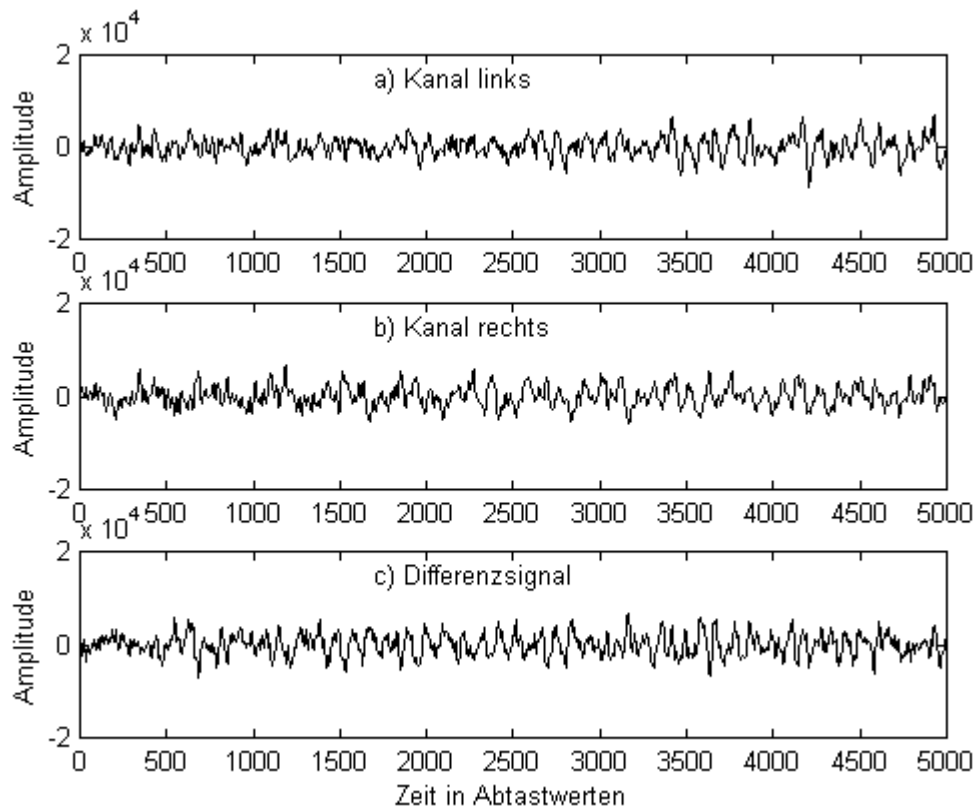


Bild 2.4: Ein Ausschnitt über 5000 Abtastwerte (113 ms) aus dem Stereosignal eines Orchester-Musik-Titels ohne besondere Ähnlichkeit zwischen beiden Kanälen. a) Kanal links, b) Kanal rechts und c) Differenzsignal zwischen beiden Kanälen.

Eine Korrelationsuntersuchung des linken Kanals mit sich selbst (Autokorrelationsfunktion aus Bild 2.5) und zwischen beiden Kanälen (Kreuzkorrelationsfunktion aus Bild 2.5), zeigt, dass die Signalverläufe von beiden Kanälen des Stereosignals nicht übereinstimmend ähnlich sind. Der maximale Kreuzkorrelationswert liegt weit unter dem maximalen Autokorrelationswert. Der Amplitudenunterschied zwischen beiden Maxima liegt hierbei bei ungefähr $3,5 \cdot 10^6$.

Die Tatsache, dass sich im Kreuzkorrelationsverlauf erst bei einer Verschiebung von 250 Abtastwerten das absolute Maximum befindet, deutet auf einen Laufzeitunterschied zwischen beiden Kanälen hin.

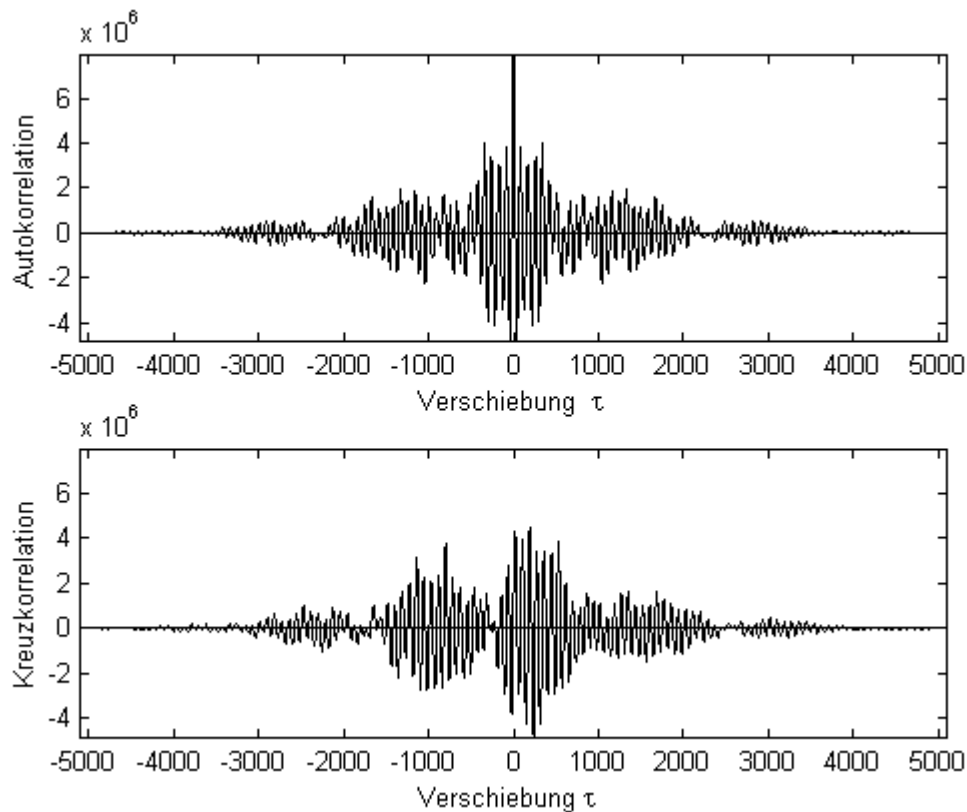


Bild 2.5: Autokorrelations- und Kreuzkorrelationsfunktion des Stereosignals aus Bild 2.4. Die Autokorrelationsfunktion beschreibt die Korrelation des linken Kanals mit sich selbst und die Stereokorrelation die Korrelation zwischen beiden Kanälen. Die Korrelationsuntersuchung erfolgte für einen Ausschnitt der Länge 5000 Abtastwerte.

2.2 Einteilung von Audiosignalen

Bei der Übertragung akustischer Signale in modernen Kommunikationssystemen kann man im allgemeinen davon ausgehen, jede Art von Audiosignalen, wie z.B. Sprache oder Musik, übertragen zu müssen. Die Übertragung von Videosignalen erfordert noch die Einbeziehung von Begleitton wie z.B. Straßenverkehrslärm oder Explosionen. Schon bei der Betrachtung von Musikstücken kann man einige Kategorien unterscheiden: Klassik, Rock, Pop, Techno usw.. Jede Kategorie von Musik besitzt eigene Eigenschaften. Diese unterschiedlichen Eigenschaften wirken sich auch unterschiedlich auf die Wahrnehmbarkeit von Störungen und Artefakte aus.

Sprachsignale lassen sich einfacher analysieren als andere akustische Signale. Es existiert ein Spracherzeugungsmodell, welches für die Codierung mit geringen Bitraten verwendet wird [Epp93].

Bild 2.6 zeigt eine mögliche Einteilung von Audiosignalen. Bei der Betrachtung von Musik ist die Melodie die bedeutendste Komponente. Die Melodie ist eine Folge von Noten oder Grundfrequenzen (*Pitch*) [Mas96]. Die Pitchfrequenz (Tonhöhe) kann als Rate der Wiederholung einer Signalform angesehen werden. Die sich wiederholenden Signalformen, als perfekt periodisch angenommen, können durch eine Fourierreihenentwicklung dargestellt werden. Die Komponenten der Fourierreihe sind sinusförmige Schwingungen mit ganzzahligen Vielfachen der Grundfrequenz. Diese werden harmonische Obertöne (Harmonische) oder auch *Partials* genannt.

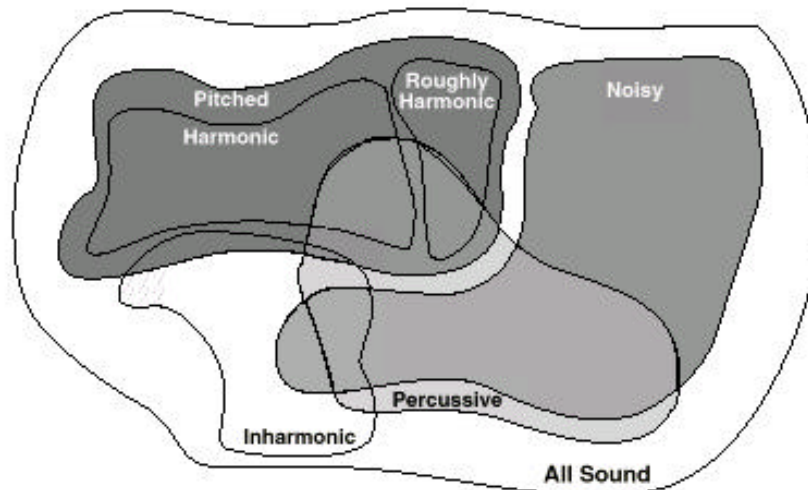


Bild 2.6: Einteilung von Audiosignalen [Mas96]

Oft liegen die Harmonischen nicht genau bei den ganzzahligen vielfachen der Grundfrequenz, in diesem Fall spricht man von Frequenzdehnung (*Frequency stretching*). Hier sollte dann der Begriff Obertöne für die Harmonischen verwendet werden. Das Piano beispielsweise streckt die Frequenzen so, dass das Frequenzverhältnis der Obertöne leicht größer als das der Harmonischen ist. Dies gibt den Eindruck einer gestreckten harmonischen Struktur der Musik [Pie92]. Bei starkem Abweichen von der harmonischen Struktur wird von grober Harmonizität (*roughly harmonic*) gesprochen [Mas96].

Audiosignale enthalten außer den harmonischen Komponenten noch Rauschsignale oder Anschlag-Komponenten (*Percussions*), die im Audiosignal kurze Transienten ohne Grundfrequenz erzeugen und so den Rhythmus z.B. in einem Musikstück bestimmen.

Zu den nicht harmonischen Audiosignalen (*unpitched; inharmonic*), die kein eindeutiges Verhalten im Frequenzbereich aufweisen, gehören beispielsweise Regen oder Ozean-Geräusche.

Zahlreiche Arbeiten haben sich mit der Modellierung und der Analyse von Audiosignalen meistens zwecks Datenkompression beschäftigt. Ein bekannter Begriff in diesem Bereich ist die sinusoidale Modellierung von Audiosignalen, die eine abstrakte mathematische Darstellung von Audiosignalen ermöglicht (Kap. 2.5).

Im Laufe der Jahre wurden viele Modelle entwickelt, welche die sinusoidale Analyse/Synthese als Grundmodell aufnahmen [Pur98], [Rod97], [Mas96]. Im Bereich der Audio-codierung kann so Bandbreite gespart werden, indem nur die Parameter der Sinusschwingungen (*Sinusoids*), die Anzahl der im analysierten Segment vorhandenen Harmonischen und die Rauschkomponenten (parametrische Codierer, z.B. HILN), anstatt aller Abtastwerten des Audiosignals übertragen werden.

Im folgenden Abschnitt werden einige Analysemöglichkeiten von Audiosignalen, die direkt auf das Zeitsignal angewendet werden können, vorgestellt.

2.3 Analyse im Zeitbereich

Die Signalform eines Audiosignals wie z.B. eines Musikinstruments ist meist sehr komplex und lässt sich nicht auf einfache Weise beschreiben. Sie besteht meistens aus vielen überlagerten Schwingungen, wobei sich die Anzahl der Schwingungen, deren Frequenzen und Amplituden schnell ändern können.

2.3.1 Nulldurchgangsrate

Die einfachste Möglichkeit, ein Audiosignal im Zeitbereich zu analysieren, ist es, dessen Nulldurchgangsrate (*zero crossing rate zcr*) zu bestimmen. Die zcr-Berechnung erfolgt segmentweise und ergibt eine Aussage über die Anzahl der Nulldurchgänge (Vorzeichenwechsel) des Signals.

Der zcr-Wert eines Segmentes m ist mathematisch wie folgt definiert:

$$zcr(m) = \frac{1}{2} \sum_{n=0}^{L-2} |\text{sign}(x(n)) - \text{sign}(x(n+1))|, \quad (2.1)$$

wobei

$x(n)$: Abtastwerte des Signals im Zeitbereich,

L : Anzahl der Abtastwerte im Segment m

$\text{sign}()$: Vorzeichenfunktion mit der folgenden Definition:

$$\text{sign}(x(n)) = \begin{cases} 1 & x(n) \geq 0 \\ -1 & x(n) < 0 \end{cases} \quad (2.2)$$

darstellen.

Die Nulldurchgangsrate liefert eine beschränkte Aussage über die Frequenz eines Audiosignals. Audiosignale sind breitbandig und haben daher unterschiedliche zcr-Werte. Hohe Frequenzen führen zu hoher zcr und niedrige Frequenzen zu niedriger zcr (s. Bild 2.7).

Bild 2.7 zeigt ein Beispiel eines zcr-Verlaufs. Das dabei untersuchte Audiosignal ist ein Ausschnitt der Länge 10000 Abtastwerte aus einem Pop-Musik-Titel. Die Ermittlung der zcr-Werte erfolgte segmentweise, wobei die Segmentlänge 100 Abtastwerte beträgt.

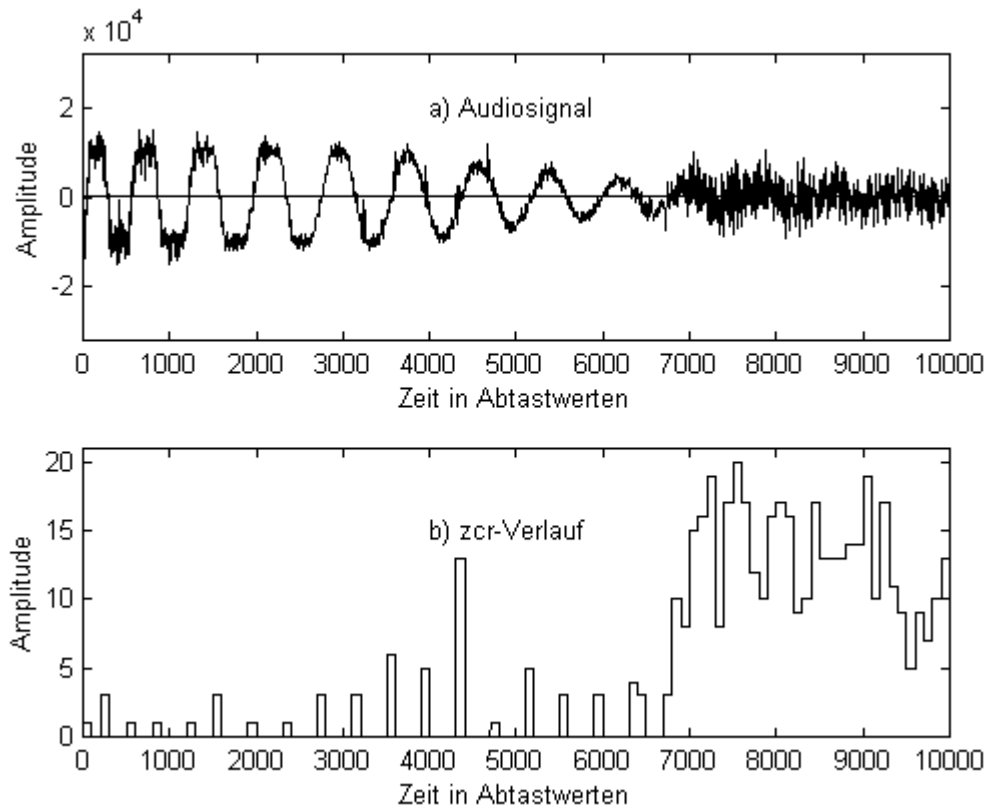


Bild 2.7: zcr-Verlauf eines Audiosignals der Länge 10000 Abtastwerte aus einem Pop-Musik-Titel. Die zcr-Werte wurden segmentweise ermittelt. Die Segmentlänge beträgt hierbei 100 Abtastwerte. a) Audiosignal, b) zcr-Verlauf.

2.3.2 Autokorrelation

Die Autokorrelationsfunktion zählt zu den wichtigsten Verfahren der Signalanalyse im Zeitbereich. Sie gibt eine Aussage über die Ähnlichkeit eines Signals zu sich selbst. Die AKF wird auf zeitdiskrete¹ und zeitkontinuierliche Signale angewendet.

Die AKF eines zeitdiskreten Signals der Länge L ist wie folgt definiert [Epp93]:

¹ Z.B. ein abgetastetes Audiosignal

$$R(k) = \frac{1}{L} \sum_{n=0}^{L-k-1} x(n) \cdot x(n+k) \quad k \geq 0 \quad (2.3)$$

Die AKF ist deswegen so wichtig für die Signalanalyse, weil sie besondere Merkmale aufweist [Epp93]:

- Die AKF hat ihr Maximum bei $k = 0$. $R(0)$ ist auch gleichzeitig die Energie des Signals (Bild 2.8-a und -b)
- Es ist unerheblich, ob die Verschiebung der Funktion nach links oder nach rechts erfolgt. Dies bedeutet, dass die AKF eine gerade Funktion ist und es gilt [Epp93]:

$$R(k) = R(-k) \quad (2.4)$$

- Die AKF weist an der Stelle k ein weiteres Maximum auf, wenn das unverschobene Signal dem um k verschobenen Signal ähnlich ist (Bild 2.8-a).

Aus dem vorhergehenden Punkt ergibt sich:

- Die AKF eines periodischen Signals ist ebenfalls periodisch und hat die selbe Periodendauer. Die Phasenlage des Signals hat keine Auswirkung auf die AKF. Diese Eigenschaft ist von Bedeutung für die Erfassung der Periodizität eines Signals. Sie wird vor allem in der Sprachsignalverarbeitung bei der Bestimmung der Grundfrequenz verwendet (s. Kap. 2.3.3).
- Die AKF eines stochastischen Signals, das keine Regelmäßigkeiten aufweist, hat ihr Maximum bei $k = 0$ und fällt sonst sehr schnell ab (Bild 2.8-b). Auch bei weiterer Verschiebung tritt keine Übereinstimmung mehr auf.

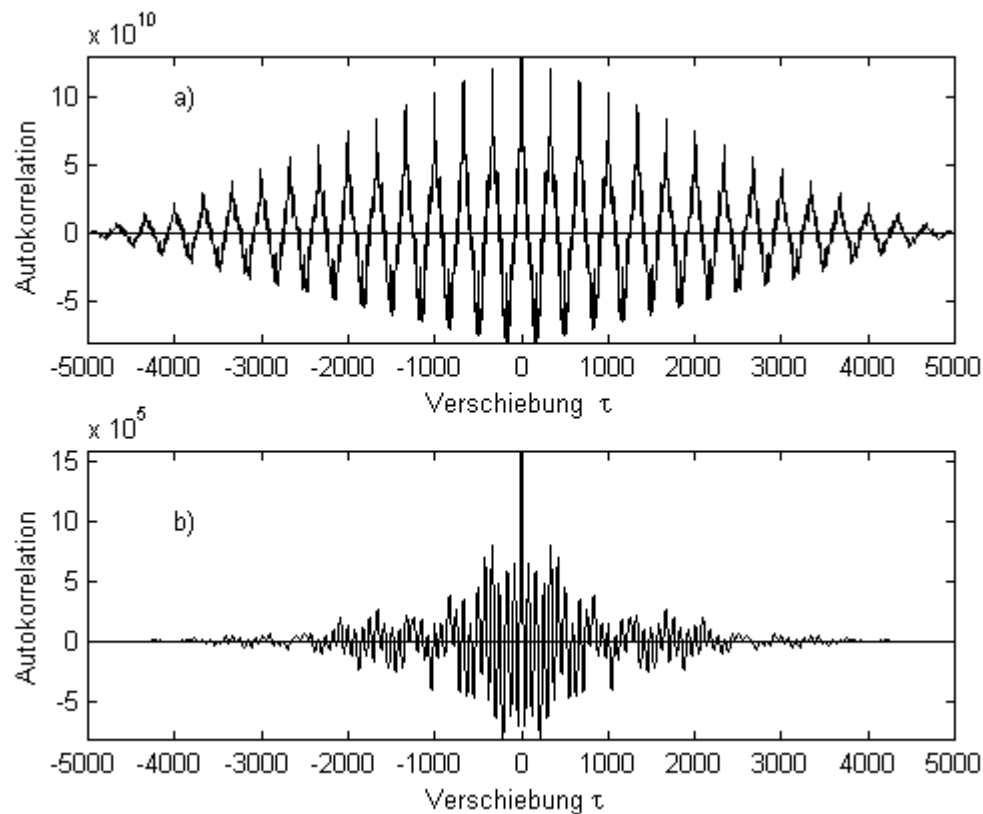


Bild 2.8: Autokorrelationssignale eines näherungsweise periodischen Audiosignalausschnittes aus einem Einzelinstrument-Musik-Titel (Bild a) und eines Audiosignalausschnittes aus einem klassischen Musik-Titel (Bild b), das an dieser Stelle keine besonderen Merkmale aufweist. Beide Audiosignale sind 5000 Abtastwerte lang.

2.3.3 Grundfrequenz

Die Bestimmung der Grundfrequenz spielt bei der Sprachsignalverarbeitung, insbesondere bei der Sprachcodierung, eine besonders wichtige Rolle, da sie der wichtigste Träger prosodischer Informationen (z.B. Betonung, Sprechergeschwindigkeit, etc.) ist [Epp93].

Die Analyse der Grundfrequenz von Audiosignalen wird in dieser Arbeit bei Verfahren der Rekonstruktion von Paketverlusten mittels pitchsynchroner Paketwiederholung (s. Kap. 4.4.1.2) benötigt und aus diesem Grund hier erläutert.

Im Laufe der Zeit ist eine Vielzahl verschiedener Verfahren zur Bestimmung der Grundfrequenz entwickelt worden, die alle ihre Stärken und Schwächen haben [Hes83] [Rab78] [Epp93]. Es gibt leider kein Verfahren, das für alle Signale einwandfrei funktioniert [Epp93]. Im wesentlichen unterscheidet man zwei verschiedene Vorgehensweisen: Die Grundfrequenz wird entweder im Zeitbereich oder im Frequenzbereich bestimmt [Epp93]. In diesem Kapitel

wird das Prinzip der Grundfrequenzbestimmung mit der Autokorrelationsfunktion im Zeitbereich erläutert.

Die Autokorrelation eines Audiosignals mit stark periodischen Anteilen weist Spitzen auf, deren Abstand der Periodendauer der Grundfrequenz entspricht. In diesem Abstand ist das Signal in Phase mit einem verzögerten Signal gleicher Signalform, was zu einer hohen Korrelation führt.

Für die Bestimmung der Grundfrequenz mit der AKF-Methode ist es notwendig, einen Ausschnitt des Audiosignals über mehrere Abtastwerte zu puffern. Die Pufferlänge muss mindestens das zweifache der zu ermittelnden Grundfrequenz sein, damit eine korrekte Ermittlung der Grundfrequenz gewährleistet werden kann. Es wird eine AKF über den Pufferinhalt gebildet und das Maximum der AKF gesucht. Die Pitch-Periode (Grundfrequenz des Signals) wird gleich dem Index dieses Maximums gesetzt. Für eine detaillierte Beschreibung des Verfahrens wird auf [Epp93] und [Clü98a] verwiesen.

Die Grundfrequenz von Sprachsignalen variiert bei ein und demselben Sprecher als auch zwischen verschiedenen Sprechern. Die Werte liegen zwischen 50 und 400 Hz.

2.3.4 Hüllkurven-Detektion

Das dynamische Verhalten eines akustischen Ereignisses im Zeitverlauf kann durch eine Hüllkurve beschrieben werden. Die komplexe Hüllkurve von natürlichen Klängen kann durch eine sogenannte ADSR-Hüllkurve (**A**ttack **D**ecay **S**ustain **R**elease) angenähert werden (s. Bild 2.9).

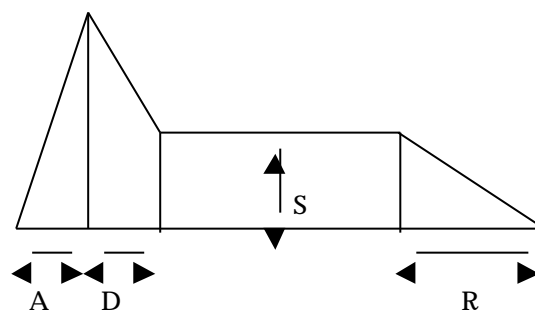


Bild 2.9: ADSR-Hüllkurvenverlauf [WWW1]

ADSR hat die folgende Bedeutung [WWW1]:

- *Attack*: Die Dauer des ersten Abschnitts der Einschwingphase bis zum Erreichen des Zeitpunktes größter Schallintensität.
- *Decay*: Dauer des zweiten Abschnitts der Einschwingphase zwischen dem Erreichen des Zeitpunktes größter Schallintensität und dem Erreichen des *Sustain-Levels*.
- *Sustain*: Pegel des Klanges in seiner stationären Phase (quasistationärer Klangabschnitt)
- *Release*: Dauer der Ausklingphase.

Die Einhüllende des Amplitudenverlaufs eines Audiosignals spielt eine große Rolle bei der Beschreibung und Parametrisierung des Signals. Der Attackenbereich in einem Audiosignal hat den oben gezeigten ADSR-Verlauf.

Im folgenden wird eine Methode zur Attackenerkennung in Audiosignalen beschrieben [Mas96]. Diese Methode wird im Zeitbereich angewendet, dabei wird die Signaluntersuchung blockweise durchgeführt, die untersuchten Blöcke sind nicht überlappend.

Schritt 1:

Es wird das absolute Maximum $M(m)$ des untersuchten Blocks m gesucht. Für $M(m)$ gilt:

$$M(m) = \max_{i=0}^{L-1} \{ |x(mL + i)| \}, \quad (2.5)$$

- L: Anzahl der Abtastwerte eines Blockes,
 $x(n)$: zeitdiskretes Signal,
 i : Abtastwert-Index in jedem Block,
 m : Index des untersuchten Blockes (erster Block $m=0$).

Schritt 2:

Ein Feld $P(m)$ wird bei dem ersten Block erzeugt und während der Untersuchung fortlaufend aktualisiert. $P(m)$ wird *Peak follower* genannt und wie folgt bestimmt:

Ist der Wert $M(m)$ aus Schritt 1 größer als $P(m-1)$ so ist $P(m) = M(m)$, andernfalls entspricht $P(m)$ das Produkt von $P(m-1)$ mit dem Decay-Faktor K_{decay} , der die Steigung der Gerade im Decay-Bereich (Bild 2.9) beschreibt.

Für $P(m)$ gilt also folgende Bestimmungsgleichung:

$$P(m) = \max \{ M(m), P(m-1) \cdot K_{decay} \} \quad (2.6)$$

Schritt 3:

Im letzten Schritt wird untersucht, ob es sich um eine Attacke handelt. Es wird eine Attacke erkannt, wenn

$$\frac{P(m)}{P(m-1)} > Th_{\text{Detection}} \quad (2.7)$$

$Th_{\text{Detection}}$ ist die Schwelle, ab der eine Attacke registriert wird.

2.4 Analyse im Frequenzbereich

In der Zeitdarstellung ist ein Audiosignal meist viel zu komplex, um dessen Eigenschaften ausreichend erfassen zu können. Viel aussagekräftiger ist die Frequenzdarstellung, welche die im Signal enthaltenen Frequenzen abbildet.

2.4.1 Fourieranalyse

Komplexe periodische Klangwellen lassen sich zerlegen. Die nach dem Mathematiker Fourier benannte Fourieranalyse erlaubt es, jede Welle als Summe von Sinuswellen verschiedener Frequenzen und Phasen darzustellen. Ein komplexes Audiosignal kann beispielsweise eine große Anzahl von Teilwellen mit verschiedenen Frequenzen, die alle in einem bestimmten Verhältnis zueinander stehen, enthalten. Man spricht von einem Grundton und von Obertönen, wobei die Obertöne Frequenzen mit ganzzahligen Vielfachen der Frequenz des Grundtones aufweisen. Dieses Ganzzahligkeitsverhältnis ergibt einen Klang, der als harmonisch empfunden wird.

Die Fourier-Transformation eines zeitkontinuierlichen periodischen Signals führt zu einem diskreten Spektrum in der Frequenzebene. Die Transformation einer endlichen Folge zeitdiskreter Werte ergibt ebenfalls ein diskretes Spektrum. Diese Transformation wird als Diskrete Fourier Transformation (DFT) bezeichnet und ist folgendermaßen definiert:

$$X(k) = \sum_{n=0}^{L-1} x(n) e^{-j2\pi nk/L} \quad k = 0 \dots L-1 \quad (2.8)$$

wobei

$x(n)$: Signal in Zeitbereich und

L: Anzahl der Abtastwerte des Signals $x(n)$.

Die Transformation einer diskreten Wertefolge von der Frequenzebene in die Zeitebene erfolgt mit der inversen DFT (IDFT) und ergibt sich wie folgt:

$$x(n) = \frac{1}{L} \sum_{k=0}^{L-1} X(k) e^{j \frac{2\pi nk}{L}} \quad n = 0 \dots L - 1 \quad (2.9)$$

Die DFT basiert auf der Annahme, dass das zu transformierende Signal im Zeitbereich periodisch ist mit der Periodenlänge L des Analysefensters. Das Ergebnis im Frequenzbereich würde genau der Transformierten des abgetasteten periodischen Eingangssignals entsprechen. Im günstigsten Fall, wenn man das Fenster im Zeitbereich an den Anfang einer Periode legt und den Analysebereich gerade ein Vielfaches der Periodendauer einer Sinus-schwingung lang wählt, bekommt man idealerweise nur einen Peak beim Maximum der Sine-funktion¹ und die restlichen Abtastwerte im Frequenzbereich liegen genau bei den Nullstellen der Funktion. Doch diesen Idealfall wird man in der Praxis so gut wie nie erreichen, so dass man keinen einzelnen Peak nach der Transformation mit der DFT bekommt, sondern schlimmstenfalls zwei gleichgroße Peaks und noch viele andere, die langsam zu den Frequenzbereichsgrenzen in der Amplitude abnehmen.

Zur spektralen Analyse eines Audiosignals wird eine überlappende DFT verwendet. Die DFT wird auf gefensterete Abschnitte des Signals durchgeführt, wobei die Fensterlänge die DFT-Länge entspricht. Die Verschiebung des Fensters geschieht überlappend, um die im Audiosignal enthaltenen Signalübergänge besser erfassen zu können.

Wahl des Fensters

Bevor eine DFT implizit durchgeführt werden kann, muss das Audiosignal mit einem Fenster multipliziert werden. Dies ist notwendig, um den sogenannten Leckeffekt zu vermindern. Der Leckeffekt entsteht, wenn die Fenstergröße nicht mehr der Grundperiode des Audiosignals entspricht. Es treten Seitenschwingungen auf, die das Spektrum verfälschen. Ursache hierfür ist das langsame Abklingen des Rechteckspektrums, was auf die steile Flanke des Rechteckfensters im Zeitbereich zurückzuführen ist. Um Nebenschwingungen soweit wie möglich zu dämpfen, sollte das Fenster nicht rechteckig, sondern geformt sein.

Diesen Effekt kann man durch andere Formen der Fenster im Zeitbereich beeinflussen. Dadurch verändert sich die Transformierte des Fensters im Frequenzbereich und man erhält andere Eigenschaften. Solche Fenster gewichten das Zeitsignal, so dass bestimmte Bereiche des Fensters stärker auf das Signal auswirken als andere. Dabei gibt es verschiedene Varianten von Fenster, bei denen zum Beispiel die Hauptkeule der Transformierten sehr schmal ist,

¹ In diesem Fall wurde davon ausgegangen, dass ein Rechteckfenster verwendet wird.

dafür die Nebenmaxima erst langsam nach außen abfallen. Oder man wählt ein Fenster, bei dem die Hauptkeule breiter ist. Die Nebenmaxima fallen bei diesem Fenster schneller ab. Einige wichtige Eigenschaften der verschiedenen Fenster sind Beeinflussung der Leistung, sowie eine Verbreiterung der Hauptkeule. Am Rand zu Null abflachende Fenster erzeugen eine leichte Frequenzerhöhung bei tiefen Frequenzen. Die verschiedenen Fenster und die entsprechenden spezifischen Eigenschaften können [HAR78] entnommen werden.

Ungenauigkeit der DFT

Mit der DFT werden L Abtastwerte in $L_{\text{DFT}} = L$ diskrete Spektralkoeffizienten im Frequenzbereich transformiert. Die diskreten Frequenzen mit ihren Frequenzindizes k werden gleichmäßig mit dem Abstand Δf auf der Frequenzachse verteilt:

$$\Delta f = \frac{1}{T \cdot L_{\text{DFT}}} \quad (2.10)$$

wobei T den zeitlichen Abstand zwischen benachbarten Abtastwerten darstellt. Δf wird als Auflösungsvermögen der DFT bezeichnet.

Der ungünstigste Fall bei der Untersuchung eines Audiosignals tritt ein, wenn die wirkliche Frequenz genau zwischen zwei diskrete Frequenzen fällt. In diesem Fall zeigen beide Fourierkoeffizienten hohe Werte an. Die genaue Frequenz kann nur annähernd mit Interpolation bestimmt werden (Anhang B.2). Um die im Signal enthaltenen Frequenzen genauer bestimmen zu können, wird *Zeropadding* benutzt.

Zeropadding

Zeropadding bedeutet allgemein, dass man einen Datenbereich, für den man nicht genügend Werte zur Verfügung hat, mit Nullen auffüllt. Eine typische Anwendung für *Zeropadding* ist bei der DFT gegeben. Transformiert man nämlich damit einen Datensatz vom Zeitbereich in den Frequenzbereich, so bekommt man dieselbe Anzahl an Werten im Frequenzbereich heraus. Oftmals ist eine höhere Auflösung im Frequenzbereich erwünscht als Datenwerte im Zeitbereich zur Verfügung stehen. In diesem Fall füllt man einfach den Datenblock bis zur gewünschten Größe mit Nullen auf.

Mit *Zeropadding* erhält man keine neuen Informationen über das Eingangssignal und es treten damit auch keine neuen Frequenzanteile dieses Signals auf, denn mit dieser Operation erhöht man nur die Frequenzauflösung, was mathematisch einer Interpolation der Frequenzen entspricht.

¹ $T = 1 / f_a$. f_a ist die Abtastfrequenz des Audiosignals.

2.4.2 Diskrete Cosinustransformation

Die Diskrete Cosinustransformation (DCT) stellt einen speziellen Fall der DFT dar. Es handelt sich um eine reellwertige Transformation und sie weist wesentlich geringe Blockgrenzeffekte als die DFT auf: Die DCT eines L -dimensionalen Signalvektors entspricht einer DFT eines $2L$ -dimensionalen Vektors, der zu Hälfte die L Komponenten des Signalvektors und zur anderen Hälfte dieselben Komponenten in umgekehrter Reihenfolge enthält. Bei der inversen DFT ergibt sich eine periodische Fortsetzung des $2L$ -dimensionalen Vektors und damit abwechselnd die ursprüngliche und die gespiegelte Signalfolge. Dadurch treten keine Sprünge an den Blockgrenzen der DFT mehr auf, woraus sich für die DCT geringeren Störungen ergeben als bei einer DFT gleicher Ordnung [Clü98b].

Die Basisvektoren bestehen aus abgetasteten Cosinus-Verläufen. Die Komponenten des Basisvektors t_i ergeben sich zu

$$t_{ik} = A_i \sqrt{\frac{2}{L}} \cos \frac{i(k + \frac{1}{2})\pi}{L} \quad i, k = 0 \dots L - 1, \quad (2.11)$$

mit

$$A_0 = \frac{1}{\sqrt{2}}, \quad A_i = 1 \quad \text{für } i = 1 \dots L - 1 \quad (2.12)$$

2.5 Die sinusoidale Modellierung von Audiosignalen

Die Entwicklung von Modellen, welche beschreiben, wie sich akustische und im besonderen musikalische Signale zusammensetzen, spielen bei der Synthese und Transformation von Musik, sowie bei deren Datenkompression, eine große Rolle.

Bei heutigen Audiosignalen haben wir es nicht nur mit natürlichen Klängen, sondern immer mehr auch mit synthetisch erzeugten Audiosignalen zu tun. Es existieren Synthesetechniken, die auch natürliche Klänge täuschend echt nachbilden können. Dies legt es nahe, eine Analyse von Klängen vorzunehmen, um diese dann nur noch parametrisch, d.h. mit den Anregungsparametern des passenden Synthesemodells, zu repräsentieren [Ohm01].

Bei der Analyse eines Audiosignals, welches oft durch die Überlagerung einer Anzahl von sinusoidalen Schwingungen angenähert werden kann, wird das Signal in seine Sinusschwingungen zerlegt und seine Parameter werden extrahiert. Bei der Synthese bzw. Rekonstruktion dieses Signals werden die einzelnen Sinusschwingungen mit ihren Amplituden,

Frequenzen und Phasen rekonstruiert und aufsummiert. Dieser Vorgang wird additive Synthese genannt.

Das sinusoidale Modell bietet die Möglichkeit einzelne Komponenten zu manipulieren, um Effekte wie z.B. *Time-Stretching* oder *Pitch-Shifting* in einem Audiosignal zu erzeugen.

Die Idee dieses Modells wird auch im Bereich der Audiocodierung verwendet. Als Beispiel einer parametrischen Codierung von Audiosignalen ist das im Standard MPEG-4 definierte **Harmonic and Individual Lines plus Noise** (HILN)-Verfahren zu nennen. Mit diesem Verfahren kann Bandbreite gespart werden, indem nur die Parameter der sinusoidalen Schwingungen und die der Harmonischen sowie die Rauschkomponenten übertragen werden [Pur98].

Im Laufe der Jahre wurden viele Modelle entwickelt, welche die sinusoidale Analyse/Synthese als Grundmodell annahmen. Es sind u.a. folgende Beispiele zu nennen:

- Der Phasenvocoder [Kah98].
- Die MQ-sinusoidale Analyse/Synthese. 1986 durch McAulay und T. Quartieri entstanden [McA86].
- Die wichtigste Erweiterung dieses Modells ist die Analyse durch Synthese, auch als *Deterministic plus stochastic decomposition* bekannt [Ser89].

2.5.1 Grundlagen des sinusoidalen Modells

In diesem Kapitel werden einige grundsätzliche mathematische Ansätze der sinusoidalen Modellierung von Audiosignalen vorgestellt.

Ein effizienter Ansatz zur Lösung der Problematik von Paketverlusten in Audiosignalen basiert auf der Idee des MQ-Modells. In diesem Modell wird ein Audiosignal $x(t)$ als Summe von Sinus-Schwingungen mit der Augenblicks-Phase $\varphi_i(t)$ und Amplitude $A_i(t)$, beschrieben. Diese ändern sich nur langsam mit der Zeit.

$$x(t) = \sum_{i=1}^{I(t)} A_i(t) \exp(j\varphi_i(t)) \quad (2.13)$$

$$\varphi_i(t) = \int_{-\infty}^t \omega_i(\tau) d\tau. \quad (2.14)$$

$I(t)$ ist eine ganzzahlige Funktion, welche die Anzahl der betrachteten Sinuskomponenten in Abhängigkeit von der Zeit beschreibt; ein Audiosignalausschnitt, der dem Zeitpunkt t zugeordnet ist, wird durch eine ganzzahlige Anzahl $I(t)$ von Sinusschwingungen angenähert.

$A_i(t)$, $\varphi_i(t)$ und $\omega_i(t)$ sind die Augenblicks- Amplitude, -Phase und -Frequenz der i -ten Sinuskomponente.

Ein Audiosignal besteht meistens nicht nur aus Sinuskomponenten, sondern auch aus stochastischen Komponenten $e(t)$.

$$x(t) = \sum_{i=1}^{I(t)} A_i(t) \exp(j\varphi_i(t)) + e(t) \quad (2.15)$$

Die Exponential-Funktion kann auch durch eine Cosinus-Funktion mit dem Zusatz einer Anfangsphase φ_{0i} ersetzt werden. Unter der Annahme der Quasiperiodizität wird davon ausgegangen, dass die Amplituden- und Phasen-Änderung konstant bleiben, was sich in einer konstanten Frequenz äußert, denn es gilt:

$$\omega_i(t) = \frac{d\varphi_i(t)}{dt}. \quad (2.16)$$

Für zeitdiskrete Audiosignale bedeutet das:

$$x(n) = \sum_{i=1}^{I(n)} A_i \cos(\omega_i \cdot n + \varphi_{0i}) + e(n) \quad (2.17)$$

$I(n)$ beschreibt die Anzahl der Sinuskomponenten zum diskreten Zeitpunkt n . Die Berechnung der Augenblicksphase erfolgt gemäß:

$$\varphi_i(n) = \omega_i \cdot n + \varphi_{0i}, \quad n = 0 \dots L - 1. \quad (2.18)$$

L ist die Länge des untersuchten Audiosignalsegments in Abtastwerten.

2.5.2 Sinusoidale Analyse

Mit dem sinusoidalen Modell werden alle Sinuskomponenten eines Audiosignals extrahiert (analysiert) und danach wieder rekonstruiert (synthetisiert).

Bild 2.10 beschreibt das Prinzip der sinusoidalen Analyse.

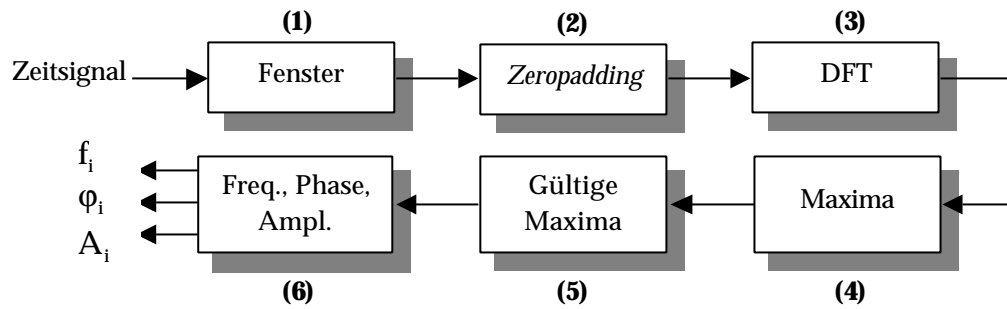


Bild 2.10: Prinzip der sinusoidalen Analyse von Audiosignalen

Die Analyse des Audiosignals erfolgt segmentweise. Das Audiosegment wird gefenstert (1) und mit der DFT in den Frequenzbereich transformiert (3). Um die Frequenzauflösung zu erhöhen wird vorher *Zeropadding* durchgeführt (2). Aus dem Betragsspektrum werden alle lokalen Maxima bestimmt (4). Die lokalen Maxima werden dabei anhand einer Schwelle für gültig oder ungültig bewertet (5). Im Originalmodell wird die Schwelle auf -80 dB festgelegt [Mas96]. Alle Maxima unterhalb dieser Schwelle werden als nicht gültig betrachtet. Es werden dann die Amplituden, Phasen und Frequenzen dieser Maxima ermittelt (6).

2.5.3 Sinusoidale Synthese

Nachdem ein Segment analysiert und dessen Sinuskomponenten ermittelt wurden, müssen nun die aufeinander folgenden Sinuskomponenten verfolgt werden, bevor sie wieder zu einem Audiosignal synthetisiert werden. Da sich die Sinuskomponenten in ihren Parameter über der Zeit ändern können, müssen bei aufeinanderfolgenden Analysen zusammengehörige Sinuskomponenten zu einem Paar zugeordnet werden.

In jedem Zeitabschnitt werden aus dem Betragsspektrum die lokalen Maxima und aus ihnen die Sinuskomponenten bestimmt. Unter der Annahme, dass sich die Sinuskomponenten über mehrere Segmente erstrecken, können Sinuskomponenten gleicher bzw. ähnlicher Frequenz als Spur (*Track*) betrachtet werden. Mit ähnliche Frequenz ist gemeint, dass sich die Frequenz nur langsam gegenüber der Zeit ändert. Beim Entstehen und Ausklingen von Sinuskomponenten wird von *Birth* und *Death* gesprochen [McA86].

Für die Synthese müssen die *Tracks* der analysierten Segmente miteinander verbunden werden. Ein einzelner *Track* $y_i(n)$ eines Segments wird mit einer Cosinus-Schwingung erzeugt, dessen Augenblicksamplitude $A(n)$ und Augenblicksphase $\varphi_i(n)$ bei jedem Abtastwert n aktualisiert werden.

$$y_i(n) = A_i(n) \cdot \cos(\varphi_i(n)) \quad n = 0, 1 \dots L - 1 \quad (2.19)$$

Die Synthese des Audiosignals erfolgt dann gemäß

$$y(n) = \sum_{i=1}^I y_i(n) = \sum_{i=1}^I A_i(n) \cdot \cos(\phi_i(n)), \quad (2.20)$$

wobei I die Anzahl der *Tracks* bezeichnet.

Da die Parameter von aufeinander folgenden Segmenten nicht unbedingt genau gleich sind, müssen diese stetig ineinander übergehen. Für die Synthese bietet sich eine lineare oder kubische Amplituden- und Phaseninterpolation an. Im folgenden werden diese erläutert, da sie auch später in Kapitel 4 für die Rekonstruktion der Lücken in Audiosignalen verwendet werden.

Bei der Synthese müssen die *Tracks* des letzten Segments $k-1$ $\{f_i^{k-1}, \phi_i^{k-1}, A_i^{k-1}\}$ und des aktuellen Segments k $\{f_i^k, \phi_i^k, A_i^k\}$ miteinander verbunden werden. Die Amplituden des analysierten Segments beziehen sich auf das Zentrum des Segments, falls eine symmetrische Fensterung vor der Frequenztransformation angewendet wurde (z.B. Hamming oder Hanning). Die Phase bezieht sich aber, wenn man vor der DFT ein *Zeropadding* am Ende des Segments anwendet, auf den Anfang des betrachteten Segments. Bei überlappenden Fenstern verschiebt sich der Bezugspunkt.

Lineare Amplitudeninterpolation

Für die zu rekonstruierenden Amplituden bietet sich die lineare Interpolation an.

$$A_i^k(n) = A_i^{k-1} + (A_i^k - A_i^{k-1}) \cdot \frac{n}{L}, \quad n = 0, 1, \dots, L-1 \quad (2.21)$$

Für die Amplitude wird folgende Bezeichnung verwendet: $A_i^k(n)$

wobei

- A: Amplitude,
- n: aktueller Abtastwert im Segment,
- k: k-te Segment im Audiosignal und
- i: Nummer des Frequenztracks.

Es handelt sich also um die Augenblicksamplitude zum diskreten Zeitpunkt n des i -ten Frequenztrack im k -ten Segment.

Für die Phaseninformation bieten sich die lineare [Mas94] oder die kubische Interpolation [McA86] an.

Kubische Phaseninterpolation

Bei der Interpolation der Phase wird hier von einem kubischen Funktionsverlauf ausgegangen. Die Augenblicksphase wird mit Gleichung 2.18 folgendermaßen definiert:

$$\varphi_i^k(n) = \varepsilon + \kappa \cdot n + \alpha \cdot n^2 + \beta \cdot n^3, \quad n = 0 \dots L. \quad (2.22)$$

Mit $\varphi_i^k(n)$ wird die Augenblicksphase des i-ten Tracks im k-ten Segment bezeichnet.

Für die Synthese gelten folgende Nebenbedingungen:

$$\varphi_i^k(0) = \varphi_{i0}^k = \varepsilon \quad (2.23)$$

und

$$\dot{\varphi}_i^k(0) = \omega_i^k = \kappa \quad (2.24)$$

φ_{i0}^k ist die Anfangsphase in Segment k.

$\dot{\varphi}_i^k$ ist erste Ableitung von φ_i^k nach der Zeit.

ω_i^k ist die Frequenz in Segment k.

Die Aufstellung des Gleichungssystems zur Bestimmung der Parameter α und β ist in [McA86] und [Kah98] erläutert. Die Berechnung der Phase sollte immer Modulo 2π erfolgen, was bei der Berechnung die Addition von einem $2\pi M$ -Faktor bedeutet. Mit dieser Maßnahme wird die Frequenz maximal flach gehalten (*maximally smooth* [McA86]).

M ist ein ganzzahliger Wert und entspricht den nächsten Integerwert von x aus der folgenden Gleichung [McA86],[Kah98].

$$x = \frac{1}{2\pi} \left[(\varphi_0^k + \omega^k \cdot L - \varphi_0^{k+1}) + (\omega^{k+1} - \omega^k) \cdot \frac{L}{2} \right] \quad (2.25)$$

für α und β , die jetzt von M abhängig sind, gilt:

$$\begin{pmatrix} \alpha(M) \\ \beta(M) \end{pmatrix} = \begin{bmatrix} \frac{3}{L^2} & -\frac{1}{L} \\ -\frac{2}{L^3} & \frac{1}{L^2} \end{bmatrix} \cdot \begin{bmatrix} \varphi_0^{k+1} - \varphi_0^k - \omega^k \cdot L + 2\pi M \\ \omega^{k+1} - \omega^k \end{bmatrix} \quad (2.26)$$

Mit den nun bekannten Parametern ε , κ , α und β lässt sich die Phase vollständig bestimmen.

3 Audioübertragung über paketvermittelnde Netze

Der Schwerpunkt dieser Arbeit liegt in der Untersuchung und Bewertung von Verfahren, mit denen eine optimale Rekonstruktion von Lücken in Audiosignalen ermöglicht wird. Dazu ist es zunächst mal wichtig, die möglichen Ursachen und Anwendungen, bei denen solche Lücken entstehen können, zu betrachten.

In diesem Kapitel werden das Prinzip der Übertragung von Audiodaten über paketvermittelnde Netze betrachtet sowie deren Vorteile und Nachteile benannt.

Die Übertragung von digitalisierten Daten wie Sprache, Audio- oder Videodaten über paketvermittelnde Netze (**PSN: Packet Switched Networks**) wie das Internet hat in den letzten Jahren zunehmend an Bedeutung gewonnen. Im Bereich der Audiotechnik, die jahrzehntlang überwiegend analog konzipiert wurde, findet zur Zeit ein großer Umbruch statt. Die Übertragung von digitalen Audiodaten über das Internet eröffnet völlig neue Perspektiven. Videokonferenzen, Internet-Telephonie, Internet-Radio und Online-Multimedia-Services sind solche Anwendungen. Sie erfordern eine Echtzeit-Übertragung der Daten und stellen deswegen besondere Anforderungen an das Übertragungs-Medium und das Protokoll.

In den Anfängen der Audioübertragung über das Internet mussten die Daten zuerst vollständig auf den anfordernden Computer (*Client*) übertragen werden, um dann abgespielt zu werden, d.h. die Audiodaten wurden *offline* abgehört. Diese Übertragungsart wurde mit „*File Transfer*“ bezeichnet, sie war anfangs aufgrund der geringen Übertragungsgeschwindigkeit und vor allem der in der Regel schlechten Durchgängigkeit des Übertragungskanals notwendig. Auch heute noch werden viele Audiodaten *offline* gehört, daneben wird aber das *online*-Hören immer üblicher, d.h. die individuell angeforderten Audiodaten werden –wie beim terrestrischen Radio- in einen Datenstrom gewandelt, der bei seinem Empfang sofort abgehört wird. Diese zweite Variante der Audioübertragung wird mit „*Real-Time file streaming*“ bezeichnet.

Für die Verbreitung von Audiosignalen über das Internet werden diese komprimiert (zumeist nach dem MPEG-1 oder MPEG-2 Standard). Von einem Computer (*Server*) werden die Audiodaten dann auf Anforderung in einzelnen Paketen an den anfordernden Computer (*Client*) gesendet, der sie mittels Software dekomprimiert und über eine Soundkarte wieder in hörbare Audiosignale wandelt oder sie auf dem PC ablegt.

Die von dem *Server* generierten Audiopakete bestehen im allgemeinen aus den folgenden Komponenten:

- Einem Paketkopf, der Zusatzinformationen enthält, z.B. Empfangsadresse, Sendeadresse, Paketlänge, Paketnummer und Kontrollbits (*cyclic redundancy check* CRC) zur Fehler-

lererkennung (s. Bild 3.1). Die genaue Zusammensetzung des Paketkopfes kann von Netz zu Netz variieren.

- Einem Pakettrumpf, der die Teilnachricht enthält.

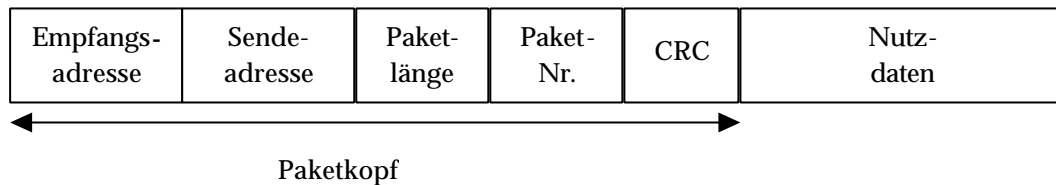


Bild 3.1: Prinzipieller Aufbau eines Paketes [Nol86].

Die zu übertragenden Audiodaten stellen hinsichtlich der Bitrate einen Bereich von wenigen kbit/s bis zu einigen Mbit/s dar. Um unterschiedliche und zeitvariable Informationsmengen wirtschaftlich in einem Kommunikationsnetz zusammenfassen zu können, muss das Netz unterschiedliche Bitraten zur Verfügung stellen und dabei die Kapazität der Übertragungskanäle so gut wie möglich ausnutzen.

Zur Übermittlung unterschiedlicher Bitraten müsste ein Netz mit synchroner Übertragung der binären Signale mehrere Bitratenklassen anbieten, für die getrennte Vermittlungseinrichtungen erforderlich wären. Aus Gründen des Aufwandes und der Wirtschaftlichkeit könnten nur wenige Klassen realisiert werden. Das hätte zur Folge, dass viele Dienste die jeweilige Kanalkapazität nicht voll ausnutzen würden. Eine bessere Kanalausnutzung bei einheitlicher Netzstruktur ergibt sich bei asynchroner Übertragung.

Der *Asynchronous Transfer Mode* (ATM) und das **Internet-Protokoll** (IP) stellen heute die interessantesten und wichtigsten Netzwerkprotokolle der Breitbandkommunikation dar.

3.1 Asynchronous Transfer Mode (ATM)

ATM ist ein Verfahren zur Datenkommunikation, das auf einem einfachen verbindungsorientierten Paketvermittlungsverfahren beruht. Aufgrund hoher Bandbreiten und großer Übertragungsgeschwindigkeiten ist man mit ATM in der Lage, sehr viele Anwendungen zu ermöglichen. Im Bereich des *Mobile Office* existiert bereits eine verstärkte Nachfrage nach Videokonferenzsystemen, die den Einsatz breitbandiger Technologien wie ATM fordern. In der Medizintechnik wird die ATM-Technologie bereits schon seit einiger Zeit im Bereich der Operationssäle angewandt. Gemeinschaftliches Arbeiten via Video ist hier ein wesentliches Einsatzgebiet. Die ATM-Technologie stellt für die kommerziellen Anwendungen wie z.B. das Virtuelle *Shopping* eine wichtige technische Grundlage.

ATM wurde weltweit von den Normierungsgremien als Multiplex- und Vermittlungsprinzip für interaktive Dienste und Anwendungen gewählt, die eine Bitrate von deutlich über 64 kbit/s erfordern, und es werden als Basis des Breitband-ISDN definiert.

ATM-Daten werden nach der Empfehlung I.361 des CCITT in Zellen á 53 Byte assembliert und mit einer Bitrate von 155 Mbit/s übertragen [Aßm91]. Die Bitrate der zu übertragenden Daten bestimmt die Anzahl der pro Zeiteinheit gesendeten Zellen. Unterschiedliche Bitraten werden dadurch realisiert, dass die Anzahl der Zellen sich nach dem Bitratenbedarf des Quellsignals richtet. Die Nachrichtenquelle schickt je nach Bedarf des Empfängers die geforderten Informationspakete über das Netz. Es besteht ein **asynchroner** Zellenbetrieb. Weiterhin werden nur dann Zellen gesendet, wenn auch Informationen übertragen werden. In den Zwischenzeiten steht die nicht benötigte Kanalkapazität für andere Benutzer zur Verfügung. Dies erhöht die mögliche Kanalausnutzung, die bereits durch die Verschachtelung von Zellen unterschiedlicher Signale auf einem physikalischen Kanal relativ hoch ist. Außerdem können örtliche Überlastungen des Netzes durch entsprechende Zuteilung der Übertragungswege vermieden werden [Clü93].

3.1.1 Aufbau einer ATM-Zelle

Von den 53 Byte einer ATM-Zelle bilden 5 Byte den Zellenkopf und die übrigen 48 Byte das Informationsfeld. Die Zellen werden aufeinanderfolgend jeweils mit dem Zellkopf beginnend übertragen. Der Zellenkopf enthält Zielangabe und Steuerinformationen, die für die Wegfindung zwischen dem Sender und Empfänger zuständig sind. Jede Zelle ist für sich bestrebt das Ziel zu erreichen.

Den Aufbau des Informationsfeldes einer Zelle vom Typ 1 zeigt Bild 3.2. Das Erste Byte des Informationsfeldes besteht aus einer Sequenznummer (SN) von 4 Bit Länge und vier Fehler-schutzbits (*Sequence Number Protection* SNP). Die 4-bit-Sequenznummer SN dient zur Erkennung von Zellenverlusten oder fehlgeleiteten Zellen.

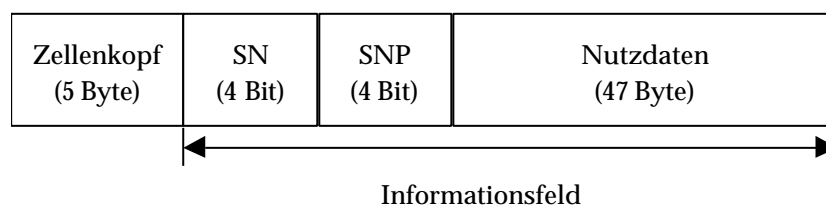


Bild 3.2: Prinzipieller Aufbau einer ATM-Zelle vom Typ 1 [Clü93].

Auf dem Weg vom Sender zum Empfänger passieren die Zellen meist mehrere Netzknoten, die den Zellenkopf lesen und die Zellen zum nächsten Knoten weiterleiten. Die Übertragung der ATM-Zellen erfolgt über virtuelle Pfade, so dass - anders als in herkömmlichen Paket-

netzen, in denen es in variablen Übertragungswegen zu Überholvorgängen kommen kann - die Reihenfolge der gesendeten Zellen erhalten bleibt [Gol77]. Das Netzwerk garantiert somit, dass die Zellen einer Verbindung den selben Weg verfolgen und demzufolge in der gesendeten Reihenfolge beim Empfänger eintreffen.

Die Datenübertragung findet unter Beachtung genormter Vorschriften statt. Ein Protokoll legt fest, welche Informationen und in welcher Reihenfolge die Informationen ausgetauscht werden müssen, damit eine sinnvolle und definierbare Kommunikation erfolgen kann.

3.1.2 ATM-Referenzmodell

Die Assemblierung, Wegbestimmung und Übertragung der ATM-Zellen wird nach einem für ATM festgelegten Referenzmodell durchgeführt. Dieses Referenzmodell beinhaltet eine Bitübertragungsschicht, auch physikalische Schicht genannt (*Physical layer PL*), eine ATM-Schicht (*ATM layer AL*) und ATM-Anpassungsschicht (*ATM adaption layer AAL*).

Die physikalische Schicht beinhaltet die Funktionen zur Bitübertragung der Zellen. Diese Schicht ist für die Übermittlung und den Zusammenbau des Bitstroms in Zellen sowie für die Signalwandlung, Synchronisierung und Verstärkung des physikalischen Übertragungssignals verantwortlich. Die physikalische Schicht übernimmt unter anderem die Zellensynchronisation und die Erzeugung von Leerzellen, wenn von der darüber liegenden ATM-Schicht keine Zellen zur Übertragung vorliegen.

Die ATM-Schicht übernimmt den Zellentransport inklusive der Vermittlung der Zellen. Sie hat die Aufgabe, die von der übergeordneten Anpassungsschicht übergebenen Daten an den Bestimmungsort zu transportieren. AL übernimmt in der Übertragungsrichtung das Multiplexen der Zellen verschiedener Verbindungen. Dabei wird zwischen logischen Verbindungen (*virtual channel VC*) und Verbindungsbündeln (*virtual path VP*) unterschieden. Die logische Verbindung entspricht den Verbindungen, die von der ATM-Anpassungsschicht AAL, der nächst höheren Schicht, zwischen den Endpunkten der Verbindung verwendet werden. Erhält die ATM-Schicht von der übergeordneten AAL-Schicht eine Informationseinheit, so erzeugt sie den dazugehörigen ATM-Header. Eine wichtige Aufgabe ist dabei die Umwandlung der Netzwerkadresse der höheren Schichten in VCI (*virtual channel identifier*) und VPI (*virtual path identifier*).

In der AAL sind alle Aspekte der Anpassung der Nutzdaten an das Zellenformat berücksichtigt. Diese Schicht bildet also das Bindeglied zwischen den Anwendungen der übergeordneten Schichten und der ATM-Schicht. Der von den Anwendungen erzeugte Datenstrom kann verschiedenste Eigenschaften haben. Sie können beispielsweise einen Datenstrom mit konstanter oder variabler Datenrate erzeugen. Die Aufgabe der Anpassungsschicht ist es, die

verschiedenen Anforderungen der Datenübertragung auf der Sendeseite jeweils auf ATM-Zellen abzubilden und auf der Empfangsseite aus den empfangenen Zellen wieder die Nutzdaten der Anwendungen zu erzeugen. Abhängig von der Art des Dienstes kann eine weitere Aufgabe im Ausgleich der Zellenlaufzeiten bestehen. Dies gilt insbesondere auch für Sprachübertragungen, wo auf der Sendeseite das Informationsfeld der ATM-Zellen mit dem zeitkontinuierlichen Nutzdatenstrom gefüllt wird und auf der Empfangsseite eine Umwandlung der empfangenen Zellen in eine kontinuierliche Information erfolgt.

Die Funktionen des AAL hängen von den Anforderungen der oberen Schichten ab. Der AAL unterstützt mehrere Protokolle. Aufgrund der Breite der angebotenen Dienste, kam die ITU-T nicht mit der Definition eines einzigen AAL aus. Es wurden vier Typen definiert Typ1, Typ2, Typ 3 / 4 und Typ5. Die Definition der verschiedenen AAL-Typen basierte auf den Anforderungen hinsichtlich des Zeitbezuges (zeitkontinuierlich, nicht zeitkontinuierlich), der Bitrate (konstante, variable) und der Verbindungsart (verbindungsorientiert, verbindungslos).

Der AAL-Typ1 wurde für Anwendungen mit konstanter Datenrate und Echtzeitanforderungen eingeführt. AAL-Typ2 wurde für die Übertragung von zeitkontinuierlichen Nutzdaten mit variabler Bitrate vorgesehen. AAL Typ3/4 und Typ5 spezifizieren die verbindungsorientierte und verbindungslose Übertragung von Datenpaketen, für die keine zeitliche Korrelation zwischen Sender und Empfänger besteht. Die dabei aufgebauten Verbindungen können sowohl Punkt-zu-Punkt- als auch Punkt-zu-Mehrpunkt-Verbindungen sein.

In [Hän94] und [Sie93] befindet sich eine weitergehende Beschreibung der verschiedenen AAL-Typen.

3.2 Internet-Protokoll

Das folgende Kapitel gibt eine Übersicht über verschiedene Protokolle, die IP Multicast nutzen. Ursprünglich war das Internet dazu gedacht, Text und Grafiken mit keiner oder nur minimaler Verzögerung zu übertragen. Das dafür entwickelte TCP/IP Protokoll (*Transmission Control Protocol/ Internet Protocol*) erfüllt diese Ansprüche auch. Der nächste Schritt mit Multimedia-Übertragungen hat aber andere Besonderheiten, die neue Protokolle und neue Dienste erfordern. Deshalb wurden neue Protokolle entwickelt, um die Internet-Architektur zu erweitern und um Anwendungen mit Audio- und Videodaten und Konferenzen zu ermöglichen. Diese realzeitorientierten Protokolle (RTP, RTCP, RSVP, RTSP) unterstützen die bisherigen Unicast-Strukturen, als auch die neuen Multicast-Netze.

Internetkommunikation stützt sich auf die TCP/IP Protokoll-Familie (Bild 3.3). Diesen Protokolle fehlen also sämtliche Besonderheiten, die für die Übertragung von Breitbandaudio von Interesse wären [Rah97]. Sie verfügen z.B. nicht über die Fähigkeit, Bandbreite zu reservieren. Beispielsweise verursacht ein Mangel an Bandbreitenkontrolle bei einer Breitband-

Audioübertragung mit CD-Musikqualität eine schlechte Audioqualität des Empfangssignals. Dieses Bandbreitenproblem wird bei Einsatz von datenkomprimierenden Verfahren minimiert. Es werden also nur komprimierte (quellencodierte) Audiopakete übertragen. Es müssen dafür effiziente Codierverfahren, welche die Datenmenge stark reduzieren und dabei eine gute Audioqualität beibehalten, eingesetzt werden.

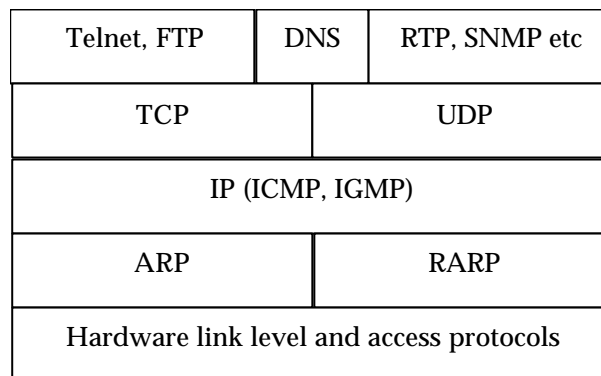


Bild 3.3: TCP/IP Protokoll-Familie [Sta91].

Zu dem Bandbreitenproblem im Internet kommt das Problem der Netzwerkverzögerung. Bei Audioübertragung ist es sehr wichtig, eine konstante oder sich minimal verändernde Verzögerung (*Delay*) zu behalten. Kontrollmechanismen, die solche Verzögerungen schätzen, sind daher notwendig. TCP/IP verfügt nicht über solche Kontroll-Mechanismen, auch nicht die letzte Protokoll-Generation *Real Time Protocol* (RTP).

TCP verwendet eine *Automatic Repeat Request*- (ARQ) Technik mit einem *Positive Acknowledgment*. Diese erhöht die Zuverlässigkeit der Übertragung: Es wird vom Empfänger eine Empfangsbestätigung des Paketes erwartet, andernfalls wird das Paket erneut übertragen. Die mit diesem Verfahren verbundene effektive Übertragungsrate hängt von der Zahl der Wiederholungen und damit von der Netzbelastung ab und kann relativ gering sein.

Der Kern der TCP/IP Protokoll-Familie ist das IP-Protokoll. Die Audiodaten werden in Datagramme (Pakete) unterteilt und zum Empfänger geschickt. IP ist ein verbindungsloses Protokoll; Das bedeutet, dass die Pakete unabhängig voneinander weitergeleitet werden. IP-Pakete können daher sogar in falscher Reihenfolge beim Empfänger eintreffen, wenn es während der Übertragung durch variable Übertragungswege zu Überholvorgänge kommt. Es besteht auch die Möglichkeit, dass die Audiopakete mehrmals oder gar nicht den Empfänger erreichen. Die Weiterleitung der IP-Pakete durch die IP-Router erfolgt anhand der IP-Adresse, die in jedem Paket enthalten ist [Rah97].

Real Time Audio im Internet wird im allgemeinen unter Verwendung der Kombination IP und *User Datagram Protocol* (UDP) realisiert. UDP ist im Gegensatz zu TCP ein verbindungsloses weniger zuverlässiges Protokoll. Seine Hauptaufgabe besteht darin, Pakete so

schnell wie möglich zum Empfänger zu übertragen. Das *User Datagram Protocol* UDP sorgt für eine hohe Übertragungsrate, es hat den Nachteil, dass es schnell das Netzwerk überlastet und die Datenübertragung nicht garantiert.

3.3 Real-Time Audio-Protokolle

Real-Time Audio im Internet ist ohne Zweifel für viele interessant. Die Basisstruktur vom Internet wurde leider nicht für diese Anwendung entworfen. Die Übertragung von zeitkritischen Informationen wie Audio ist schwierig und nicht sehr zuverlässig. So entsteht ein dringender Bedarf nach Verbesserung der Protokoll- und Netzwerk-Technologien. Die Hauptprobleme der Echtzeitübertragung von Audiodaten über das Internet sind die dabei zur Verfügung stehenden Kanalbandbreite, die Bandbreitenreservierung und die Latenzkontrolle (Zeitverzögerung) [Rah97].

Bandbreiten-Quelle können mit Unterstützung fortgeschrittener Netzwerk-Technologien wie ATM zu der zur Verfügung stehenden Bandbreite hinzugefügt werden. Diese löst aber nicht das „*Internetworking*“-Problem, denn das Internet ist ein Netzwerk, das aus unterschiedlichen Netztechnologien besteht und nicht nur aus reinen ATM-Netzen. Diese Lösung ist also nur für lokale Verbindungen wie *Local Area Networks* (LAN) geeignet.

Bei *Streaming* und Echtzeitübertragung reicht ein einziges Übertragungsprotokoll nicht aus. Für den Datentransport von *Streaming*-Audio sind beispielsweise folgende Protokolle zuständig: *Real-Time Transport Protocol*, *Real-Time Streaming Protocol* und *Resource Reservation Protocol*.

3.3.1 Real-Time Transport Protocol

Das *Real-Time Transport Protocol* (RTP) ermöglicht eine Punkt-zu-Punkt-Datenübertragung von Daten mit Echtzeit-Charakteristik wie interaktives Audio oder Video. Die zur Verfügung gestellten Dienste umfassen Typidentifikation, Paketnummerierung, Zeit-Stempel und Zustellungskontrolle. RTP ermöglicht mit diesen Diensten der empfangenden Station, fehlende doppelte oder in falscher Reihenfolge empfangene Pakete zu erkennen und den Empfangsstrom zu korrigieren. Durch RTP kann die Synchronisation zwischen Audio- Video- und Dateninformationen hergestellt werden.

RTP-Pakete werden überwiegend in UDP-Paketen transportiert. RTP garantiert weder die zeitgerechte Zustellung von Paketen noch andere *Quality of Service* (QoS)-Kriterien und verlangt dies auch nicht von dem darunter liegenden Protokoll.

Im RTP-Header stehen die wichtigsten Informationen für die übertragenen Daten:

- *Payload Type Identifier* spezifiziert das *Payload* Format, also das Format der mitgeführten Daten. Dies kann sehr unterschiedlich sein und hängt von der jeweiligen Anwendung ab. Bsp.: MPEG 1/2, JPEG Video.
- Die wichtigste Information für Real-Time-Anwendungen, die RTP unterstützt, ist *Timestamping* (Zeitstempel). Mit Hilfe von *Timestamping* kann der Empfänger das *Original-Timing* des Datenstreams rekonstruieren und Video- und Audiodaten synchronisieren.
- Eine andere Information ist die *source identification*. Dies erlaubt der empfangenden Anwendung Informationen über den Sender beispielsweise den Konferenzteilnehmer zu geben.
- Da UDP die ankommenden Pakete nicht in der richtigen Reihenfolge liefert, werden noch Informationen über die *sequence numbers* mitgeliefert, welche die Pakete wieder ordnen und gleichzeitig verlorengegangene Pakete feststellen können.

In der Praxis ist RTP normalerweise in der Anwendung implementiert und bei einer *Multimedia-Session* wird auch für jedes Medium eine eigene RTP-Session eröffnet. Bei einer Konferenz z.B. werden dann Audio- und Videodaten in getrennten RTP-Paketen untergebracht und übertragen. Anwendungsbeispiele sind der Netscape *LiveMedia* und Microsofts *Netmeeting Conferencing* Software.

3.3.2 Real-Time Streaming Protocol

Das *Real-Time Streaming Protocol* (RTSP) steuert den Datenstrom. Unter der Federführung von *RealNetworks* und *Netscape* ist RTSP 1996 entstanden, das der effizienten Übertragung von Multimedia Streams über IP-Netzwerke dient. Bei den Datenquellen kann es sich um Live-Daten oder Aufzeichnungen handeln. Dabei bietet RTSP *Remote* Funktionalitäten wie *pause*, *play*, *rückwärts an*. RTSP wird auf Anwendungsebene implementiert und setzt auf RTP oder RSVP auf und ergänzt dies um einer Verbindungsüberwachung und Echtzeitübertragung-Kontrolle.

3.3.3 Resource Reservation Protocol

Das *Resource Reservation Protocol* (RSVP) ist ein Netzwerk *Control Protocol*, zuständig für den QoS einer Datenübertragung. RSVP reserviert die erforderlichen Netzwerk-Ressourcen für den Datenstrom, somit können Real-Time-Anwendungen eine Bandbreite auf dem Übertragungsweg anfordern, um eine gewisse Bandbreite garantiert belegen zu können. Gerade bei der Übertragung von Video- oder Audioinformationen haben schon geringe Paketverluste oder Schwankungen der Verzögerung (Jitter) sehr negative Auswirkungen auf

die Qualität. RSVP soll dieses Problem lösen. Es erlaubt dem Empfänger, Ressourcen für den Datenstrom zu reservieren. Die Reservierung wird von Router zu Router bis hin zum Sender vorgenommen. Mehrere Reservierungen von verschiedenen Stellen für einen Datenstrom können von Routern zusammengelegt werden.

3.4 Vorteile der Paketvermittlung

Für eine digitale Audioübertragung über paketvermittelnde Netze sprechen insbesondere folgende Vorteile:

- Die Paketvermittlung bietet mehr Flexibilität als die Leitungsvermittlung, was für den Aufbau verschiedenster Kommunikationsbeziehungen entscheidend ist. Es werden einheitliche Vermittlungseinrichtungen und Leitungen für die Übertragung einer Vielfalt digitaler Signale verwendet.
- Ein wesentlicher Vorteil der Paketvermittlung liegt in der besseren Kanalausnutzung. Es wird nur dann Kanalkapazität beansprucht, wenn Nachrichten zu übertragen sind. In Übertragungspausen kann die frei werdende Kanalkapazität für andere Verbindungen genutzt werden, ohne dass besondere Maßnahmen im Netz erforderlich werden (ATM). Durch die Paketvermittlung ist also ein Gespräch zwischen zwei Personen kostengünstiger zu realisieren, da eine Leitung für mehrere Kommunikationsbeziehungen verwendet werden kann. Während eines Gespräches kommt es zu Pausen, in denen keine Daten übertragen werden müssen. Innerhalb dieser Pausen kann die Leitung für eine andere Kommunikationsbeziehung verwendet werden.
- In Paketvermittlungsnetzen können auf einfache Weise Audio- und Sprachcodierverfahren eingesetzt werden, die mit unterschiedlichen Bitraten arbeiten. Daneben können Verfahren verwendet werden, die mit variabler Bitrate arbeiten, wobei die Steuerung der Bitrate vom Quellencodierer (signalabhängig) oder auch von den Netzknoten (belastungsabhängig) vorgenommen wird [Nol86].

3.5 Probleme der Paketvermittlung

Paketvermittelnde Netzwerke haben Schwächen, die sich gerade auf die Übertragung von Echtzeitdaten wie Audio oder Sprache nachteilig auswirken.

Während einer Übertragung über ATM- oder IP-Netze ist eine Verzögerung oder einen Verlust einiger Pakete manchmal unvermeidbar. Die Verzögerung ist auf die Übertragung und Pufferung zurückzuführen. Die Pufferungszeit hängt von der Linkbelastung ab.

3.5.1 Variable Laufzeiten

Die Übertragung über paketvermittelnde Netzwerke erfolgt blockweise. Jeder Block wird als Paket vom Sender zum Empfänger geschickt. Der Weg des Paketes und die Zeit, die es braucht, um durch ein aus mehreren Netzknoten bestehendes Netzwerk zu gelangen, ist meistens unbestimmt (insbesondere bei IP). Die Gesamtlaufzeit eines Paketes ist stochastischer Natur. Die Netzverzögerung, die ein Paket nach dem Abschicken erfährt, setzt sich zusammen aus [Che88]:

- Feste Übertragungszeit (*Propagation Delay*). Diese ist durch die endliche Ausbreitungsgeschwindigkeit elektromagnetischer Wellen bedingt.
- Variable Verzögerung. Diese wird durch Verarbeitung und Pufferung im Empfänger und in auf dem Übertragungsweg liegenden Gateways und Routern (*Processing and Queueing Delay*) verursacht.

3.5.2 Netzüberlastung

Bei Überlastung eines Pufferspeichers und *Gateways* oder *Routern* im Netzwerk werden Datenpakete unterdrückt (*Discarding/Dropping*). Die Echtzeitanforderungen erlauben keine erneuten Anforderungen von Paketen. Wenn ein Datenpaket später als sein geplanter Ausgabezeitpunkt eintrifft, kann dies dazu führen, dass das Paket nicht mehr verwendet werden kann und als verloren angesehen werden muss.

Audio-Streaming-Systeme benutzen einen kleinen Puffer, um Audiopakete zusammenzufassen und anzuordnen, bevor sie decodiert werden. Pakete, die fehlen, können erneut angefordert werden solange die Pufferungs-Zeit noch nicht überschritten ist. Pakete, die diese Zeit überschreiten und später ankommen, werden als verloren angesehen.

3.5.3 Kanalfehler

Bei der Übertragung über paketvermittelnde Netzwerke kann es bei Kanalfehlern sogar zu einem vollständigen Verlust des Paketes kommen. Durch das Auftreten von nicht korrigierbaren Bitfehlern im Zellkopf kann die Zielinformation verfälscht werden. Das Paket kann somit den richtigen Empfänger nicht erreichen.

Diese oben erwähnten Probleme der paketorientierten Übertragung können dazu führen, dass Pakete verspätet, in falscher Reihenfolge¹ oder überhaupt nicht beim Empfänger ein-

¹ nicht bei ATM

treffen. Beim Empfänger entsteht dann im Audiosignal eine Lücke, welche die empfundene Audioqualität stark beeinträchtigt. Diese fehlende Signalauschnitte können für den Hörer sehr unangenehm sein. Um einen besseren Höreindruck zu erreichen, müssen diese Signallücken mit geeigneten Daten ausgefüllt werden. Diese Aufgabe wird in dieser Arbeit mit Paketsubstitution bezeichnet (s. Kap. 4).

3.6 Verlust-Charakteristik

Aufgrund der zunehmenden Bedeutung der Übertragung von Sprache, Audio- und Video-Daten über das Internet haben sich immer mehr Forschungsarbeiten gegeben, die sich mit der Untersuchung von Paketverlusten sowie von der Statistik der auftretenden Paketverlusten im Internet befassen haben [Bol95] [Kos98] [Har95]. Aus den Untersuchungen wurde festgestellt, dass bei kleinen bis mittleren Paketverlustwahrscheinlichkeiten die Paketverluste einzeln auftreten. Bei höheren Paketverlustwahrscheinlichkeiten (über 10 %), finden häufiger Bündelfehler statt, wobei die Anzahl der Paketverluste pro Bündel relativ klein bleibt (meistens bis maximal 3 Paketverluste pro Bündel).

In einigen Untersuchungen von Bolot und García wurden die auftretenden Paketfehler an einem großen Audio-Datenstrom, welcher über das Internet geschickt wurde, gemessen. Untersucht wurde vor allem die Verbindung zwischen INRIA Sophia Antipolis in Süd Frankreich und der Universität College London (UCL) in Großbritannien. Diese Verbindung galt tagsüber als besonders belastet [Bol95]. Es wurden sowohl *Unicast* (über IP/UDP/RTP) als auch *Multicast*- (über Mbone) Verbindungen untersucht. Die Untersuchungen ergaben, dass Paketverluste meistens einzeln auftreten. Bündelfehler traten seltener und meistens nur bei zwei oder maximal drei aufeinanderfolgenden Paketen auf.

Dieses Ergebnis wurde durch eine Messung von Hardman und Handley [Har95] im Mbone bestätigt. Bei einer Übertragung einer Audio-Datei gingen 2 bis 5 % aller Pakete verloren. Dabei handelte es sich vorwiegend um einzeln auftretende Paketverluste.

4 Substitutionssignale für Paketverluste

Bei der paketorientierten Übertragung von Sprach- oder Audiosignalen können durch die Paketverluste beim Empfänger Lücken im auszugehenden decodierten Signal entstehen. Diese Lücken können sich je nach Paketlänge¹, Paketverlustwahrscheinlichkeit und Anzahl der aufeinanderfolgenden Paketverluste negativ auf den subjektiven Höreindruck auswirken.

Bei einem Paketverlust muss die Signallücke zur Verbesserung der subjektiven Audioqualität durch ein Substitutionssignal gefüllt werden. Ziel kann dabei naturgemäß nicht sein, das verlorene Audiosignal fehlerfrei zu rekonstruieren, sondern lediglich, ein Substitutionssignal zu erzeugen, das sich möglichst unhörbar in das übrige, korrekt übertragene Audiosignal einfügt. Zu diesem Zweck werden unterschiedliche Substitutionsverfahren, die unterschiedliche Komplexität und Qualität aufweisen, angewendet. Das Hauptproblem bei der Paketsubstitution bei Paketverlusten liegt in der ausreichenden Parametrisierung des Audiosignals zur Erzeugung eines künstlichen Signals, das möglichst nicht als Substitutionssignal erkannt wird.

Man unterscheidet senderseitige und empfängerseitige Techniken zur Substitution von Paketverlusten. Audioanwendungen, die auf einer paketorientierten Übertragung basieren, sollten sowohl senderseitige Schutz- als auch empfängerseitige Rekonstruktionsmechanismen anwenden, um eine maximale Übertragungs- und Empfangsqualität zu gewährleisten.

4.1 Senderseitige Verfahren

In diesem Abschnitt werden einige Verfahren vorgestellt, die im Sender auf Audiosignalen angewendet werden, um den negativen Einfluss von Paketverlusten zu mindern oder die Paketverlustwahrscheinlichkeit zu reduzieren. Senderseitige Rekonstruktion bedeutet die Teilnahme des Senders bei der Wiederherstellung verlorener Pakete. Dies kann entweder durch Verständigung mit dem Empfänger z.B. zwecks einer Neuübertragung (ARQ-Verfahren) bei einem Paketverlust, oder unabhängig vom Empfänger und Übertragungszustand erfolgen.

Man unterscheidet hauptsächlich drei Verfahren: *Retransmission* (Neuübertragung), *Interleaving* und *Forward Error Correction* (s. Bild 4.1).

¹ Bei der Übertragung von codierten Daten entspricht die Paketlänge, die Länge des uncodierten PCM-Segments.

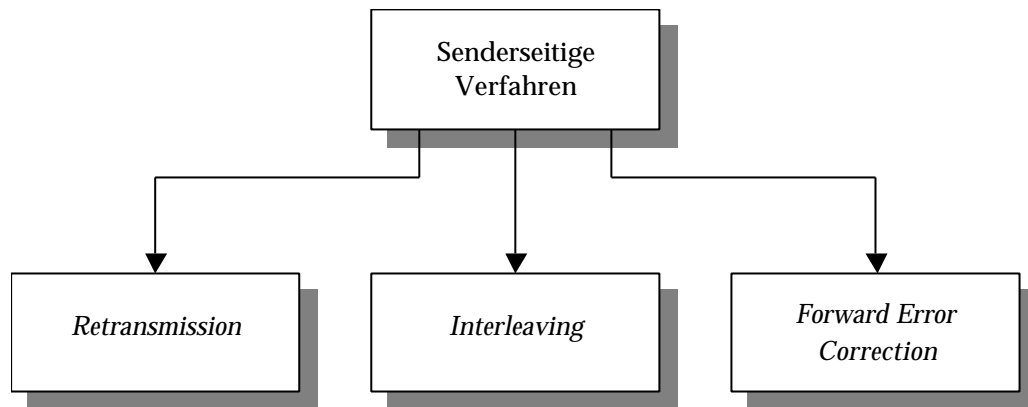


Bild 4.1: Senderseitige Verfahren zur Korrektur bzw. Rekonstruktion von Paketverlusten

4.1.1 Retransmission

Mit diesem Verfahren erfolgt die Wiederherstellung eines Paketverlustes durch Verständigung beider Seiten einer Kommunikation (Sender und Empfänger). Nach dem Senden eines Datenpaketes wartet der Sender auf eine Empfangs-Bestätigung (*Acknowledgement ACK*) durch den Empfänger. Die Bestätigung erfolgt in Form eines Paketes ohne Datenanteil (nur *Header*) mit gesetztem ACK-Bit und *Acknowledgement-Sequenz-Nummer*¹. Erfolgt nach einer gewissen Zeit (**Round Trip Time** RTT) keine Bestätigung des Empfängers, wird das Paket neu verschickt. Der Sender gleicht somit einen möglichen Paketverlust aus. Bei der Datenübertragung wird also im Fall eines Paketverlustes das Paket erneut übertragen. Dies hat den Vorteil, dass am Empfänger die tatsächlichen und nicht die geschätzten Daten vorliegen.

Diese Technik wird vom TCP-Protokoll verwendet und sorgt für die Zuverlässigkeit der Übertragung mit diesem Protokoll. Allerdings bewirkt die *Retransmission* durch die höhere Anzahl der gesendeten Pakete eine zusätzliche Netzbelastung, die zu Übertragungsverzögerung führt.

Im Fall eines Paketverlustes muss also der Empfänger auf die erneut gesendeten Pakete warten. Die Wartezeit ist stochastischer Art und hängt in erster Linie von der Entfernung zwischen Sender und Empfänger und von der Netzbelastung ab. Manche Echtzeit-Anwendungen wie z.B. *online Audio* oder interaktive Anwendungen, haben feste Verzögerungsgrenzen, die nicht überschritten werden dürfen. Beispielsweise liegt die maximal zulässige Wartezeit bei interaktiven Anwendungen (Sprachdialog-Anwendung) bei ungefähr 250 ms [Bra71]. Die maximale Verzögerungsgrenze bei Echtzeit-Anwendungen kann mit der

¹ Nummer der Sequenz, die der Empfänger als nächstes erwartet.

Retransmission schnell überschritten werden. Bei solchen Anwendungen ist die *Retransmission* also zur Behebung von Paketverlusten nicht geeignet.

4.1.2 Interleaving

Im *Interleaving*-Verfahren werden die zu übertragene Pakete vor dem Senden in kleinere Einheiten unterteilt und dann verschachtelt. Übertragene Pakete enthalten Einheiten aus anderen Paketen, so dass im Fall eines Paketverlustes nicht die gesamte Information eines Paketes, sondern nur ein Teil davon verloren geht. Der Verlust eines Paketes wird auf mehrere Pakete verteilt. Anstatt einer großen Lücke im Datenstrom entstehen bei einem Paketverlust nur noch kleine Lücken.

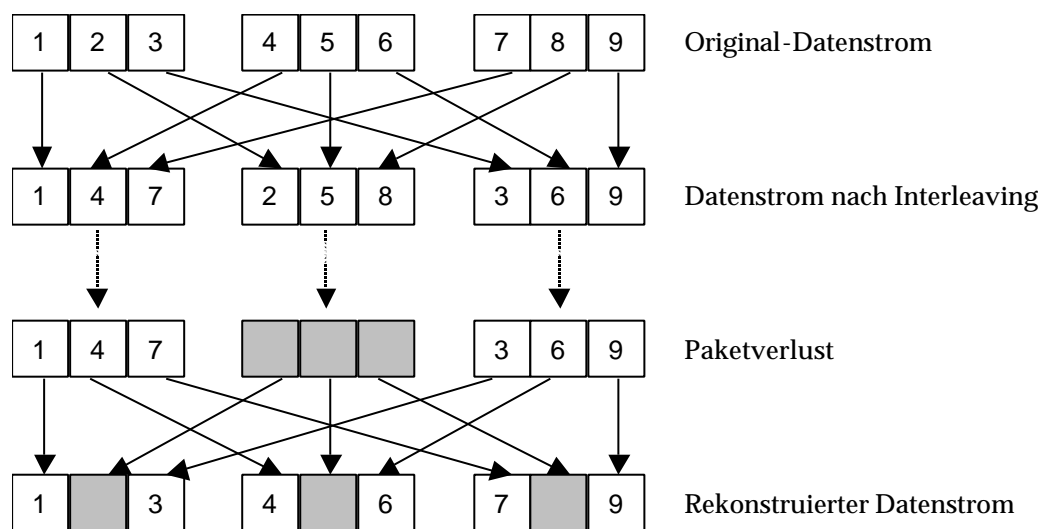


Bild 4.2: Fehlerpreizung mit *Interleaving*

In Bild 4.2 ist das Prinzip von *Interleaving* anhand eines Beispiels dargestellt. Es werden drei Pakete mit jeweils drei Einheiten übertragen. Während der Übertragung geht Paket zwei, das aus den Einheiten 2, 5 und 8 zusammengesetzt ist, verloren. Dieser Verlust wird im rekonstruierten Datenstrom auf die drei Pakete gespreizt. Es entstehen somit kleinere Lücken im rekonstruierten Datenstrom.

Interleaving hat als Nachteil den zusätzlichen Bedarf an Verarbeitungszeit für die empfangenseitige Wiederherstellung des ursprünglichen Datenstroms, was einer Begrenzung seiner Verwendungsmöglichkeiten vor allem für interaktive Applikationen bedeutet. Dieses Verfahren ist gut geeignet für nicht interaktive Anwendungen, bei denen die Echtzeit-Übertragung eine geringe Rolle spielt.

4.1.3 Forward Error Correction

Netzplaner haben bei der Kanalcodierung zwei Grundstrategien für die Fehlerbehandlung entwickelt: Fehlerkorrigierende und fehlererkennende Codes. Verfahren der *Forward Error Correction* (FEC) nutzen fehlerkorrigierende Codes. Sie werden eingesetzt, damit eine gewisse Anzahl von Übertragungsfehlern (Bitfehler) oder Paketverlusten toleriert werden kann.

FEC basiert darauf, dass der Sender von sich aus zusätzliche Pakete als Prüfbitinformation an den Empfänger übermittelt. Der Empfänger kann aus den Redundanzinformationen zusammen mit den korrekt empfangenen Paketen einzelne und teilweise auch mehrere nicht korrekt empfangene Pakete rekonstruieren. Der Empfänger kann somit den ursprünglichen Datenblock ohne Neuzusendung (*Retransmission*) wieder rekonstruieren. Mit FEC werden also im vornherein zusätzliche Datenpakete versendet, die den Verlust von Paketen ausgleichen können. Dabei werden mehrere Pakete in Blöcken zusammengefasst. Für jeden Block werden dann zusätzliche Pakete erzeugt und alle Pakete übertragen.

FEC bietet eine Alternative zu rückwärtskorrigierenden Verfahren wie ARQ und sorgt für eine höhere Zuverlässigkeit der Übertragung ohne die Verzögerung zu erhöhen. Dies ist insbesondere für Echtzeit-Applikationen wichtig. FEC ist daher besonders passend für z.B. *Live Audio* Applikationen über das Internet. Die Leistungsfähigkeit von FEC bei der Rekonstruktion von Paketverlusten hängt stark von der Paketverlustcharakteristik im Netz ab. FEC sind mehr effektiv bei einzelnen Paketverlusten die weit übereinander liegen (niedrige Paketverlustwahrscheinlichkeit).

Die Nachteile der Vorwärtsfehlerkorrektur mit FEC liegen vor allem im Codieraufwand (senderseitigem Hinzufügen von Prüfdaten, welche die empfängerseitige Korrektur ermöglichen) aber auch beim erhöhten Bandbreitenbedarf. Außerdem kann es durch das Übertragen zusätzlicher Pakete dazu kommen, dass die erhöhte Netzwerklast wiederum zu einer höheren Paketverlustwahrscheinlichkeit führt.

FEC unterteilt sich in zwei Kategorien: quellenabhängige und quellenunabhängige FEC [Per98]. Mit Quellenabhängigkeit ist die Abhängigkeit von dem Inhalt des Quellensignals (z.B. Audio oder Video) gemeint.

- **Quellenunabhängige FEC:** Es werden Prüfdaten (z.B. Paritätsdaten) zu den originalen Daten eines Paketes hinzugefügt. Bei einem Paketverlust kann der Empfänger durch die zusätzlichen Daten den Verlust erkennen und die verlorenen Daten rekonstruieren. Dieses Verfahren kann auf alle zu übertragenden Daten angewandt werden und hat, aufgrund der Einfachheit der verwendeten Algorithmen, geringe Systemanforderungen.

- **Quellenabhängige (wahrnehmungsangepaßte) FEC:** Dieses Verfahren funktioniert grundsätzlich wie das quellenunabhängige FEC, allerdings wird hier das Wissen über den Aufbau der zu sendenden Daten zur Erstellung der redundanten Daten genutzt. Die Prüfdaten werden den Eigenschaften des Quellensignals angepasst. Es werden z.B. nur die wichtigsten Bits übertragen.

4.2 Empfängerseitige Verfahren

Die Paketsubstitution bei Paketverlusten, die während einer Übertragung von Audio-Daten über paketvermittelnde Netze auftreten können, kann auf der Empfängerseite erfolgen. Hier unterscheidet man decodiererinterne und decodiererexterne Verfahren:

- **Decodiererinterne** Substitutionsverfahren erfordern eine Modifikation des Decodierers selbst (Modul 3 aus Bild 1.2). Die Substitution erfolgt hierbei auf der Ebene codierter Audiosignale. Der Decodierer muss also versuchen, anhand der beispielsweise vor dem Paketverlust vorhandenen Parameter der zuvor korrekt empfangenen Pakete, neue Parameter zu schätzen, und diese zur Decodierung, also Rekonstruktion des verlorenen Paketes, heranzuziehen.

Das größte Problem bei decodiererinternen Paketsubstitutionsverfahren entsteht, wenn die verwendete Audiocodierung auf einer Rückwärtsadaption von Parametern basiert. In diesem Fall ist die Decodierung der dem Paketverlust nachfolgenden korrekt empfangenen Parameter von den geschätzten Parametern des verlorenen Paketes abhängig. Die Decodierung korrekt empfangener Parameter wird somit von der Güte des verwendeten Substitutionsverfahrens abhängig. Dies kann bei einer schlechten Rekonstruktion eine Fortpflanzung der Fehler bedeuten.

In den meisten Audio- und Sprachdecodierern sind Fehlerverschleierungsmodule (*Concealment*) integriert, die bei einem Paketverlust versuchen, die codierten Parameter des verlorenen Paketes z.B. durch Wiederholung oder Interpolation der vor dem Verlust vorhandenen Parameter zu schätzen. Beispiele hierfür sind: *Concealment* in dem Sprachdecodierer ITU-T G.729 und in dem Audiodecodierer MPEG2-AAC (*Advanced Audio Coding*). Ein besseres Verständnis des decodiererinternen *Concealment*-Verfahren erfordert Kenntnis der entsprechenden Codierverfahren und der Zusammensetzung des codierten Datenstroms. Dieses würden den Rahmen dieser Arbeit übersteigen. Wesentliche Strategien werden aber im Folgenden angesprochen.

Concealment im ITU-T G.729-Sprachdecodierer

Das Codierverfahren des ITU-T G.729 Sprachcodierers beruht auf einer Modellierung der menschlichen Spracherzeugung. In diesem Modell wird der Vokaltrakt durch ein lineares Filter (Synthesefilter) dargestellt. Sprachsignale werden produziert, indem dieses Filter durch einen Anregungsvektor angesteuert wird. Beim G. 729 Codierer dauert ein Sprachsegment 10 ms korrespondierend zu 80 PCM Abtastwerten. Für jedes Segment analysiert der G.729 Codierer die Eingabedaten und extrahiert die CELP (*Code **E**xited **L**inear **P**rediction*) Modellparameter wie Filterkoeffizienten der linearen Prädiktion und Anregungsvektoren.

Bei Codierverfahren, die auf einer Datenreduktion mittels linearer Prädiktion (z.B. G.729) oder Transformationscodierung basieren, ist es dem Decodierer möglich, bei Paketverlust die Decodierung mit wiederholten oder geschätzten Parametern durchzuführen.

Bei einem Paketverlust fehlen im Decodierer die für die Decodierung des Signals notwendigen Parameter. Der Decodierer versucht, mit seinem internen *Concealment*-Modul anhand der Parameter des letzten Paketes die Parameter des verlorenen Paketes wie folgt zu schätzen:

- Die LSP-Koeffizienten aus dem letzten Paket werden wiederholt.
- Der adaptive und feste Codebuch-Gewinn werden aus dem letzten Paket wiederholt.
- Die Anregung hängt von der stimmhaft/stimmlos-Klassifikation des im letzten Paket vorhandenen Sprachsignals ab. Wenn das letzte Paket als stimmhaft klassifiziert wurde, wird der Beitrag des festen Codebuchs auf Null gesetzt und die Pitch-Verzögerung entspricht der des letzten Paketes. Ist das letzte Paket stimmlos, so wird der Beitrag des adaptiven Codebuchs auf Null gesetzt und das feste Codebuch wird zufällig erzeugt [Sann00] [Ros01].

Concealment im MPEG2-AAC-Decodierer

Für die Synthese verlorener oder fehlerhafter Segmente eines Audiosignals werden im MPEG2-AAC-Decodierer *Error Concealment* Techniken verwendet. Diese erfordern eine empfängerseitige Erkennung der Fehler.

Der MPEG2-AAC Decodierer ist durch seine *Error Resilience Tools* VCB11 (*Virtual Codebook*) , HCR (*Huffman Codeword Reordering*) und RVLC (*Reversibles Variable Length Codes*) in der Lage, Fehler in den codierten spektralen Daten zu erkennen und Informationen für die genaue Lokalisierung dieser Fehler im Datenstrom zu liefern [Lau01] [Spe00]. Um eine Fehlerausbreitung im decodierten Audiosignal zu vermeiden, werden die als fehlerhaft identifizierten spektralen Daten verworfen und mit *Concealment*-Techniken geschätzt. Dazu muss zuerst eine Klassifikation des Audiosignals in „tonal“

oder „rauschähnlich“ erfolgen. Die Klassifikation erfolgt mit einer Kreuzkorrelationsuntersuchung aufeinanderfolgender Abtastwerte [Lau01].

Rauschähnliche Komponenten¹ werden mittels Rauschsubstitution² rekonstruiert. Dazu wird die Energie an die unmittelbar vor dem Fehler vorhandenen Signaleigenschaften angepasst.

Tonale Komponenten werden mit linearer Prädiktion rekonstruiert [Lau01] [Her95]. Es wird eine LMS-Prädiktion (*Least Mean Square*) der Ordnung 32 verwendet. Aufgrund der mit der hohen Frequenzauflösung der MDCT (*Modified Discrete Cosine Transform*) verbundenen schlechten Auflösung im Zeitbereich kann die Prädiktion nicht direkt auf das MDCT-Spektrum angewendet werden. Das MDCT-Spektrum wird in 32 äquidistante Teilbänder zerlegt und im Zeitbereich mit der IMDCT (inverse MDCT) zurücktransformiert. Hier wird die Prädiktion auf die Zeitsignale der unteren Teilbänder des Spektrums angewendet. Zum Schluss erfolgt eine MDCT-Transformation der Zeitsignale in den Frequenzbereich und anschließend eine Decodierung des neuen Spektrums.

In Bild 4.3 ist das Blockschaltbild des MPEG2-AAC-Concealment-Verfahrens dargestellt.

¹ In MPEG2-AAC sind diese Komponenten Spektrallinien mit 1024 Koeffizienten.

² Es werden mit einem Zufallsgenerator Random-Werte erzeugt.

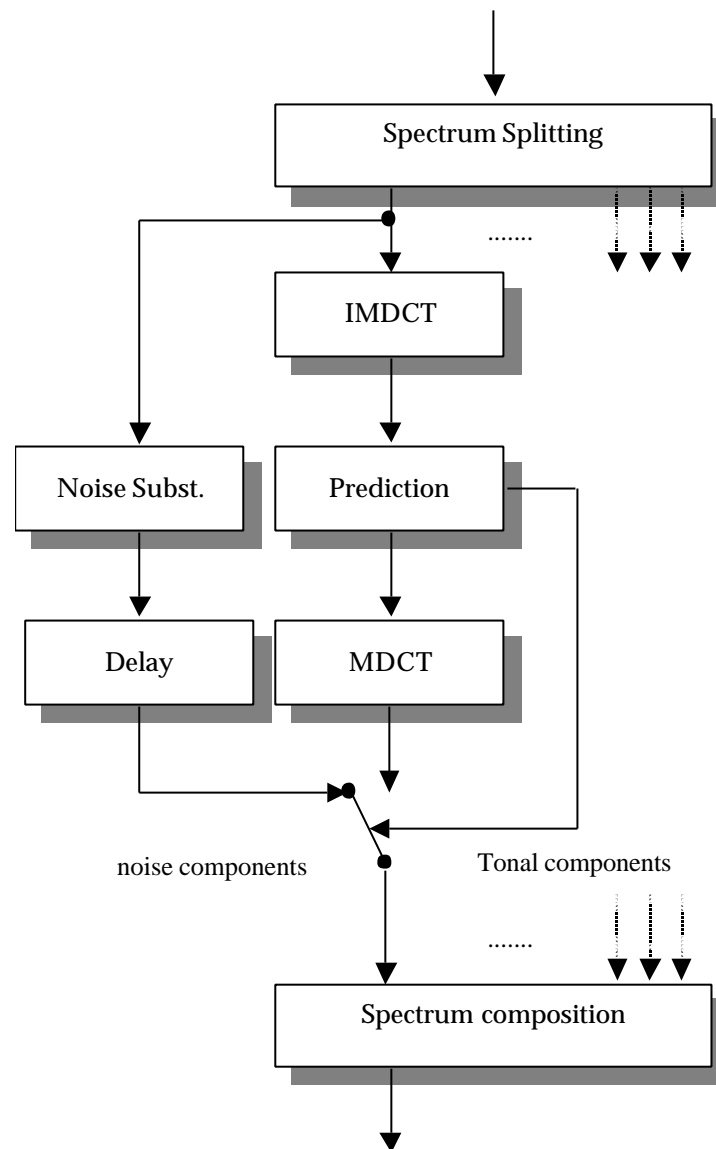


Bild 4.3: *Error Concealment* im MPEG2-AAC-Decoder [Lau01].

- **Decodiererexterne** Verfahren zur Paketsubstitution bei Paketverlusten werden unmittelbar vor (Modul (1) in Bild 1.2) oder nach (Modul (2)) dem Quellendecodierer eingesetzt. Im ersten Fall wird bei einem Paketverlust ein kompletter Satz von Parametern und Koeffizienten, die ein komprimiertes Audiosignal beschreiben, rekonstruiert, indem beispielsweise die Parameter aus dem letzten korrekt empfangenen Paket wiederholt werden. Im zweiten Fall erfolgt die Paketsubstitution nach dem Quellendecodierer. Es werden in diesem Fall unabhängig vom Codier/Decodier-Verfahren PCM-Audiosignale (PCM *Pulse Code Modulation*) rekonstruiert.

Bei der PCM-Codierung sind beide Ansätze praktisch identisch, da jedem Codewort ein Abtastwert zugeordnet wird. Bei einer gedächtnisbehafteten Quellencodierung des Audiosignals sind die Darstellungen sehr unterschiedlich und vom jeweiligen Codierverfahren abhängig. In diesem Fall ist es zumeist sinnvoll, die Rekonstruktion innerhalb des Audio-

Decodierers durchzuführen, um die Codierparameter zur Erzeugung des Substitutionssignals heranziehen zu können, und somit Decodierung und Rekonstruktion zu einem Modul zu verschmelzen.

In dieser Arbeit werden ausschließlich decodiererexterne Verfahren untersucht. Die Substitution erfolgt unmittelbar nach dem Decodierer, entweder direkt auf Signalebene (Kap. 4.3) oder auf Parameter-Ebene (Kap. 4.4, 4.5 und 4.6). Sie ist vom Codierverfahren unabhängig.

4.3 Substitution auf Signalebene

In diesem Abschnitt werden einige Substitutionsverfahren vorgestellt, die bei PCM-Formaten angewendet werden. Auch wenn Audiosignale in komprimierter Form übertragen werden, z. B. im MP3-Format, sind diese Verfahren anwendbar, da bei der Decodierung komprimierter Daten wieder eine Rückumsetzung in PCM-Formate stattfindet (s. Bild 4.4). Bei rückwärtsgesteuerten Codierern müssen allerdings zusätzliche Maßnahmen ergriffen werden.

In Bild 4.4 ist eine Audio-Übertragungsstrecke, die einen zusätzlichen Modul zur Signalrekonstruktion bei Paketverlusten enthält, dargestellt.

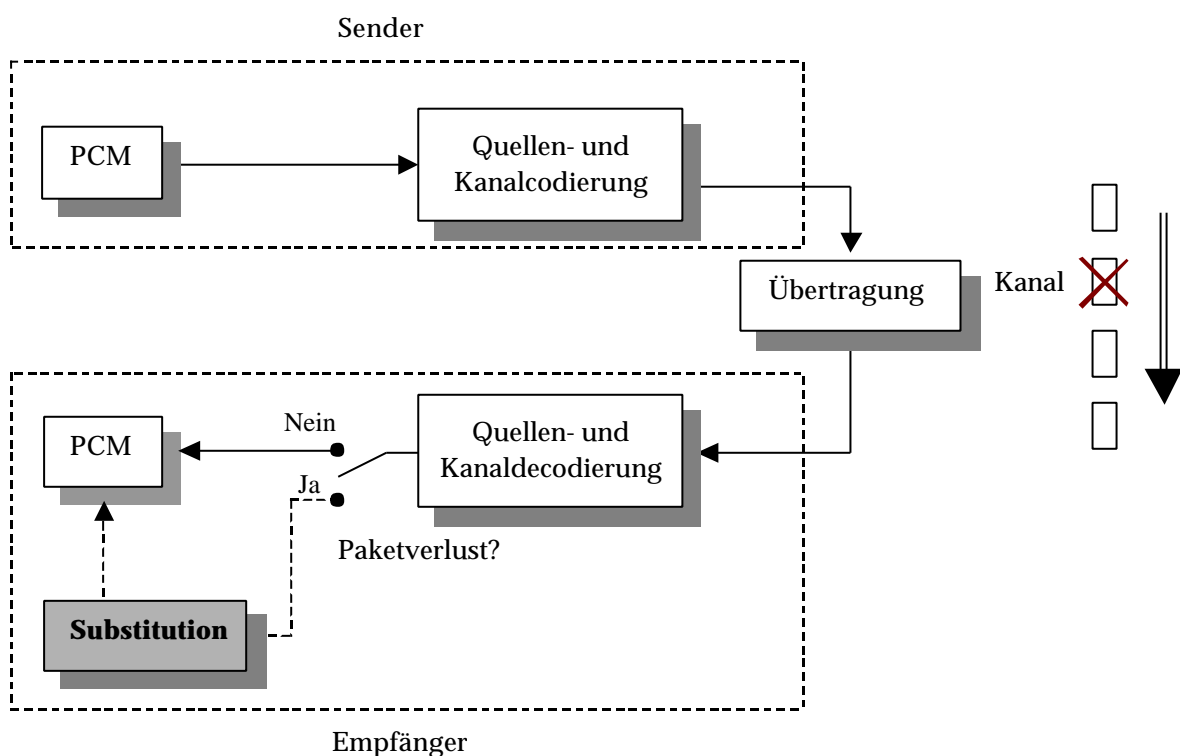


Bild 4.4: Empfängerseitige Paketsubstitution bei Paketverlusten auf PCM-Ebene.

Das Audiosignal liegt beim Sender noch im PCM-Format vor. Vor der Übertragung wird das Signal einer Quellen- und einer Kanalcodierung unterzogen. Die Quellencodierung be-

freit das Audiosignal von Redundanz und Irrelevanz, was eine Kompression des Audiosignals zur Folge hat. Ein Beispiel dazu ist die MPEG-Codierung (z.B. MP3). Die Kanalcodierung fügt im Gegensatz dazu Redundanz zur Bitfolge hinzu, so dass empfängerseitig eine Fehlererkennung und -korrektur möglich wird. Im Fall einer paketorientierten Übertragung wird im Kanalcodierer eine Sequenznummer erzeugt und ein Zellenkopf an die Nutzinformation angefügt. Die Bitfolge wird somit zur Anpassung an den Übertragungsweg in Paketen assembliert.

Die Pakete werden im Netz unabhängig voneinander übertragen. Während der Übertragung kann es zu Verlust von einem (Einzelfehler) oder von mehreren aufeinander folgenden Paketen (Bündelfehler) kommen. Diese Paketverluste werden empfängerseitig erkannt. An dieser Stelle erhält das Substitutionsmodul die Aufgabe, geeignete Substitutionssignale zu finden, um die aufgrund der Paketverluste entstandenen Lücken im PCM-Audiosignal zu füllen. Diese Lücken können beispielsweise mit einem der folgenden Verfahren gefüllt werden.

4.3.1 Nullsubstitution

Der einfachste Ansatz zur Substitution eines verlorenen Paketes ist die Nullsubstitution. Hier werden im Empfänger statt des verlorenen Paketes Nullen ausgegeben (s. Bild 4.5). Bei diesem Verfahren sind keine Schätzungen der Signaleigenschaften erforderlich.

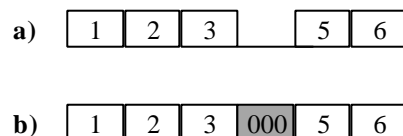


Bild 4.5: Nullsubstitution fehlender Pakete eines Signals

- a) Signal mit Paketverlust (Paket 4 fehlt)
- b) Signal nach der Substitution

Der starke Kontrast zwischen dem Audiosignal und dem Nullsignal wirkt sich negativ auf dem Höreindruck aus. Die Nullsubstitution liefert daher bei Audiosignalen keine akzeptablen Ergebnisse. Sie wird in dieser Arbeit als Referenz für die Beurteilung der untersuchten Substitutionsverfahren verwendet.

4.3.2 Paketwiederholung

Eine weitere Möglichkeit der Substitution fehlender Signalpakete besteht darin, ein fehlendes Signalstück durch ein dem Signal entnommenes vorhergehendes Segment zu ersetzen. Die einfachste Möglichkeit dabei ist die Wiederholung des letzten korrekt empfangenen Paketes. Signalverläufe und die dazu gehörenden Spektrogramme für die Paketlängen 1024 Abtastwerte (22 ms) sind in Bild 4.6 dargestellt. Hier handelt es sich um das Signal eines Einzelinstruments, das in dem dargestellten Abschnitt einen näherungsweise periodischen Verlauf aufweist. Das fehlende Paket aus Bild 4.6-b) wurde durch das vorangegangene ersetzt, was zu einem Phasenfehler führt. Phasenfehler verursachen Knackgeräusche. Eine Verbesserung dieser Substitutionstechnik wird mit der sogenannten "überlappenden Paketwiederholung" erreicht. Dabei werden die Übergänge zwischen dem Substitutionspaket und den angrenzenden Paketen noch geglättet, um die durch Phasensprünge entstehenden Knackgeräusche zu mindern (Kapitel 4.3.7).

Die im Spektrogramm des Signals mit Lücke (Bild 4.6-b) an den Rändern der Lücke auftretenden hohen Frequenzen sind auf die Überlappung der Segmente bei der Berechnung des Spektrogramms zurückzuführen. Bei der Ermittlung der Frequenzen unmittelbar vor und nach der Lücke werden, je nach verwendeter Segmentlänge, einige Abtastwerte aus der Lücke herangezogen. Der plötzliche Bruch in dem Signalpegel (scharfe Kante bei einem Paketverlust) beeinflusst hierbei die berechneten Frequenzen.

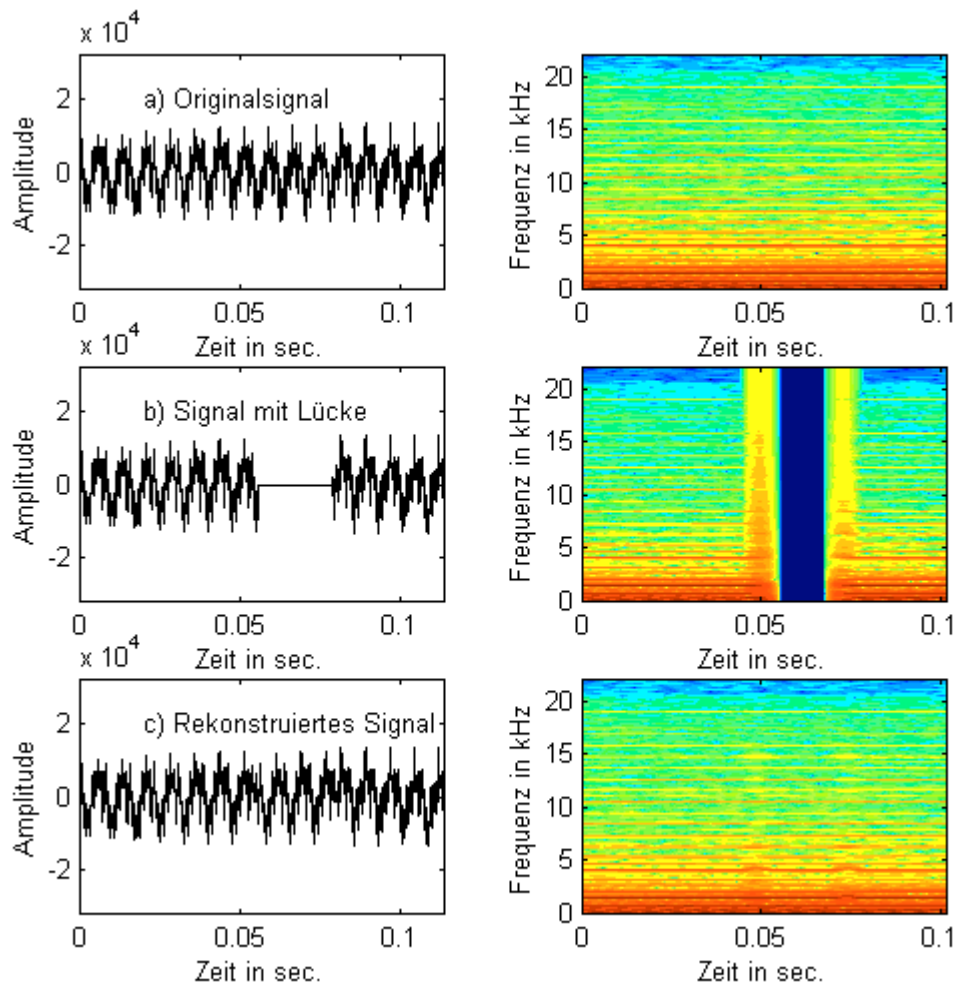


Bild 4.6: Paketsubstitution mit Paketwiederholung; a) Originalsignal (5000 Abtastwerte \cong 113 ms bei 44,1 kHz Audio) aus einem Einzelinstrument-Titel (Orgel); b) Signal mit Paketverlust. Die Paketlänge beträgt hierbei 1024 Abtastwerte (\cong 22 ms); c) rekonstruiertes Signal: Bei Paketverlust wird das letzte Paket wiederholt.

Die Paketwiederholung wird beim GSM-Mobilfunk (*Global System Mobile*) zur Signalrekonstruktion bei Paketverlusten verwendet. GSM-Pakete sind 20 ms lang. Wenn der Empfänger ein fehlerhaftes Paket erhält, versucht er das korrespondierende Signal zu rekonstruieren, indem er das letzte korrekt empfangene Paket wiederholt. Werden mehrere fehlerbehaftete Pakete empfangen, so wiederholt der Empfänger diesen Vorgang unter Absenkung der Sprachamplitude. Der Empfänger wiederholt das selbe Paket für Lücken oder Fehler bis 320 ms. Überschreitet eine Lücke diese Grenze, so wird das Signal oberhalb 320 ms auf Null gesetzt (Nullsubstitution) [Per98].

4.3.3 Substitution mit Rauschen

Zur Verdeckung von Paketverlusten ist es auch möglich, die Lücken im Signal mit Rauschwerten zu füllen. Um einen Bruch im durchschnittlichen Pegel des Signals zu vermeiden, ist eine Leistungsanpassung zum bisherigen Signal vorzunehmen. Das führt jedoch dazu, dass das Rauschen in lauten Passagen besonders stark den Höreindruck stört.

Im Signal mit der Rauschsubstitution sind sehr hohe zischende Hintergrundgeräusche zu hören, so dass der Gesamthöreindruck noch schlechter ist als bei der Nullsubstitution. Die Rauschsubstitution ist daher für die Substitution von Paketverluste bei Audiosignale ungeeignet.

4.3.4 Odd-Even-Sample Interpolation

In [Jay81] wurde ein Verfahren zur Rekonstruktion von Paketverlusten in Sprachsignalen vorgeschlagen. Dieses Verfahren wurde auf PCM- und DPCM- (Differenz-PCM) codierte Sprachsignale angewendet.

Senderseitig werden codierte Pakete (für PCM sind das Abtastwerte) der Länge $2L$ in zwei Pakete der Länge L unterteilt und unabhängig voneinander gesendet. Im ersten Paket sind die ungeraden und im zweiten die geraden Abtastwerte enthalten (s. Bild 4.7).

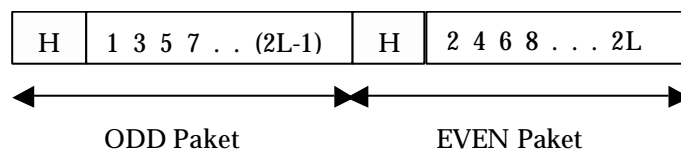


Bild 4.7: Aufteilung eines Paketes in *Odd*- und *Even*-Pakete [Jay81].

Zu jedem Paket der Länge L wird ein zusätzlicher *Header*, der Hilfsinformationen zur späteren Rekonstruktion des ursprünglichen Paketes enthält, hinzugefügt. Es werden also zwei Pakete der Länge $L+h$ unabhängig voneinander gesendet, wobei h die *Header*-Länge bezeichnet.

Wenn eines der beiden Pakete verloren geht, werden die *Odd*-Abtastwerte des ursprünglichen Paketes (bzw. *Even*) aus den *Even*-Abtastwerten (bzw. *Odd*) wie folgt geschätzt:

$$x(n) = \alpha \cdot x(n-1) + \beta \cdot x(n+1) \quad (4.1)$$

wobei

$$\alpha = \beta = \frac{R_{xx}(1)}{1 + R_{xx}(2)} \quad (4.2)$$

$x(n)$: Abtastwerte des Audiosignals.

R_{xx} : Autokorrelation.

α, β : Hilfsparameter zur Rekonstruktion des Audiosignals.

α und β werden senderseitig für jedes 2L-Paket adaptiv berechnet und als Seiteninformation in dem *Odd-* und *Even-Header* gesendet. Sollten beide Pakete nicht zur Verfügung stehen, so wird ein Nullsignal ausgegeben.

Das Paket-Interpolationsproblem lässt sich mit diesem Verfahren in ein besser zu lösendes Abtastwerte-Interpolationsproblem vereinfachen. Mit diesem Rekonstruktionsverfahren wurden bei Sprachübertragung mit Paketlängen von 16-32 ms bei Paketverlusten bis 10 % gute Ergebnisse erreicht [Jay81]. Der Nachteil dieses Verfahrens liegt in der zusätzlichen Verzögerung bei der Rekonstruktion des Audiosignals.

4.3.5 Links/Rechts-Ersatz

Mit diesem Substitutionsverfahren werden bei Stereoübertragung die Stereo-Eigenschaften eines Audiosignals ausgenutzt. In einigen Stereosignalen sind die Kanäle „links“ und „rechts“ fast gleich oder sehr ähnlich. Sprache und Schlagzeug sind z.B. praktisch immer auf beiden Kanälen gleich [Lei99].

Ausgehend von der Korrelation zwischen beiden Kanälen, werden bei Paketverlusten in einem Kanal die fehlenden Abtastwerte durch zeitgleiche Abtastwerte des Nachbarkanals ersetzt. Dieses Verfahren kann bei Stereosignalen mit großer Ähnlichkeit zwischen dem linken und dem rechten Kanal sehr gute Ergebnisse liefern. Ein Beispiel dafür liefern die Signale aus Bild 4.8. Hier handelt es sich um ein 5000 Abtastwerte (113 ms) Audiosignalausschnitt eines Pop-Musik-Titels. Beide Kanäle des Stereosignals sind in Bild 4.8-a) (Kanal links) und Bild 4.8-b) (Kanal rechts) dargestellt. Bild 4.8-c) stellt das Audiosignal des linken Kanals mit einem Paketverlust. Die Länge der Lücke beträgt hierbei 1024 Abtastwerten (22,22 ms). Das rekonstruierte Signal für den linken Kanal ist in Bild 4.8-d) dargestellt und zeigt einen ähnlichen Verlauf und Spektrum wie das Originalsignal aus Bild 4.8-a).

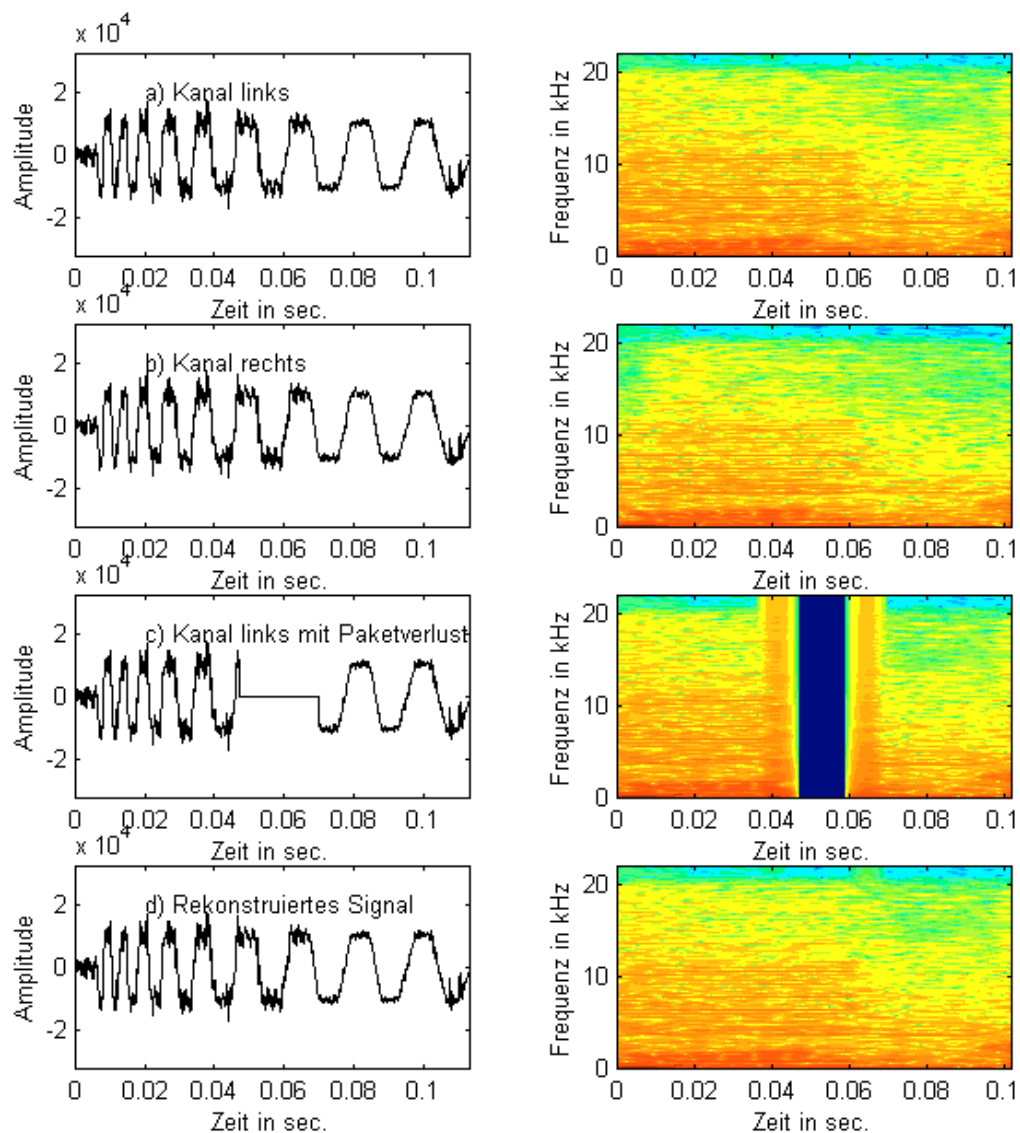


Bild 4.8: Paketsubstitution mit Links/Rechts-Ersatz. Originalsignal (5000 Abtastwerte \cong 113 ms bei 44,1 kHz Audio) aus einem Pop-Musik-Titel mit hoher Ähnlichkeit zwischen beiden Kanälen; a) Linker Kanal des Stereosignals; b) Rechter Kanal des Stereosignals; c) Linker Kanal mit Paketverlust. Die Paketlänge beträgt hierbei 1024 Abtastwerte (22 ms); d) Rekonstruiertes Signal des linken Kanals: Bei Paketverlust wird das zeitgleiche Paket aus dem Nachbarkanal verwendet.

Bild 4.9 zeigt ein Beispiel eines Stereosignals, mit unterschiedlichem Verlauf der Signale aus beiden Kanälen (Kanal links in Bild 4.9–a) und Kanal rechts in Bild 4.9–b)). Bei solchen Signalen eignet sich der Links/Rechts-Ersatz der Pakete nicht für die Signalsubstitution bei Paketverlusten. Dies zeigt das rekonstruierte Signal für den linken Kanal (Bild 4.9–d), das an der Fehlerstellen einen unterschiedlichen Verlauf als das Originalsignal aus Bild 4.9–a) hat. Im spektralen Bereich treten durch das ungeeignete Substitutionssignal für die Lücke ganz andere Frequenzen als die des Originalsignals auf.

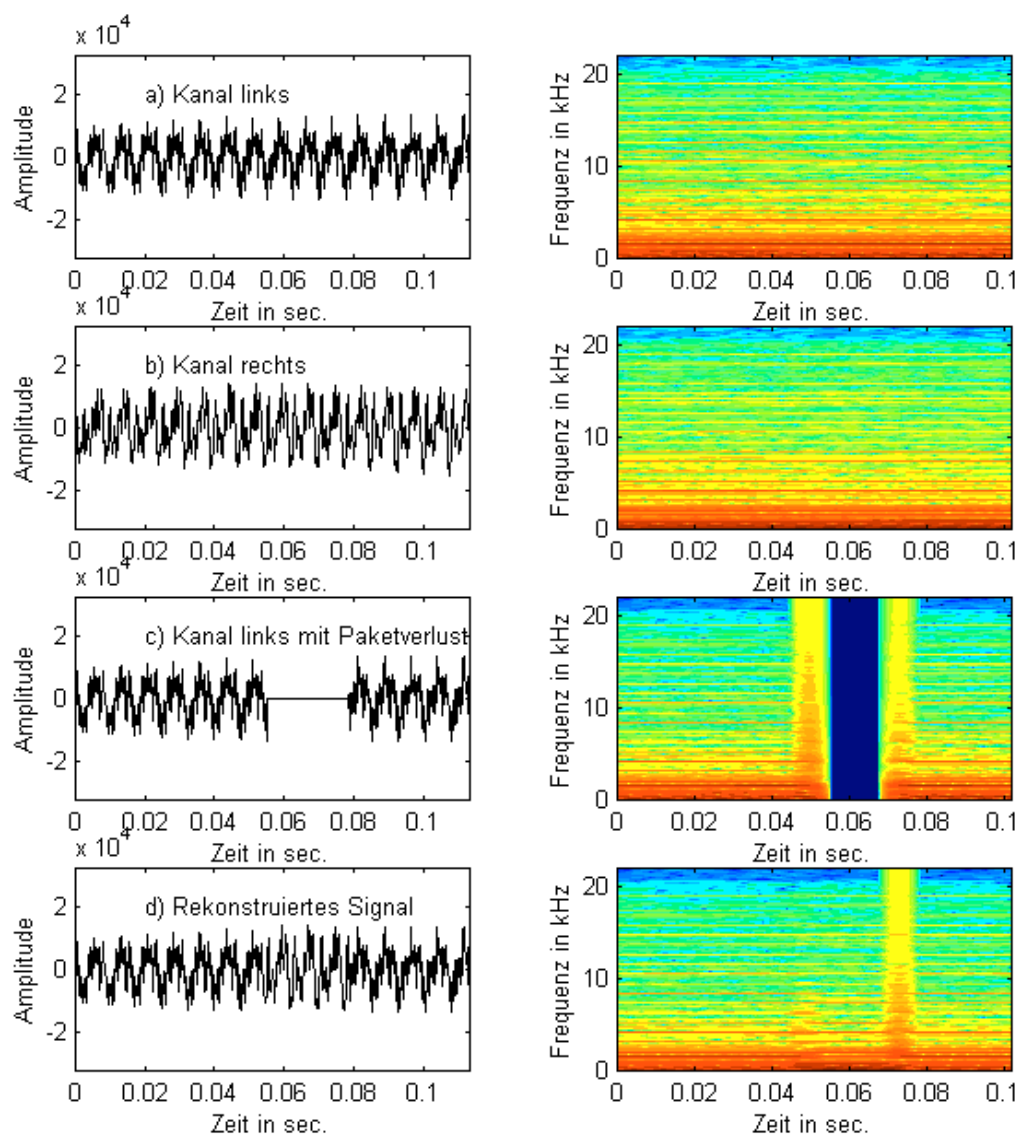


Bild 4.9: Paketsubstitution mit Links/Rechts-Ersatz. Originalsignal aus einem Einzelinstrument-Titel (Orgel) (5000 Abtastwerte \approx 113 ms bei 44,1 kHz Audio). Die Signalverläufe beider Kanäle sind nicht ähnlich; a) Linker Kanal des Stereosignals; b) Rechter Kanal des Stereosignals; c) Linker Kanal mit Paketverlust. Die Paketlänge beträgt hierbei 1024 Abtastwerte (22 ms); d) Rekonstruiertes Signal des linken Kanals: Bei Paketverlust wird das zeitgleiche Paket aus dem Nachbarkanal verwendet.

4.3.6 Verfahren der Maximalen Kreuzkorrelation

In diesem Abschnitt wird ein neues Verfahren zur Rekonstruktion von PCM-Audiosignalen bei Paketverlusten vorgestellt. Das Verfahren der maximalen Korrelation (hier MK-

Verfahren genannt) ist ein Verfahren, das die maximale Ähnlichkeit mit den verlorenen Abtastwerten der Lücke schätzen soll. Die Untersuchung beschränkt sich hierbei nicht auf Monosignale, sondern wird auch mit Stereosignalen durchgeführt, denn oftmals weist ein Audiosignal eines Kanals keine wesentliche Korrelationen mit seiner Vergangenheit, sondern eher mit dem Audiosignal des Nachbarkanals. Diese Ähnlichkeit beider Kanäle eines Stereosignals soll soweit wie möglich auch für die Rekonstruktion der Lücke in einem der Kanäle ausgenutzt werden.

Das MK-Verfahren untersucht bei einem Paketverlust die Kreuzkorrelation zwischen einem *Template*¹, welches die letzten t korrekt empfangenen PCM-Abtastwerte unmittelbar vor dem Paketverlust enthält, und einem Puffer², bestehend aus den letzten l empfangenen Abtastwerten. Bei Stereosignalen stehen zwei Puffer zur Verfügung (*Kanal1Puffer* und *Kanal2Puffer* aus Bild 4.10), welche Daten aus beiden Kanälen des Stereo-Systems enthalten. Es ergeben sich in diesem Fall zwei Kreuzkorrelationssignale KKF1 und KKF2.

Für die Kreuzkorrelationsuntersuchung sollte nicht die gesamte Pufferlänge l herangezogen werden, da es in diesem Fall nicht auszuschließen ist, dass die zur Verfügung stehenden Abtastwerte nicht für die Verdeckung der gesamten Lücke ausreichen würden. Die Kreuzkorrelationsfunktion wird nur für die ersten l Abtastwerte des Puffers ermittelt, wobei

$$l_2 = l_1 - (L + 2 \cdot \Delta L) - t \quad (4.3)$$

L: Paketlänge

ΔL : Länge des *Merging*-Intervalls (s. Kap. 4.3.7)

Nach der Ermittlung der Kreuzkorrelation des Templates mit den beiden gepufferten Audiosignalen, werden die entstandenen Kreuzkorrelationssignale auf Maxima untersucht, um die optimale Position im Puffer, welche die maximale Ähnlichkeit des Audiosignals an dieser Stelle mit dem *Template* beschreibt, zu bestimmen. Das Substitutionssignal der Länge $L+2\Delta L$ ergibt sich dann aus den PCM-Abtastwerten, welche dieser Position um t Abtastwerte folgen.

¹ Die Länge t des Templates kann variiert werden. Optimale Ergebnisse wurden bei 44,1 kHz Audio mit $t = 50 - 100$ AW (ungefähr 1 bis 2 ms) erreicht.

² Optimaler Wert für l ist $3(L + 2\Delta L)$; L entspricht der Paketlänge und ΔL der Länge des *Merging*-Intervalls.

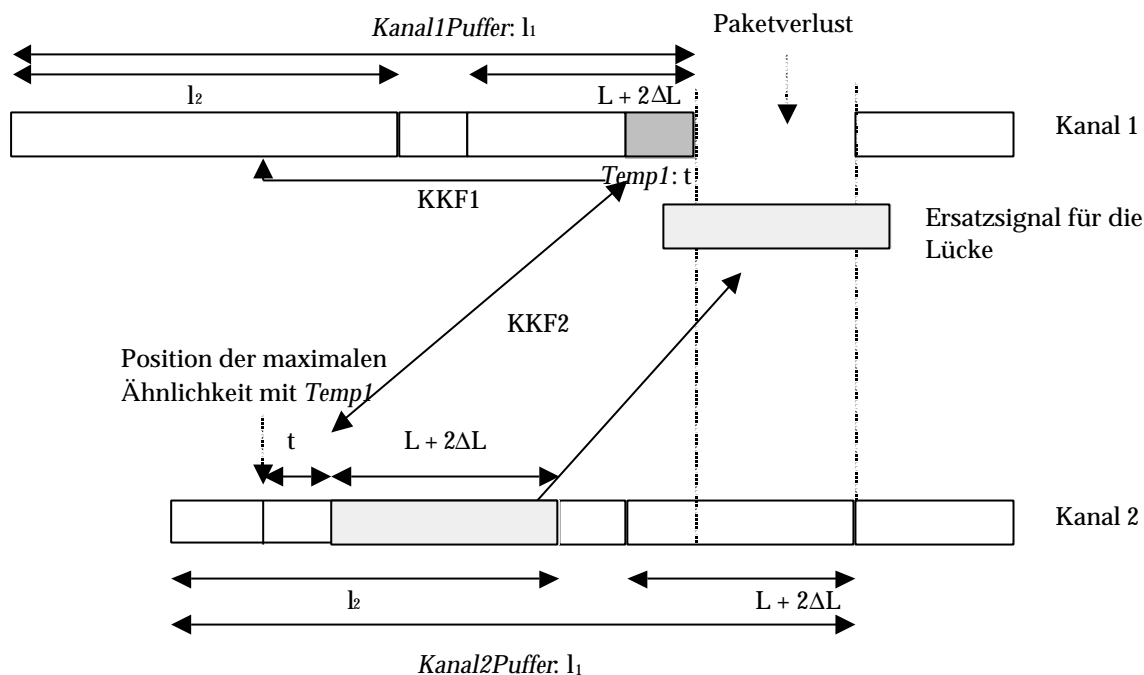


Bild 4.10: Kreuzkorrelationsfelder für das Verfahren der maximalen Kreuzkorrelation. Hier tritt ein Paketverlust in Kanal 1 auf. Kanal 2 bleibt fehlerfrei. Das Verfahren untersucht die Kreuzkorrelation von $Temp1$ mit den ersten l_2 Abtastwerten aus $Kanal1Puffer$ (KKF1) und $Kanal2Puffer$ (KKF2).

Die Idee der Kreuzkorrelations-Untersuchung zwecks Paketsubstitution bei Paketverlusten in Audiosignalen wurde 1986 von Goodman eingeführt und mit *Pattern Matching* (PM) bezeichnet. *Pattern Matching*-Verfahren wurden speziell für stimmhafte Sprache entwickelt [Goo86].

Goodman hat mehrere Möglichkeiten zur Untersuchung der Ähnlichkeiten vorgeschlagen: Als erstes hat er ein Kreuzkorrelationsmaß $C(n)$ definiert und untersucht:

$$C(n) = \frac{\sum_{m=1}^t x(m)y(n+m)}{\sum_{m=1}^t [y(n+m)]^2}, \quad n = 1, 2, \dots, l_1 \quad (4.4)$$

wobei

- t: Länge des *Templates*,
- l_1 : Pufferlänge,
- $x(i)$: Abtastwerte des *Templates* und
- $y(i)$: Abtastwerte des Puffers.

Das Ergebnis der Suche ist der Index n , bei dem $C(n)$ maximal ist.

Eine weitere Variante des PM-Verfahrens basierte auf die Untersuchung der Signalverlaufdifferenz. Goodman hat hier auf drei Berechnungsmethoden hingewiesen:

$$D_1(n) = \sum_{m=1}^t \left| \frac{x(m)}{\sqrt{\sum_{j=1}^t [x(j)]^2}} - \frac{y(n+m)}{\sqrt{\sum_{j=1}^t [y(n+j)]^2}} \right| \quad (4.5)$$

$$D_2(n) = \sum_{m=1}^t \left| \frac{x(m)}{\sum_{j=1}^t |x(j)|} - \frac{y(n+m)}{\sum_{j=1}^t |y(n+j)|} \right| \quad (4.6)$$

$$D_3(n) = \sum_{m=1}^t \left| \frac{x(m)}{x_{\max} - x_{\min}} - \frac{y(n+m)}{y_{\max} - y_{\min}} \right| \quad (4.7)$$

wobei,

$$x_{\max} = \max\{x(1), x(2), \dots, x(t)\}, \quad x_{\min} = \min\{x(1), x(2), \dots, x(t)\} \quad (4.8)$$

$$y_{\max} = \max\{y(n+1), y(n+2), \dots, y(n+t)\} \text{ und} \quad (4.9)$$

$$y_{\min} = \min\{y(n+1), y(n+2), \dots, y(n+t)\}$$

Das Ergebnis der Suche ist hier der Index n , bei dem $D(n)$ minimal ist.

Die Signalverlaufdifferenz-Methode (insbesondere D_1 und D_2) lieferte bei Sprache die besten Ergebnisse [Goo86].

Der Signalverlauf stimmhafter Sprache ist meist quasiperiodisch. Es ist deswegen denkbar, eine Ähnlichkeit des Signalverlaufs mit dessen Vergangenheit zu finden. Diese Ähnlichkeit wird mit dem *Pattern Matching*-Verfahren entweder mit einem Maximum des Kreuzkorrelationsmaßes oder Minimum der Differenzsignale erfasst.

In dieser Arbeit wurde untersucht, ob das *Pattern Matching*-Verfahren auch für die Signalrekonstruktion bei Paketverlusten in Audiosignalen geeignet ist, die einen komplexeren Verlauf aufweisen als stimmhafte Sprache. Die Untersuchungen haben ergeben, dass -im Gegensatz zu dem Einsatz bei stimmhafter Sprache- das absolute Maximum der Kreuzkorrelation häufig nicht mit der Position der maximalen Ähnlichkeit mit dem *Template* übereinstimmt. Stimmhafte Sprache weist üblicherweise keine sprunghaften Änderungen des Signalverlaufs und der Signalstatistik auf. Deswegen ist es hier denkbar, dass *Pattern Matching*-Verfahren für die Paketsubstitution bei Paketverlusten in Sprachsignalen gut funktionieren.

Audiosignale haben oft unerwartete, sprunghafte Änderungen des Signalverlaufs, insbesondere, wenn es sich dabei um Schlaginstrumente handelt. Der Kreuzkorrelationswert kann stark von der Energie beider Signale beeinflusst werden. Das absolute Maximum des KKF-Wertes ist daher nicht unbedingt ein Maß für die maximale Korrelation beider Signale. Dies erklärt, wieso das *Pattern Matching*-Verfahren bei solchen Signalen nicht funktionieren kann.

Für eine Entscheidung, ob es sich bei dem Kreuzkorrelationsmaximum tatsächlich um eine maximale Ähnlichkeit beider Signale handelt, sollte daher zu dem Kreuzkorrelationswert noch ein weiteres Maß verwendet werden. Die Energie der Differenz beider Signale soll an der Stelle auch ihr Minimum haben.

In Bild 4.11 ist zur Veranschaulichung ein Beispiel dargestellt. Bilder c) und d) stellen Ausschnitte aus den ermittelten Kreuzkorrelationssignalen KKF1 und KKF2 dar, wobei KKF1 (Bild a) die Kreuzkorrelation des Template mit dem Puffer des eigenen Kanals und KKF2 (Bild b) Kreuzkorrelation des Template mit dem Puffer aus dem Nachbarkanal darstellen. In diesem Beispiel handelt es sich um ein Paketverlust im linken Kanal eines Pop-Stereosignals, der rechte Kanal ist fehlerfrei.

In Bildern e) und f) sind die dazugehörigen Energiewerte dargestellt. Hierbei handelt es sich um die Energiewerte der Differenzsignale zwischen dem Template und einem Signalausschnitt der selben Länge aus dem Puffer, an der Positionen der lokalen Maxima der Kreuzkorrelationsfunktionen (Bild 4.11-e für KKF1 und Bild 4.11-f für KKF2).

Das Ergebnis der Kreuzkorrelationsuntersuchung ergab, dass das optimale Substitutionssignal für die Lücke aus dem Nachbarkanal zu entnehmen ist. Es wird also bei diesem Kanal eine Stereokorrelation festgestellt. Die Energie des Differenzsignals hat ein absolutes Minimum bei dem Abtastwert 1505 (Bild 4.11-f). Die gesamte Länge des Kreuzkorrelationsfeldes beträgt hierbei 2306¹ Abtastwerte.

Die Position des absoluten Minimums der Energie stimmt nicht mit der Position des absoluten Maximums der Kreuzkorrelationsfunktion KKF2 aus Bild 4.11-b) überein; Die Energie des Differenzsignals ist an dieser Stelle viel größer als das Minimum.

Ein Vergleich beider Kreuzkorrelationsfunktionen ergibt, dass sich der maximale Kreuzkorrelationswert im Signal KKF1 (Bild 4.11-a) bei einer Verschiebung von 200 Abtastwerten befindet. Dieses deutet im allgemeinen auf eine Monokorrelation (Ähnlichkeit des Signals mit seiner Vergangenheit) hin. Das dabei untersuchte Audiosignal ist ein Stereosignal aus einem Pop-Musik-Titel mit einer hohen Ähnlichkeit der Signalverläufe beider Kanäle. Eine Monokorrelation in jedem Kanal ist in diesem Signal nicht vorhanden.

¹ Die Pufferlänge beträgt hierbei 3600 Abtastwerte, die Paketlänge 1024 und die Länge des Mergingintervalls 100 Abtastwerte.

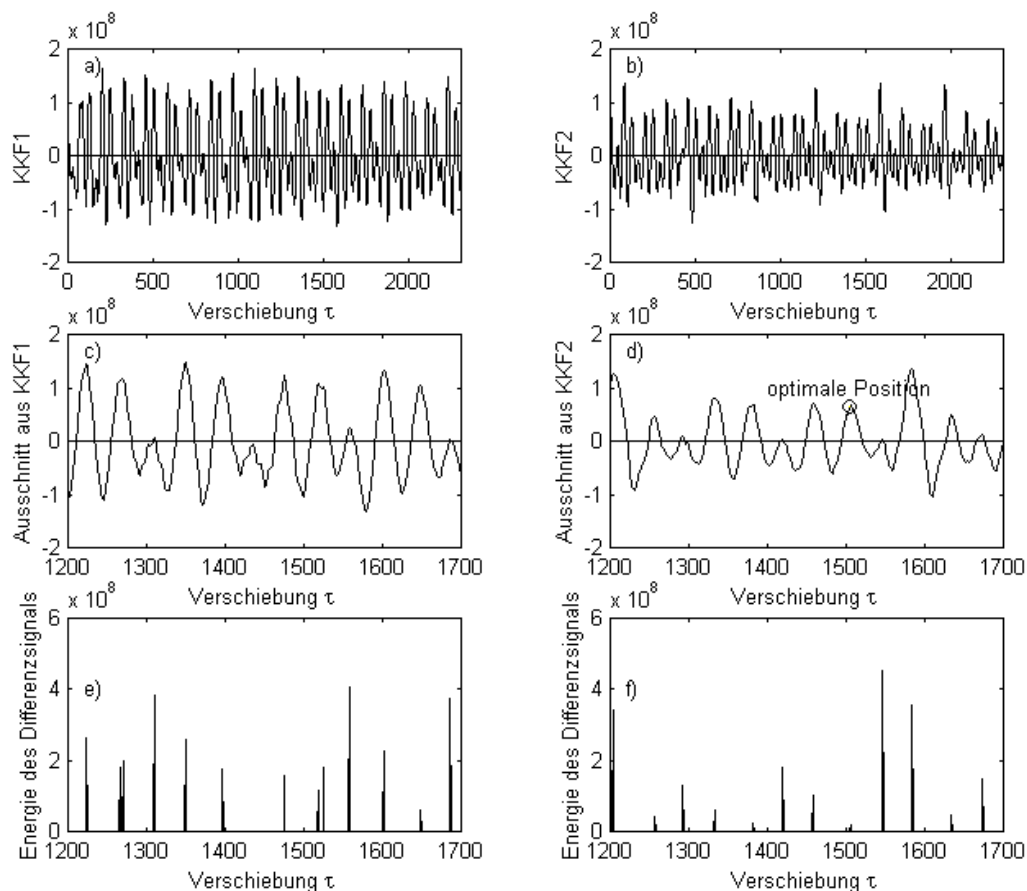


Bild 4.11: Kreuzkorrelationsfunktionen und die dazugehörigen Energiewerte der Differenzsignale zwischen dem Template und ein Audiosignalausschnitt aus dem Puffer an der Position des lokalen Maximums der Kreuzkorrelationsfunktion.

a) Kreuzkorrelationsfunktion des Template und Puffer aus dem selben Kanal: KKF1; c) Energie des Differenzsignals an den lokalen Maxima von KKF1; b) Kreuzkorrelationsfunktion des Template und Puffer aus dem Nachbarkanal :KKF2; d) Energie des Differenzsignals an den lokalen Maxima von KKF2.

Die Tatsache, dass ein Maximum des Kreuzkorrelationswertes nicht ausreichend für die Beschreibung der Ähnlichkeit zwischen Audiosignalen ist, wird in Bild 4.12 bestätigt. Es handelt sich hier um das Audiosignal, aus dem die Kreuzkorrelations- und Energiewerte aus Bild 4.11 ermittelt wurden.

Für die Rekonstruktion des Audiosignals aus Bild 4.12-d) wurde im MK-Verfahren sowohl die Kreuzkorrelation als auch die Energie der Differenzsignale untersucht und als Maß für die Bestimmung der optimalen Position im Puffer verwendet. Diese optimale Position entspricht der Verschiebung, bei der die Kreuzkorrelationsfunktion ein **lokales** Maximum und die Energie des Differenzsignals ein **absolutes** Minimum besitzen (Position 1505 in KKF2 aus Bild 4.11-d). Das Verfahren gewährleistet mit dieser Kombination und bei diesem Audi-

obeispiel eine perfekte Rekonstruktion des Signalverlaufs im Vergleich zu dem Originalsignal. Akustisch war an der Fehlerstelle kein Unterschied zwischen Original- und Substitutionssignal wahrnehmbar.

Bei dem in Bild 4.12-e) dargestellten Audiosignal wurde die Rekonstruktion nur mithilfe der Kreuzkorrelationsuntersuchung durchgeführt (Prinzip des *Pattern Matching*-Verfahren). Die optimale Position im Puffer zeichnet sich durch ein **absolutes** Maximum der Kreuzkorrelationsfunktion. In diesem Beispiel ist es die Position 200 in KKF1 (Bild 4.11-a). Der Signalverlauf des Rekonstruktionssignals weist keine Ähnlichkeit mit dem Verlauf des Originalsignals auf. Die Audioqualität ist an dieser Stelle störend.

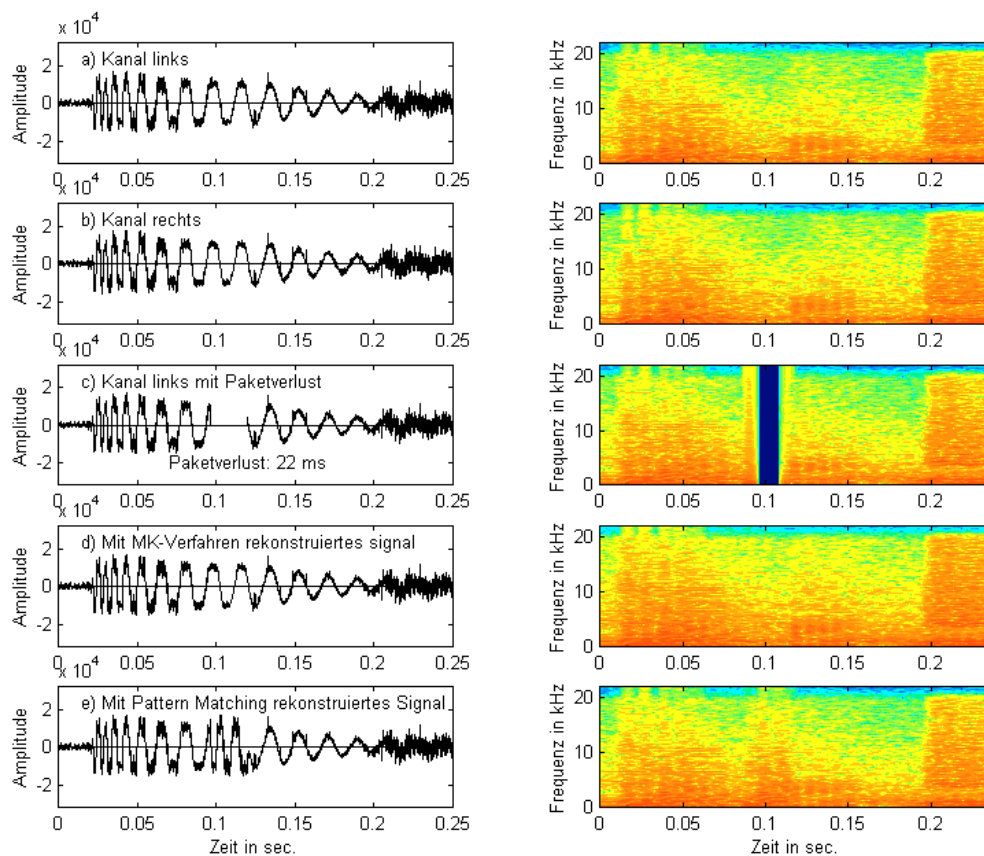
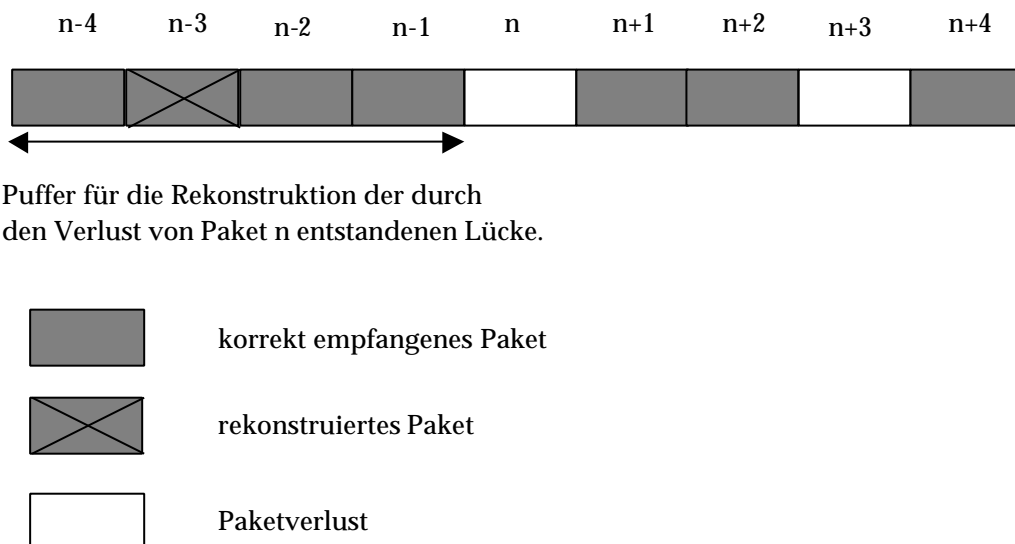


Bild 4.12: Signalverläufe und Spektrogramme bei dem MK-Verfahren. Originalsignal aus Pop-Musik-Titel (250 ms); a) Linker Kanal des Stereosignals; b) Rechter Kanal des Stereosignals; c) Linker Kanal mit Paketverlust; d) Mit dem MK-Verfahren rekonstruiertes Signal für den linken Kanal. Hier wird zur Rekonstruktion sowohl die Kreuzkorrelationsfunktion als auch die Energie der Differenzsignale untersucht; e) Mit *Pattern Matching* rekonstruiertes Audiosignal.

Hinweis:

Das MK-Substitutionsverfahren erfordert eine kontinuierliche Pufferung der Audiodaten in jedem Kanal. Es werden sowohl korrekt empfangene als auch rekonstruierte Signalaus-

schnitte gepuffert. Das Audiosignal eines Puffers kann also auch bereits rekonstruierte Signale bei früher auftretenden Paketverlusten enthalten (s. Bild 4.13). Das kann zu Diskontinuitäten in dem Signalverlauf im Puffer führen, wenn der Übergang zwischen korrekt empfangene und rekonstruierte Audiodaten nicht geglättet wird (Merging aus Kapitel 4.3.7). Diese Diskontinuitäten beeinflussen die Qualität der Audiosignalrekonstruktion von später auftretenden Paketverlusten. z.B. wird die Audioqualität für das Substitutionssignal von Paket n von der früher erfolgten Substitution des Paketes $n-3$ beeinflusst. Dies beeinflusst wiederum die Rekonstruktion der Lücke, entstanden durch den späteren Verlust von Paket $n+3$. Um diese Diskontinuitäten im Puffer zu vermeiden, sollte im fehlerfreien Fall nicht das Eingangssignal $x(n)$ aus Bild 4.14 sondern das Ausgangssignal $y(n)$ gepuffert werden. $y(n)$ unterscheidet sich bei einem fehlerfreien Empfang eines Audiopaketes nur in der Umgebung eines Paketverlustes (unmittelbar vorher oder nachher), wenn Merging verwendet wird.



Puffer für die Rekonstruktion der durch den Verlust von Paket n entstandenen Lücke.

Bild 4.13: Puffer für die Rekonstruktion der Lücke aus Paket n mit dem MK-Verfahren. Der Puffer enthält bereits rekonstruierte Substitutionssignale (Paket $n-3$).

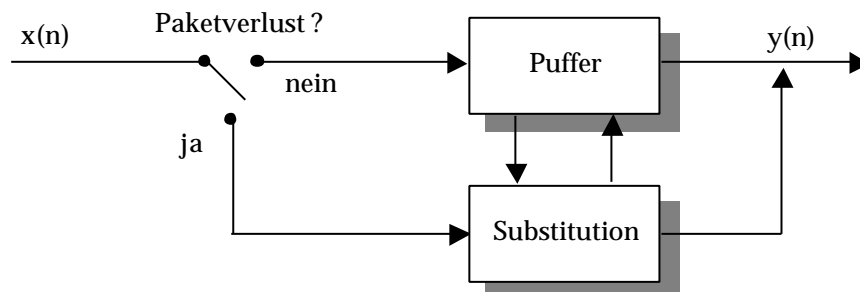


Bild 4.14: Pufferung der Audiodaten bei Verfahren der maximalen Kreuzkorrelation.

4.3.7 Packet Merging

Um die aufgrund von Diskontinuitäten in dem Signalverlauf zwischen korrekt empfangenen und rekonstruierten Audiosignalen bei Paketverlusten verbleibenden Knackgeräusche zu vermindern, können die Übergänge zwischen dem Substitutionspaket und den angrenzenden Paketen durch Überlappung einiger Abtastwerte geglättet werden.

In [Goo86] wird dafür ein „*Packet Merging*“-Verfahren vorgeschlagen. Bei der Substitution eines verlorenen Paketes der Länge L wird ein Paket der Länge $L' = L + 2\Delta L$ verwendet. Goodman verwendet ein *Merging*-Intervall der Länge $\Delta L = 1\text{ms}$. Die ersten und letzten ΔL Abtastwerte des Substitutionspaketes werden den ΔL Abtastwerten des letzten bzw. folgenden Paketes mit einer entsprechenden Gewichtung überlagert (s. Bild 4.15)

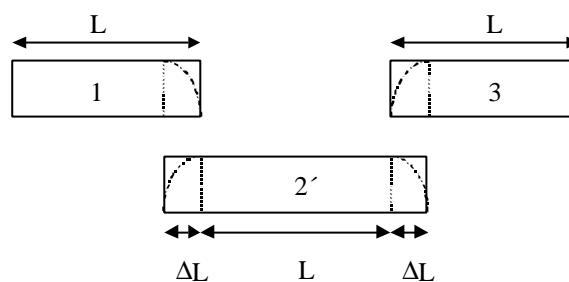


Bild 4.15: *Packet Merging*

Das Substitutionspaket (hier Paket 2') wird an seinen Grenzen mit einem von 1 auf 0 fallenden Gewichtungsfaktor versehen. Das nächste Paket (Paket 3) wird im Gegensatz dazu mit einem von 0 auf 1 steigenden Gewichtungsfaktor versehen, so dass die Summe der Gewichte immer eine 1 ergibt.

Packet Merging kann in allen Substitutionsverfahren, auch in den trivialen Methoden, wie z.B. Nullsubstitution oder Paketwiederholung, verwendet werden. Die Überblendung liefert aber bei diesen einfachen Substitutionsverfahren keine große Verbesserung der Signalqualität, da die Störgeräusche hier durch das primitive Substitutionssignal bedingt sind.

Im folgenden wird gezeigt, wie diese oben genannten Substitutionstechniken auf der Ebene des Differenzsignals nach einer Dekorrelation des Signals mit einer linearen Prädiktion angewendet werden können.

4.4 Substitution mittels linearer Prädiktion

Die Rekonstruktion fehlender Audiosignalblöcke beruht auf der Annahme, dass die Eigenschaften des Audiosignals eines verlorengegangenen Paketes und die des Signals aus dem zuletzt korrekt empfangenen Paketes ähnlich sind. Hier wird von einer Stationarität des

Audiosignals über kurze Zeitabschnitte (Quasistationarität) ausgegangen. Zur Rekonstruktion fehlender Audiosignalblöcke könnte die Annahme der Quasistationarität auf die Dauer des Verlustes ausgedehnt werden, so dass eine Extrapolation des Signals unter Verwendung der unmittelbar vor dem Verlust bestimmten Parameter (z.B. die Parameter eines linearen Prädiktors) zur Verdeckung der Lücke im Audiosignal ausreichen würde.

Auf der Annahme der Quasistationarität von Audio- und Sprachsignalen beruhen einige in der Audio- und Sprachcodierung bekannten Verfahren. Ein Beispiel ist die lineare Prädiktion, die oft zur Codierung von Sprachsignalen verwendet wird (**L**inear **P**redictive **C**oding LPC). Die lineare Prädiktion nutzt die bekannte Vergangenheit eines Sprachsignals, um seinen aktuellen Wert zu schätzen. Für einen Wert $x(n)$ des Signals kann ein Schätzwert $\hat{x}(n)$ aus vergangenen Signalwerten $x(n-i)$ wie folgt bestimmt werden:

$$\hat{x}(n) = \sum_{i=1}^K a_i x(n-i). \quad (4.10)$$

Dies entspricht einer Filterung des Sprachsignals mit einem Transversalfilter (Prädiktor) k -ter Ordnung. a stellen dabei sind die Koeffizienten des Prädiktors, die für eine optimale Schätzung über kurze Zeitabschnitte adaptiv ermittelt werden. Es existieren verschiedenen Methoden zur Bestimmung der Filterkoeffizienten, die sich sowohl im erforderlichen Rechenaufwand als auch in der Qualität der Analyse unterscheiden. Das Verfahren der kleinsten Fehlerquadrate über die Autokorrelationsmethode liefert gute Analyse-Ergebnisse bei geringem Rechenaufwand im Vergleich zu anderen Verfahren [Mar76].

Durch Subtraktion des Schätzwertes $\hat{x}(n)$ vom ursprünglichen Signal $x(n)$ erhält man das folgende Differenzsignal (Innovation) $i(n)$, auch Prädiktionsfehler genannt, das bei hinreichend guter Schätzung wesentlich kleiner als $x(n)$ selbst ist.

$$i(n) = x(n) - \hat{x}(n) = x(n) - \sum_{i=1}^K a_i x(n-i). \quad (4.11)$$

$i(n)$ ist hier gleichbedeutend mit einer konstanten Einhüllenden des Fehlerspektrums [Clü98a].

Die Rekonstruktion des mit LPC codierten Sprachsignals erfolgt mittels rekursivem Synthesefilter, welcher die bei der Codierung verwendeten Filterkoeffizienten kennt.

In dieser Arbeit wird untersucht, ob die Parametrisierung von Audiosignalen mittels linearer Prädiktion und bei einem Paketverlust die Synthese eines Substitutionssignals unter Verwendung der LPC-Parameter der zuletzt korrekt empfangenen Daten zu qualitativ guten Verdeckung der durch den Paketverlust entstandenen Lücke im Audiosignal führt.

4.4.1 Synthese eines Substitutionssignals aus der Innovation

In diesem Abschnitt wird gezeigt, wie die lineare Prädiktion zur Rekonstruktion von Audiosignalen bei Paketverlusten verwendet werden kann (Bild 4.16).

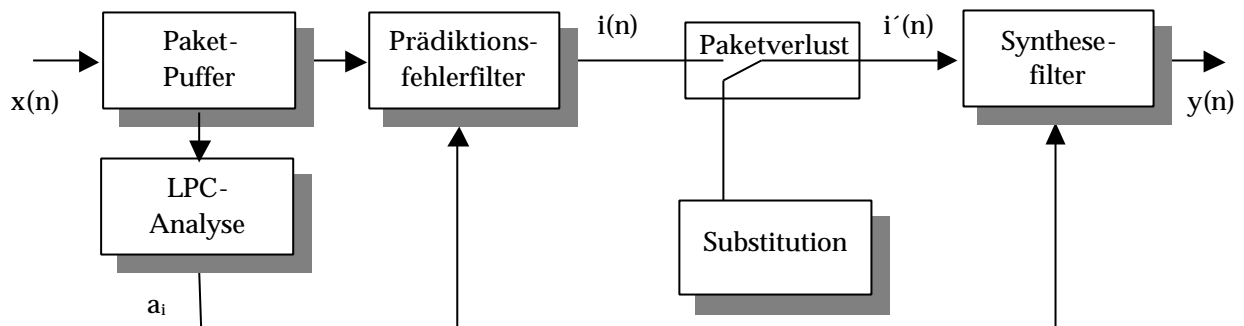


Bild 4.16: Rekonstruktion von Lücken in Audiosignalen mittels Prädiktionsanalyse und -synthese.

Mit der LPC-Analyse werden optimale Filterkoeffizienten a aus dem zuletzt gepufferten korrekt empfangenen Audiosignal berechnet. Die berechneten Filterkoeffizienten werden zur Adaption von zwei zueinander inversen Filtern, des Prädiktionsfehlerfilters und des Synthesefilters, herangezogen.

Das decodierte Audiosignal $x(n)$ durchläuft zunächst das Prädiktionsfehlerfilter, welches die Innovation $i(n)$ erzeugt (s. Bild 4.16). Im Fall eines gültigen Paketes ist $i'(n)$ identisch mit $i(n)$. Am Ausgang des Synthesefilters liegt wieder das Eingangssignal unverändert vor ($y(n) = x(n)$).

Bei einem Paketverlust können keine neuen Filterkoeffizienten berechnet werden. Die Koeffizienten des Synthesefilters, die im letzten Zyklus berechnet wurden, werden beibehalten. Am Eingang dieses Filters muss in diesem Fall eine Innovation erzeugt werden, da auch diese nicht zur Verfügung steht.

Von [Clü98a] wurden für den Fall einer digitalen **Sprachübertragung** über paketvermittelnde Netze mehrere Verfahren zur Erzeugung einer Innovation untersucht. In dieser Arbeit werden diese Möglichkeiten auch für Audiosignale, insbesondere für Stereosignale, untersucht und modifiziert.

Die Substitution der fehlenden Innovation im Fall eines Paketverlustes erfolgt mit einem der folgenden Verfahren.

4.4.1.1 Einfache Verfahren

Im einfachsten Fall wird eine **Nullsubstitution** vorgenommen. Hier wird die Innovation durch Einfügen von Null-Abtastwerten ersetzt. Am Eingang des Synthesefilters liegt das Anregungssignal $i'(n)$ mit:

$$i'(n) = 0, \quad n \in \{0 \dots L + 2 \cdot \Delta L - 1\}$$

L ist hierbei die Länge der fehlenden Innovation.

Bei der **Paketwiederholung** wird die Innovation des zuvor korrekt empfangenen Paketes wiederholt; $i'(n) = i(n - L - 2 \cdot \Delta L)$.

Bei der Substitution mit **Rauschen** besteht das Anregungssignal des Synthesefilters in Bild 4.16 im Fall eines Paketverlustes aus einem weißen Rauschen, das vom Synthesefilter, das noch die Filterkoeffizienten des letzten Paketes enthält, spektral gefärbt wird. Für die Innovation $i'(n)$ gilt dann:

$$i'(n) = (\text{random}(n) - 0.5) \cdot \sqrt{12} \cdot \sigma_i, \quad (4.12)$$

$\text{random}(n)$ ist eine gleichverteilte Zufallsfunktion, die Zufallswerte zwischen 0 und 1 erzeugt. σ_i ist die Standardabweichung der Innovation im letzten korrekt empfangenen Paket.

4.4.1.2 Pitchsynchrone Paketwiederholung

Die pitchsynchrone Signalwiederholung zur Rekonstruktion von Audiosignalen wurde 1986 von Goodman als Substitutionsverfahren für fehlende Sprachsignalsegmente vorgeschlagen [Goo86]. Das Verfahren nutzt die Tatsache, dass stimmhafte Sprachsignalsegmente einen näherungsweise periodischen Verlauf aufweisen können. Die Kenntnis der Periodendauer¹ wird hier bei der Wahl des Rekonstruktionssignals für die Lücke berücksichtigt.

In [Clü98a] wurde dieses Substitutionsverfahren ergänzt. Neben der Grundfrequenz wurde die spektrale Einhüllende des Sprachsignals, die in Form von linearen Filterkoeffizienten vorliegt, bei der Rekonstruktion fehlender Sprachblöcke verwendet. Das Verfahren basiert auf linearer Prädiktion.

In dieser Arbeit wird das Verfahren auch für Audiosignale wie z.B. klassische Musik, einzelne Instrumente, etc. untersucht. Es folgt eine kurze Beschreibung des Verfahrens (s. Bild 4.17):

¹ In der Sprachsignalverarbeitung Grundfrequenz oder auch *Pitch* genannt.

Im fehlerfreien Fall werden aus dem korrekt empfangenen Audiosignalausschnitt die Filterkoeffizienten a_i und die Innovation $i(n)$ berechnet. $i'(n)$, die in diesem Fall mit $i(n)$ identisch ist, wird für die spätere Grundfrequenzanalyse gepuffert. Das Ausgangssignal ist mit dem Eingangssignal identisch. Es wird hier, im Gegensatz zu dem in [Clü98a] vorgeschlagenen Verfahren keine, Synthese durchgeführt.

Im Fall eines Paketverlustes wird aus dem Puffer der Innovation die Grundfrequenz ermittelt (s. Kap. 2.3.3). Es wird hier die Autokorrelationsmethode zur Bestimmung der Grundfrequenz verwendet. Mit der ermittelten Pitchperiode G wird die Innovation $i'(n)$ folgendermaßen bestimmt:

$$i'(n) = i'(n - G - \Delta L), \quad n = 0 \dots L + 2\Delta L - 1. \quad (4.13)$$

L: Länge der fehlenden Innovation (Paketlänge)

ΔL : Länge des Mergingintervalls.

Dieses Substitutionsverfahren erfordert eine kontinuierliche Berechnung und Pufferung der Innovation sowohl im fehlerfreien als auch im fehlerbehafteten Fall.

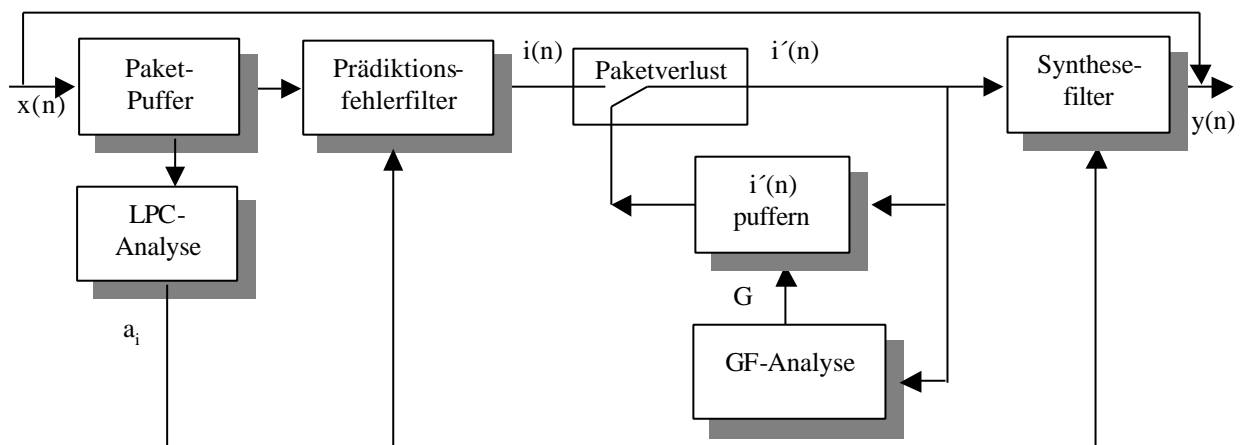


Bild 4.17: Pitchesynchrone Rekonstruktion der Innovation

Bild 4.18 zeigt ein Beispiel eines Audiosignals mit einem näherungsweise periodischen Verlauf. Es handelt sich hier um den linken Kanal eines Stereosignal aus einem Einzelinstrument-Titel (Orgel). Der Signalverlauf von Orgel-Musik ist über längere Segmente fast periodisch (s. Kapitel 5.4). Mit der pitchesynchronen Paketwiederholung wird diese Periodizität erfasst. Das Substitutionssignal (Bild 4.18-c) wird phasenrichtig in das übrige Audiosignal eingefügt.

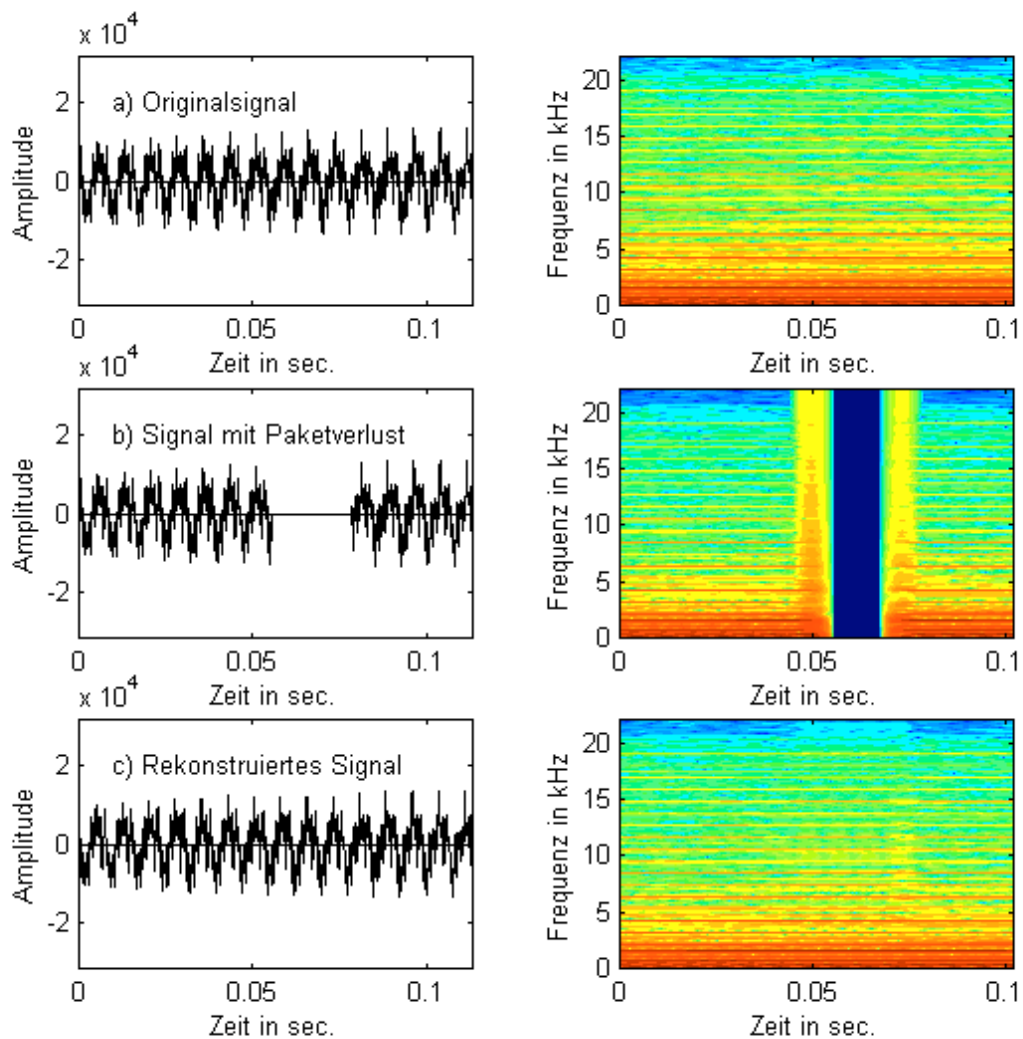


Bild 4.18: Paketsubstitution mit pitchsynchrone Paketwiederholung. a) Originalsignal (5000 Abtastwerte \cong 113 ms bei 44,1 kHz Audio) aus einem Einzelinstrument-Titel mit einem näherungsweise periodischen Signalverlauf; b) Signal mit Paketverlust. Die Paketlänge beträgt hierbei 1024 Abtastwerte (22 ms); c) Mit der pitchsynchrone Paketwiederholung rekonstruiertes Signal. Rechts sind die Spektrogramme der einzelnen Signale dargestellt.

Die Substitution der Lücke mit dem Verfahren der pitchsynchrone Paketwiederholung liefert im nächsten Beispiel (Bild 4.19) ein Substitutionssignal, das an der Fehlerstelle (Paketverlust) keine Ähnlichkeit mit dem Verlauf des Originalsignals aufweist. Es handelt sich hier um einen Ausschnitt von über 5000 Abtastwerte (113 ms) aus dem linken Kanal eines Stereosignals eines klassischen Musik-Titels (Beethoven), das keine Periodizitäten in seinem Verlauf enthält (Bild 4.19-a). Mit der pitchsynchrone Paketwiederholung wird eine Pitchperiode aus dem Signal unmittelbar vor der Lücke ermittelt und das Signal der Lücke periodisch fortgesetzt. Das Audiosignal ist nicht periodisch und kann also keine eindeutige Pitchperiode besitzen, das Substitutionssignal kann also mit dem Originalsignal nicht übereinstimmen (Bild 4.19-c).

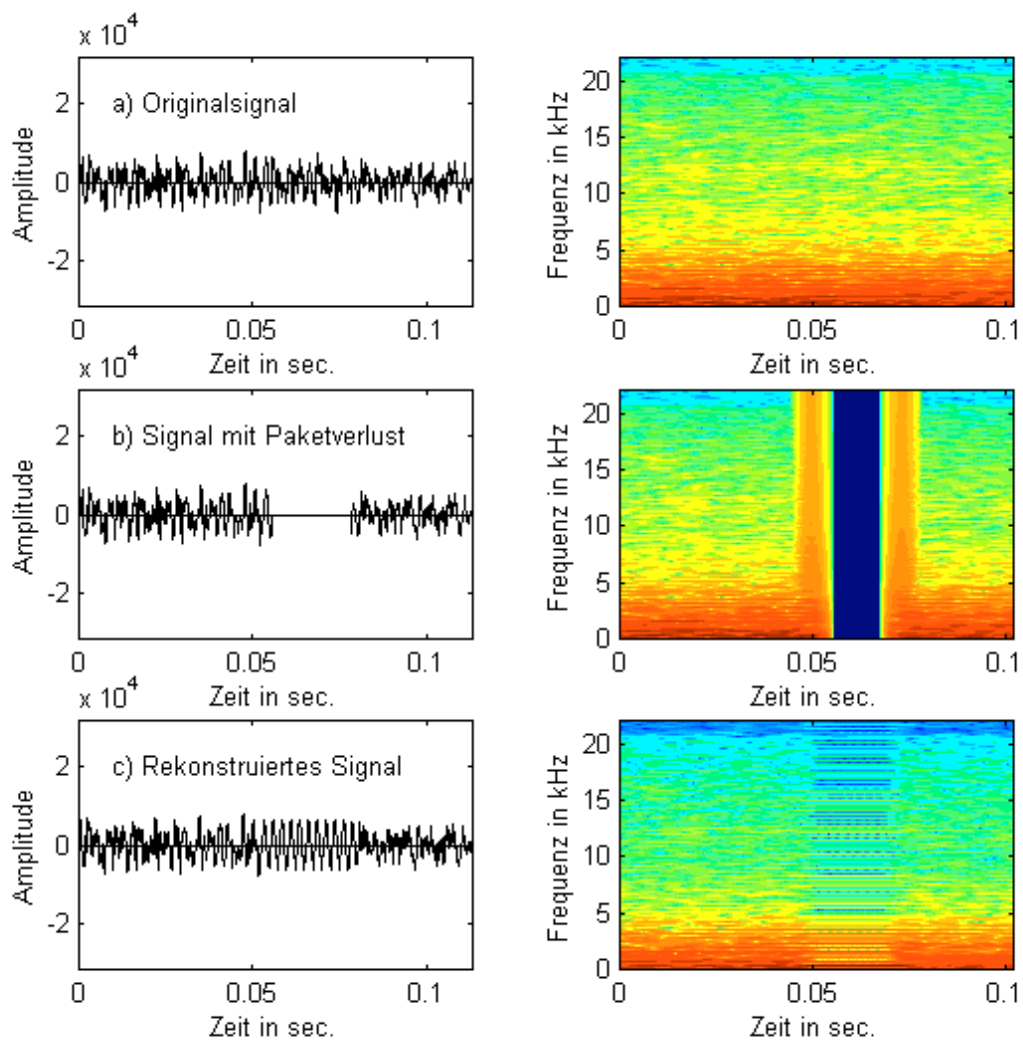


Bild 4.19: Paketsubstitution mit pitchsynchrone Paketwiederholung. Originalsignal (113 ms) aus einem klassischen Musik-Titel; a) Linker Kanal des Stereo-Audiosignals; b) Linker Kanal mit Paketverlust. Die Paketlänge beträgt hierbei 1024 Abtastwerte (22 ms); c) Mit der pitchsynchrone Paketwiederholung rekonstruiertes Signal. Rechts sind die Spektrogramme der einzelnen Signale dargestellt.

Das Verfahren der pitchsynchrone Paketwiederholung eignet sich nur zur Rekonstruktion von Lücken bei Signalen mit periodischen Segmenten, wie z.B. bei einigen Einzelinstrumenten. Pop- oder klassische Musiksignale lassen sich mit diesem Verfahren nur schlecht rekonstruieren.

4.4.1.3 Links/Rechts-Ersatz der Innovation

Die Idee der Rekonstruktion von Audiosignalen bei Paketverlusten auf der Ebene der LPC-Innovation wurde in dieser Arbeit auf Stereo-Audiosignale erweitert, um von der in einigen Stereosignalen eventuell vorhandenen Korrelation zwischen beiden Kanälen für die Rekonstruktion von Lücken in Audiosignalen zu profitieren.

Aufnahmen von Audiosignalen in Stereotechnik weisen offensichtlich gewisse - von der Aufnahme abhängige - Gemeinsamkeiten zwischen dem Inhalt des linken und des rechten Stereokanals auf [Her95]. Zur Ausnutzung einer solchen Korrelation scheint es naheliegend, die entsprechenden Werte (bei diesem Verfahren die Innovation) aus dem jeweils anderen Stereokanal als Ersatz für die fehlenden Werte im fehlerbehafteten Kanal heranzuziehen. Dazu muss die Innovation, also das Paket, aus dem Nachbarkanal zur Verfügung stehen.

Im fehlerfreien Fall werden aus den in beiden Kanälen (Kanal K_1 und Kanal K_2 aus Bild 4.20) korrekt empfangenen Paketen die Filterkoeffizienten Koz_1 , Koz_2 und die Innovationen $i_1(n)$, $i_2(n)$ berechnet. Bei einem Paketverlust in einem der beiden Kanäle werden zur Synthese eines Rekonstruktionssignals zur Verdeckung der Lücke in dem fehlerbehafteten Kanal die Innovation aus dem Nachbarkanal und die Koeffizienten des zuletzt korrekt empfangenen Paketes desselben Kanals genommen.

Sollte z.B. ein Paket nur in Kanal K_1 fehlen, so werden für die Synthese eines Substitutionssignals zur Verdeckung der durch den Paketverlust entstandenen Lücke im Audiosignal dieses Kanals die Innovation $i_2(n)$ aus dem fehlerfreien Kanal K_2 und die Filterkoeffizienten Koz_{1a} des zuletzt korrekt empfangenen Paketes in Kanal K_1 verwendet.

Dies ist allerdings nur dann von Vorteil, wenn die Signalverläufe in beiden Kanälen ähnlich sind. Andernfalls würde das Verfahren nur zu einer zusätzlichen Verschlechterung der Audioqualität führen.

Bei einem Paketverlust in beiden Kanälen kann diese Substitutionstechnik nicht mehr angewendet werden, hier wird die fehlende Innovation mit einem Rauschsignal ersetzt. Zur Synthese der PCM-Substitutionssignale werden in jedem Kanal die in dem letzten Zyklus berechneten Filterkoeffizienten verwendet.

Ähnlich wie die Substitution der Lücken mit dem Links-/Rechts-Ersatz der PCM-Signale (Kapitel 4.3.5), werden mit diesem Verfahren nur bei korrelierten Stereosignalen gute Ergebnisse erreicht.

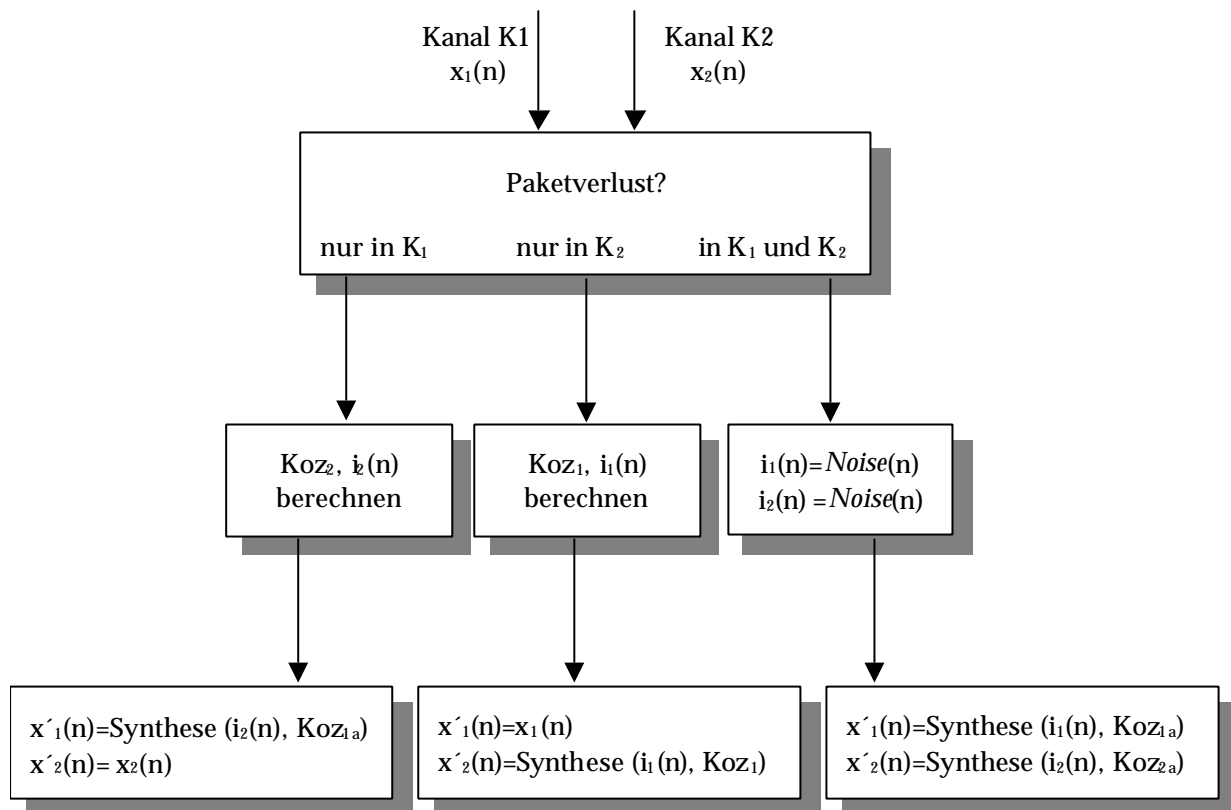


Bild 4.20: Links-/Rechts-Ersatz der Innovation bei Paketverlust

- Koz_1 : Filterkoeffizienten aus dem aktuellen Paket in Kanal 1.
 Koz_2 : Filterkoeffizienten aus dem aktuellen Paket in Kanal 2.
 Koz_{1a} : Filterkoeffizienten aus dem letzten Paket in Kanal 1.
 Koz_{2a} : Filterkoeffizienten aus dem letzten Paket in Kanal 2.
 $i_1(n)$: Innovation in Kanal 1.
 $i_2(n)$: Innovation in Kanal 2.
 $Noise(n)$: Funktion, mit der ein Rauschsegment erzeugt wird.

4.4.1.4 Gewichtete Interpolation

In diesem Abschnitt wird ein neues Signalsubstitutionsverfahren bei Paketverlusten vorgestellt. Die Substitution basiert weiterhin auf der Parametrisierung des Signals mittels linearer Prädiktion und der Schätzung einer für die Synthese des Substitutionssignals geeigneten Innovation. Das Synthesefilter erfordert für die Erzeugung eines Substitutionssignals sowohl neue Filterkoeffizienten als auch eine passende Innovation als Anregungssignal. Es handelt sich hierbei um ein Interpolationsverfahren, das zur Substitution die Statistik des Signals sowohl vor als auch nach (bei Bedarf) der Lücke heranzieht. Dieses Verfahren wird *Gewichtete Interpolation* (GI) genannt. Die Substitution erfolgt auf der Ebene der LPC-Innovation und wird daher mit LPC-GI bezeichnet.

Fehlende Audiosegmente, die aufgrund von Paketverlusten entstehen, lassen sich prinzipiell mit Extrapolationsverfahren rekonstruieren. Solche Verfahren nutzen zur Rekonstruktion die Kenntnis der Signaleigenschaften kurz vor der Lücke aus. Bei kurzen Paketen kann man von einer Quasistationarität ausgehen, so dass man annehmen kann, dass die Statistik des Signals in dem fehlenden Paket und im vorhergehenden Paket ähnlich ist. Inwieweit solche Extrapolationsverfahren geeignet sind, hängt von der Lage der Lücke im Audiosignal und von den Signaleigenschaften unmittelbar vor und nach der Lücke ab.

Tritt ein Paketverlust in einem rauschähnlichen Signalausschnitt auf, so kann man davon ausgehen, dass sich die Signaleigenschaften unmittelbar vor der Lücke und nachher ähneln. Hier sind Interpolationsverfahren, die zur Rekonstruktion die Signalinformation beiderseits der Lücke ausnutzen, kaum vorteilhafter als Extrapolationsverfahren. Die Kenntnis des Signalverlaufs und dessen Statistik vor der Lücke kann für die Substitution ausreichend sein.

Einfache Extrapolationsverfahren wie z.B. die Paketwiederholung würden spätestens dann scheitern, wenn sich die Statistik des Audiosignals in dem fehlenden Segment von der Statistik des Segments davor wesentlich unterscheidet. Dies kann z.B. der Fall sein, wenn es sich im verlorenen Segment um eine *Attacke* z.B. eines Schlaginstruments handelt. In diesem Fall ist es kaum noch möglich, das Originalsignal nachzubilden. Extrapolationsverfahren können diese Veränderung innerhalb der Lücke nicht erfassen. Man könnte hier höchstens das Substitutionssignal für die Lücke an die vorherige bzw. nachfolgende Signalstatistik anpassen.

Sprachsignale sowie einzelne Musikinstrumente enthalten häufig Signalausschnitte, mit näherungsweise periodischen Signalverlauf. Bei Sprache sind das stimmhafte Sprachsegmenten. Für solche Signale ist die Rekonstruktion fehlender Audiosignal-Segmente unkompliziert und bedarf meistens auch keiner Beteiligung der Informationen aus dem nächsten Paket. Allerdings werden hier Zusatzinformationen wie z.B. eine Stimmhaft-/Stimmlos-Entscheidung für die Signalrekonstruktion benötigt. Ein geeignetes Extrapolationsverfahren, wie die pitchsynchrone Paketwiederholung (Kapitel 4.4.1.2) könnte in diesem Fall eine nahezu perfekte Rekonstruktion des Signalverlaufs gewährleisten.

Das LPC-GI-Verfahren ist ein Substitutionsverfahren, das für die Rekonstruktion der Lücke die Information aus dem Signal beiderseits der Lücke heranzieht. Die Synthese eines Substitutionssignals erfolgt mittels LPC-Synthesefilter, das die zuletzt ermittelten Filterkoeffizienten verwendet. Das Anregungssignal (Innovation) wird mithilfe einer Kreuzkorrelationsuntersuchung der Innovationen vor dem Verlust, ähnlich dem Verfahren der maximalen Kreuzkorrelation aus Kapitel 4.3.6 gewonnen.

Substitution mit LPC-GI

Es folgt eine kurze Beschreibung des vorgeschlagenen Verfahrens:

- Im fehlerfreien Fall werden die Filterkoeffizienten des korrekt empfangenen Paketes und dessen Innovation $i(n)$ berechnet. $i(n)$ ergibt sich aus der Subtraktion des geschätzten Signalausschnittes, der am Ausgang des Prädiktionsfilters vorliegt, vom Original-Signalausschnitt. Bei fehlerfreiem Empfang wird die Innovation fortlaufend berechnet und gepuffert.
- Bei einem Paketverlust muss die fehlende Innovation der Lücke geschätzt werden. Die Schätzung erfolgt nach dem Verfahren der maximalen Kreuzkorrelation aus Kapitel 4.3.6. Die substituierte Innovation muss die Länge $L+\Delta L$ haben.

L: Paketlänge und

ΔL : Länge des Merging-Intervall.

Merging wird nur am Ende des Segmentes durchgeführt, da dessen Anfang durch die Synthesefilterung keine Sprünge aufweist und somit keiner Glättung bedarf.

- Mit der Innovation und den dafür geeigneten Filterkoeffizienten¹ *KozLast* wird ein Substitutionssignal *SubstLast* der Länge $L + \Delta L$ synthetisiert (s. Bild 4.21):

$$\text{SubsLast} = \text{Synthese}(\text{KozLast}, i(n)).$$

- Als nächstes werden die optimalen Filterkoeffizienten *KozNext* des Paketes unmittelbar nach der Lücke (falls vorhanden²) bestimmt.
- Es wird ein neues Substitutionssignal *SubsNext* mit den Filterkoeffizienten des nächsten Paketes synthetisiert.

$$\text{SubsNext} = \text{Synthese}(\text{KozNext}, i(n)).$$

- Zuletzt wird der fehlende Signalausschnitt in der Lücke wie folgt rekonstruiert (s. Bild 4.21):

$$\text{Substitutionssignal} = \text{Gew}_{1-0}(\text{SubsLast}, L + \Delta L) + \text{Gew}_{0-1}(\text{SubsNext}, L + \Delta L).$$

¹ Die Ermittlung der für die Synthese eines Substitutionssignals passenden Filterkoeffizienten wird später erläutert.

² Bei Bündelfehlern wird die Substitution nur anhand der Daten des letzten Paketes durchgeführt (keine Interpolation).

Gew_{1-0} und Gew_{0-1} sind Fensterungsfunktionen, die das Signal von 1 auf 0 bzw. von 0 auf 1 gewichten. Sie sind wie folgt definiert:

$$Gew_{1-0}(x[i]) = 0,5 + 0,5 \cdot \cos\left(\frac{\pi \cdot i}{L + \Delta L}\right) \cdot x[i] \quad (4.14)$$

$$Gew_{0-1}(x[i]) = 0,5 - 0,5 \cdot \cos\left(\frac{\pi \cdot i}{L + \Delta L}\right) \cdot x[i] \quad (4.15)$$

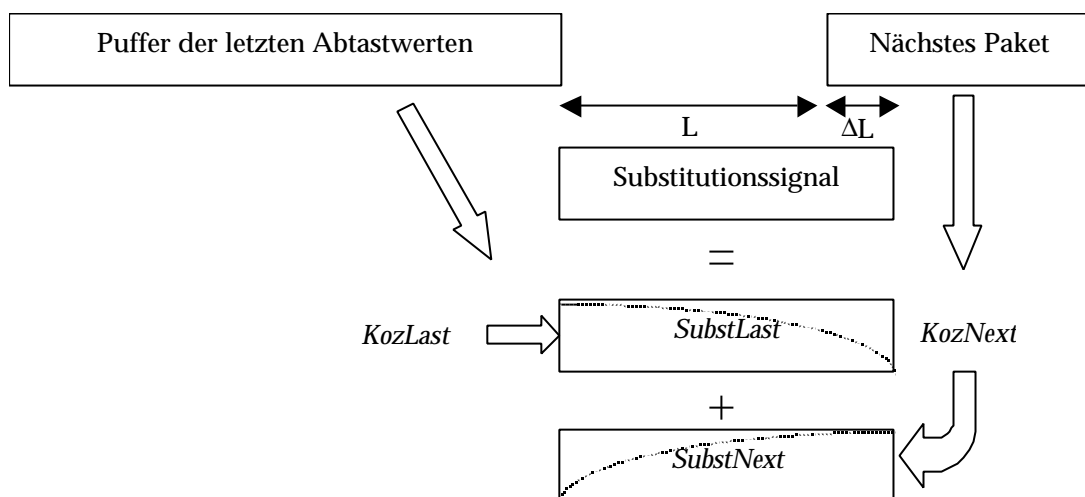


Bild 4.21: Signalsubstitution mit dem LPC-GI-Verfahren. Das Substitutionssignal ergibt sich aus der Summe von *SubstLas*, das sich aus der Synthese der optimalen Innovation aus dem Puffer mit den dazu passenden Koeffizienten ergibt, und *SubstNext*, das sich aus der Synthese dieser Innovation mit den Koeffizienten des nächsten Paketes ergibt.

Hinweise:

- Die Bestimmung der Innovation erfolgt nach dem Verfahren der maximalen Kreuzkorrelation. Dieses erfordert eine kontinuierliche Berechnung und Pufferung der Innovationen sowohl im fehlerfreien als auch im fehlerbehafteten Fall.
- Für die Korrelationsuntersuchung zwecks Ermittlung der besten Innovation für die Lücke besteht eine geringfügige Abweichung von den Parametern aus Kapitel 4.3.6. Aufgrund der einseitigen Glättung (*Merging*) und des daraus ergebenden kürzeren Länge des Substitutionssignals weicht die Länge l_2 des Korrelationsfeldes von der in Gleichung 4.3 definierten Länge ab. Für l_2 gilt jetzt:

$$l_2 = l_1 - (L + \Delta L) - t \quad (4.16)$$

- Es ist hier zu beachten, dass für die Korrelationsuntersuchung nicht das PCM-Signal selbst, sondern die gepufferte Innovation verwendet wird. Es wird damit versucht, Ähnlichkeiten im Prädiktionsfehlersignal (Innovation) zu finden.

Anwendung auf Stereosignale

- Das LPC-GI-Verfahren kann auch auf Stereosignale angewendet werden, somit werden nicht nur Korrelationen mit der Vergangenheit des Signals selbst, sondern auch Ähnlichkeiten zwischen den Kanälen eines Stereosignals untersucht.
 - Falls ein Paket i fehlt, und das zeitgleiche Paket aus dem Nachbarkanal vorhanden ist, wird für die Korrelationsuntersuchung der Puffer aus dem Nachbarkanal entsprechend aktualisiert. Es wird zudem das nächste Paket $i+1$ (falls vorhanden) aus dem Nachbarkanal mitgepuffert. Dies hat einen großen Vorteil insbesondere, weil manchmal verschobene Pakete eine große Korrelation mit dem aktuellen Paket aufweisen.
 - Falls zwischen beiden Kanälen festgestellt wird, dass die Innovation des Nachbarkanals eine höhere Korrelation liefert, wird das Substitutionssignal aus dem Nachbarkanal verwendet. In diesem Fall wird keine Interpolation und anschließende Gewichtung der Substitutionssignale vorgenommen. Es wird hier angenommen, dass die Statistik-Änderung (falls es eine gibt) aufgrund der Stereokorrelation auch in dem Nachbarkanal vorhanden ist. Eine Interpolation ist deshalb in diesem Fall nicht nötig.

Bild 4.22 zeigt die Signalverläufe und die dazugehörigen Spektrogramme des linken Kanals eines Audiosignals aus dem Orgel-Titel (a), des linken Kanals mit Paketverlust (22 ms Lücke) und des mit dem LPC-GI-Verfahren rekonstruiertes Audiosignals. Das Verfahren erkennt die in diesem Kanal vorhandenen Korrelationen und verwendet für die Synthese des Substitutionssignals bereits gepufferte Innovation und Filterkoeffizienten des selben Kanals.

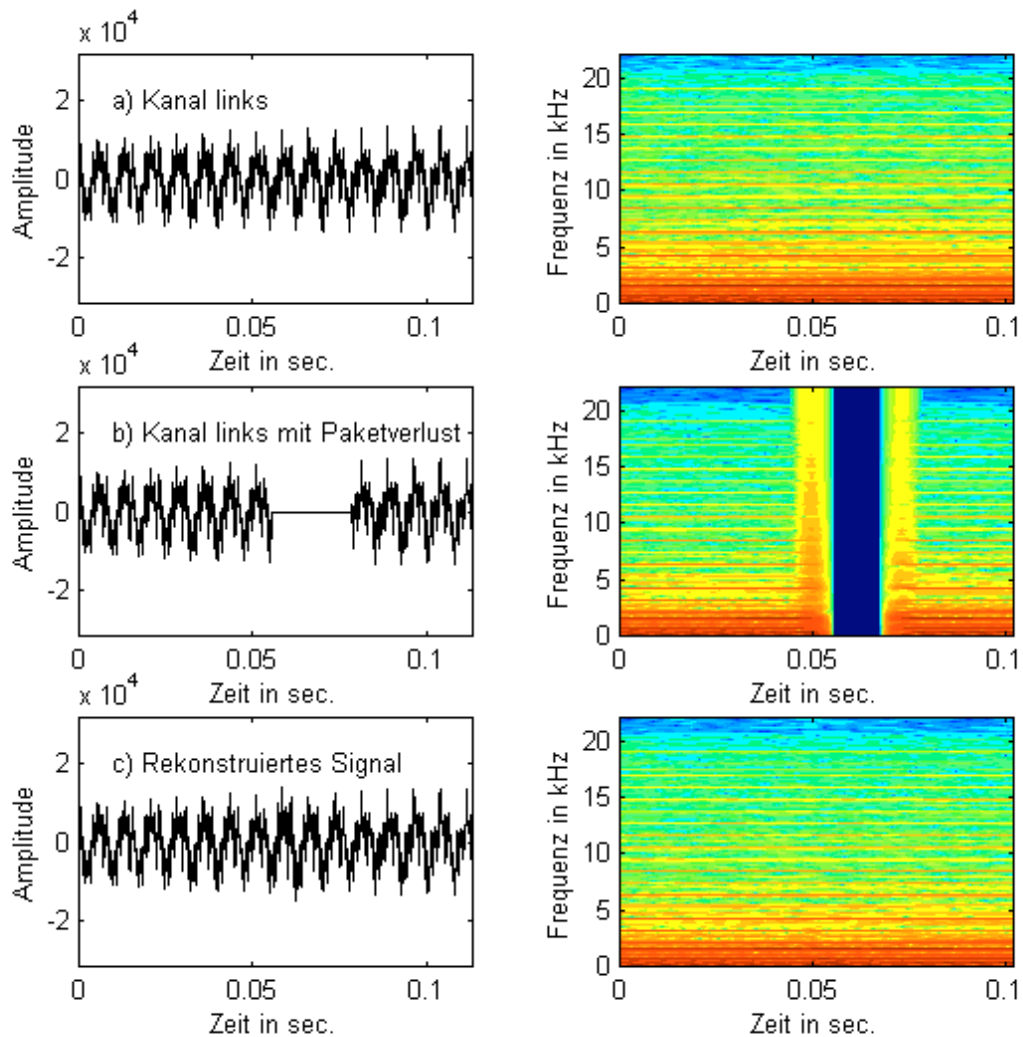


Bild 4.22: Paketsubstitution mit LPC-GI; Signalverläufe links und die dazugehörigen Spektrogramme rechts. a) Originalsignal aus einem Einzelinstrument-Titel (Orgel), hier handelt es sich um den linken Kanal des Stereosignals (5000 AW: 113ms); b) Linker Kanal mit Paketverlust. Die Paketlänge beträgt hierbei 1024 Abtastwerte (22 ms); c) Mit LPC-GI rekonstruiertes Signal.

Bild 4.23 zeigt die Signalverläufe und die dazugehörigen Spektrogramme des linken Kanals eines Audiosignals aus dem Beethoven-Titel (a), des linken Kanals mit Paketverlusten (22 ms Lücke) und des mit dem LPC-GI-Verfahren rekonstruiertes Audiosignals. Das Originalsignal hat keine besonderen Merkmale. Es ist deswegen schwierig, aus den korrekt empfangenen Paketen einen Ersatz für den Paketverlust zu schätzen. Ähnlich dem Verfahren der Maximalen Korrelation extrahiert das LPC-GI Verfahren aus den zur Verfügung stehenden Daten ein Substitutionssignal, das mit dem fehlenden Signal der Lücke am ähnlichsten sein sollte. Für die Rekonstruktion des zweiten Paketverlustes aus Bild 4.23-b werden die gepufferten Innovationen vor dem Verlust analysiert. Diese enthalten in diesem Fall ein bereits rekonstruiertes Audiosignal (Substitutionssignal für den ersten Paketverlust aus Bild 4.23-b). Die Qualität des Substitutionssignal für den zweiten Paketverlust wird von der

Qualität der Substitution des ersten Substitutionssignals beeinflusst. Dies erklärt wieso in diesem Fall das rekonstruierte Signal eine ganz andere Statistik als das Originalsignal aufweist. Das Verfahren ist sehr empfindlich gegenüber dicht aufeinander folgenden Paketverlusten.

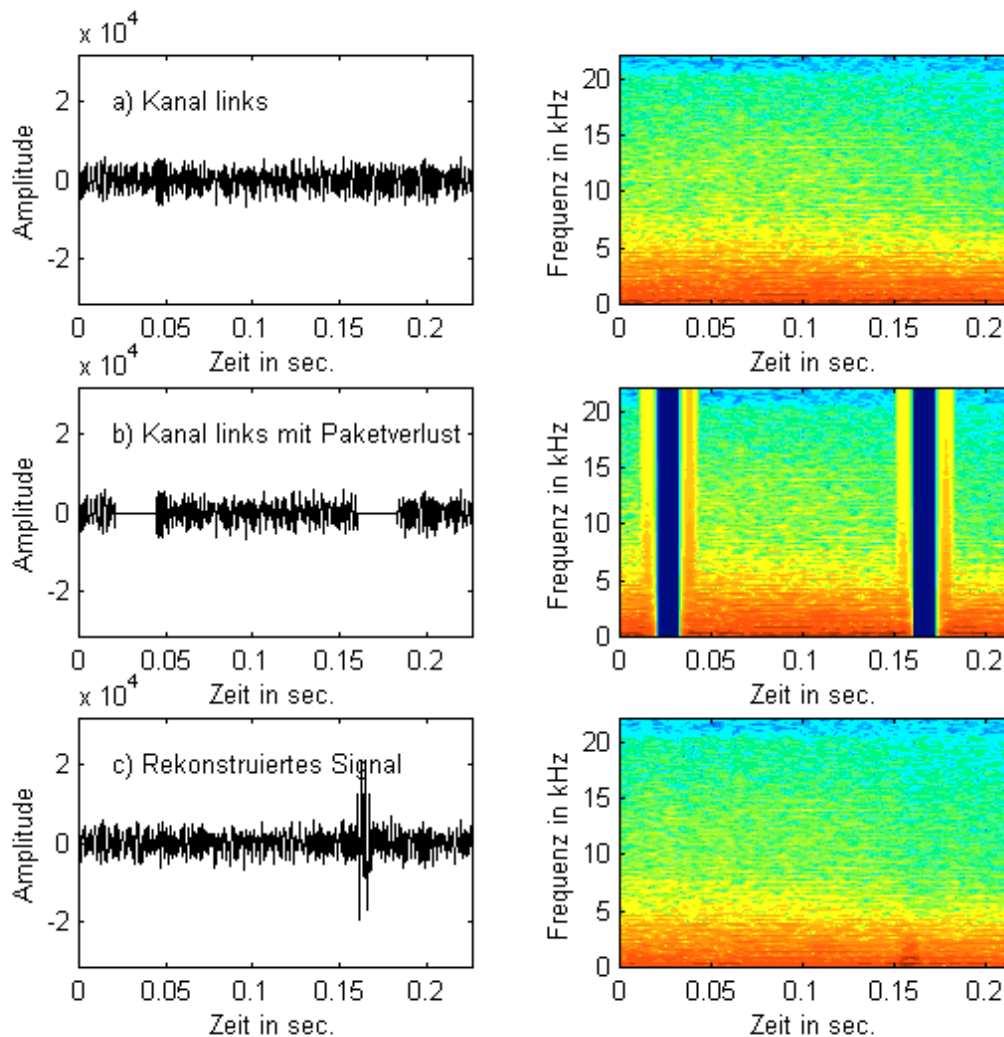


Bild 4.23: Paketsubstitution mit LPC-GI; Signalverläufe links und die dazugehörigen Spektrogramme rechts. a) Originalsignal aus einem Klassische-Musik-Titel (Beethoven), hier handelt es sich um den linken Kanal des Stereosignals (10000 AW: 227ms); b) Linker Kanal mit Paketverlusten. Die Paketlänge beträgt hierbei 1024 Abtastwerte (22 ms); c) Mit LPC-GI rekonstruiertes Signal.

Wahl der Filterkoeffizienten für die Substitution

Mit dem LPC-GI Substitutionsverfahren wird im Fall eines Paketverlustes eine optimale Innovation nach der Methode der maximalen Kreuzkorrelation geschätzt und synthetisiert. Es wird nach Bedarf eine Interpolation durchgeführt.

Eine Untersuchung des Verfahrens anhand eines einfachen Sinussignals hat gezeigt, dass die Substitution eines fehlenden Signalausschnittes zwar zumeist phasen- und amplitudenrichtig geschieht, aber an manchen Stellen eine leichte bis starke Erhöhung oder Verringerung der Amplitudenwerte bewirkt. Eine genauere Untersuchung des Signals und der Filterkoeffizienten hat gezeigt, dass die Koeffizienten des zuletzt korrekt empfangenen Paketes manchmal nicht für die Substitution geeignet sind. Aus diesem Grund wurde das Verfahren so optimiert, dass abhängig von der Lage der Innovation im Puffer die dazu passenden Filterkoeffizienten gewählt wurden. Die Wahl der für die Synthese optimalen Koeffizienten geschieht folgendermaßen:

Schritt1

Zuerst werden die Koeffizienten der korrekt empfangenen Pakete fortlaufend berechnet und gepuffert. Die Pufferung erfolgt in jedem Kanal für die letzten drei Koeffizientensätze. Es entstehen somit zwei Matrizen mit der Dimension $3 \times k$ ¹, wobei k die Filterordnung bezeichnet.

Die Matrizen *KozMatrixLinks* der gepufferten Koeffizienten des linken Kanals und *KozMatrixRechts* des rechten Kanals sind wie folgt aufgebaut:

$$KozMatrixLinks = \begin{pmatrix} La_1^n & La_2^n & \dots & La_k^n \\ La_1^{n-1} & La_2^{n-1} & \dots & La_k^{n-1} \\ La_1^{n-2} & La_2^{n-2} & \dots & La_k^{n-2} \end{pmatrix}$$

und

$$KozMatrixRechts = \begin{pmatrix} Ra_1^n & Ra_2^n & \dots & Ra_k^n \\ Ra_1^{n-1} & Ra_2^{n-1} & \dots & Ra_k^{n-1} \\ Ra_1^{n-2} & Ra_2^{n-2} & \dots & Ra_k^{n-2} \end{pmatrix}$$

La_i^j : Filterkoeffizienten für Paket j im linken Kanal ($i=1\dots k$).

¹ Es werden nur die letzten drei Koeffizientensätze gepuffert, weil die bei der Simulation verwendeten Länge l_1 des Innovationspuffers $3 \cdot (L + \Delta L)$ beträgt. Nimmt man für l_1 beispielsweise $5 \cdot (L + \Delta L)$, so bekommt die Matrix die Dimension $5 \times k$.

Ra_i^j : Filterkoeffizienten für Paket j im rechten Kanal ($i=1\dots k$).

Beispiel

Besteht beispielsweise ein Paketverlust nur im linken Kanal an der Stelle n (n : Paketnummer) so besteht die Matrix *KozMatrixLinks* aus den folgenden Koeffizientensätzen:

$La_1^{n-1} \dots La_k^{n-1}$: Filterkoeffizienten für Paket $n-1$ im linken Kanal,
 $La_1^{n-2} \dots La_k^{n-2}$: Filterkoeffizienten für Paket $n-2$ im linken Kanal und
 $La_1^{n-3} \dots La_k^{n-3}$: Filterkoeffizienten für Paket $n-3$ im linken Kanal.

Da der rechte Kanal fehlerfrei ist, werden auch in diesem Kanal die Koeffizienten des n -ten Paketes berechnet. Somit bekommen die Vektoren der Matrix *KozMatrixRechts* folgende Bedeutung:

$Ra_1^n \dots Ra_k^n$: Filterkoeffizienten für Paket n im rechten Kanal,
 $Ra_1^{n-1} \dots Ra_k^{n-1}$: Filterkoeffizienten für Paket $n-1$ im rechten Kanal und
 $Ra_1^{n-2} \dots Ra_k^{n-2}$: Filterkoeffizienten für Paket $n-2$ im rechten Kanal.

Schritt 2

Im zweiten Schritt wird mit dem MK-Verfahren die optimale Position (i_{\max}) des lokalen Maximums in beiden KKF-Vektoren KKF1 und KKF2 untersucht (Bild 4.10). Man unterscheidet drei Fälle:

- $$l_2 - \frac{L + \Delta L}{2} < i_{\max} < l_2$$

Dies hat für die Grenze (Ende) der Innovation im Puffer folgende Bedeutung:

$$l_1 - \frac{L + \Delta L}{2} \leq i_{\max} + t + L + \Delta L < l_1 \quad (4.17)$$

Die optimale Innovation befindet sich somit in Bereich 1 (s. Bild 4.24). In diesem Fall werden die Koeffizienten der ersten Matrix-Zeile entnommen.

- $l_2 - 3 \cdot \frac{L + \Delta L}{2} < i_{\max} < l_2 - \frac{L + \Delta L}{2}$

Die Innovation wird aus Bereich 2 kopiert. Die Koeffizienten werden der zweiten Matrix-Zeile entnommen.

- $0 < i_{\max} < l_2 - 3 \cdot \frac{L + \Delta L}{2}$

Die Innovation wird aus Bereich 3 kopiert. Die Koeffizienten des Synthesefilters werden der letzten Matrix-Zeile entnommen.

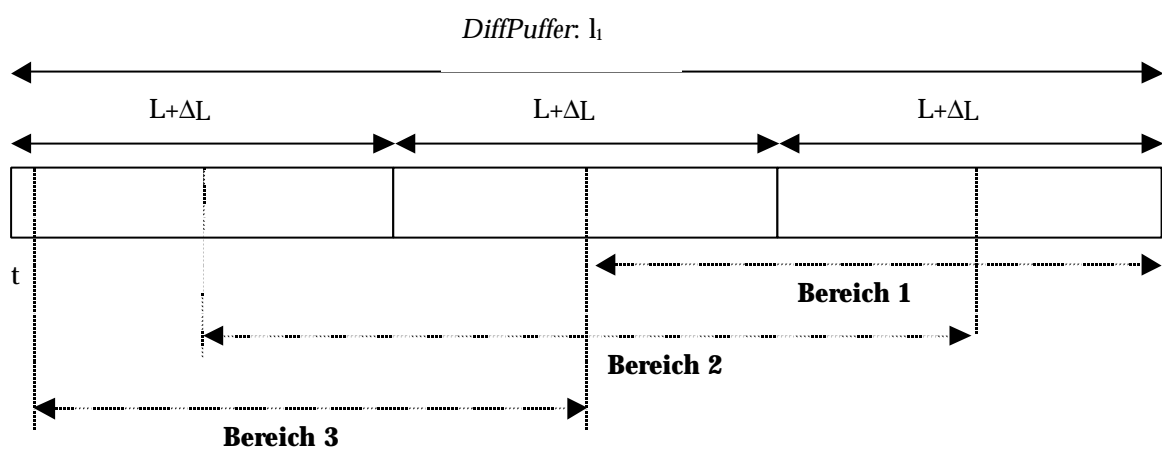


Bild 4.24: Bereichsunterteilung für die Bestimmung der Filterkoeffizienten in Abhängigkeit von der mit dem Verfahren der maximalen Kreuzkorrelation ermittelten optimalen Position der Innovation im Puffer.

Interpolation oder Extrapolation?

Eine ausreichende Verdeckung der Paketverluste erfordert in der Regel eine Analyse des korrekt empfangenen Audiosignals vor und gegebenenfalls auch nach dem verlorenen Paket.

Das oben beschriebene Substitutionsverfahren zieht zur Signalrekonstruktion zwei Pakete vor und nach der Lücke heran (Interpolation). Dies ist von Vorteil, wenn sich die Statistik des Audiosignals während des Verlustes ändert, denn dann würden Extrapolationsverfahren nicht mehr ausreichen, um diese Änderung zu erfassen.

Sollte sich die Signalstatistik aber nicht ändern, so würde in diesem Fall ein Interpolationsverfahren nur zusätzliche algorithmische Verzögerungen um die Dauer eines Paketes beanspruchen, da zur Bestimmung des Substitutionssignals das Einlaufen des auf den Verlust folgenden Paketes in den Decodierer abgewartet werden muss.

Aufgrund der erforderlichen zusätzlichen Signalverzögerung sind Verfahren mit Interpolation des Audiosignals in der Regel lediglich zur Rekonstruktion einzelner oder weniger aufeinanderfolgender Paketverluste geeignet. Treten Bündelverluste auf, so kann eine Interpolation zumeist nicht mehr durchgeführt werden. Hier wäre ein Umschalten auf eine Extrapolation sinnvoll.

Sinnvoll ist es, eine Interpolation nur dann anzusetzen, wenn dies auch nötig ist. Dies erfordert eine Untersuchung des Audiosignals vor und nach der Lücke.

Untersucht werden folgende Eigenschaften des Audiosignals:

- **Nulldurchgangsrate:**

Es werden die Anzahl der Nulldurchgänge des Audiosignals vor (*zcrLast*) und nach (*zcrNext*) der Lücke untersucht. Eine bemerkbare Erhöhung oder Verringerung des *zcr*-Wertes deutet auf einer Änderung der Frequenz des Audiosignals hin. In diesem Fall wird die Interpolation beiderseits der Lücke durchgeführt.

Sollte sich der *zcr*-Wert nur geringfügig ändern, so wird nur eine Extrapolation durchgeführt. Eine Interpolation wird nur dann vorgenommen, wenn das Verhältnis der *zcr*-Werte vor der Lücke und nachher mindestens drei beträgt und somit folgende Bedingung erfüllt:

$$\max \left\{ \frac{zcrNext}{zcrLast}, \frac{zcrLast}{zcrNext} \right\} \geq 3$$

- **Signalenergien:**

Die Energie der Signalausschnitte unmittelbar vor und nach der Lücke wird ermittelt. Eine bemerkbare Änderung des Energie-Wertes deutet auf einer Veränderung der Signal-Statistik hin, was zur Notwendigkeit einer Interpolation im Fall eines Paketverlustes führt.

- **Teilenergien**

Pro Paket werden vier Energie-Werte ermittelt (E1-E4), die jeweils über ein Viertel der Paketlänge berechnet werden (s. Bild 4.25).

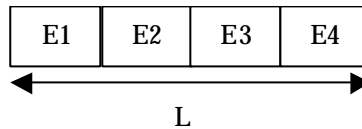


Bild 4.25: Teil-Energien eines Paketes.

Eine fortlaufende Erhöhung der Energie-Werte könnte auf das Einsetzen einer Attacke hinweisen (s. Kapitel 2.3.4). In diesem Fall wird eine Interpolation bevorzugt. Zur Veranschaulichung wird der gesamte Algorithmus zur Verdeckung der Lücke mit dem LPC-GI-Verfahren in dem folgenden Blockschaltbild zusammengefasst.

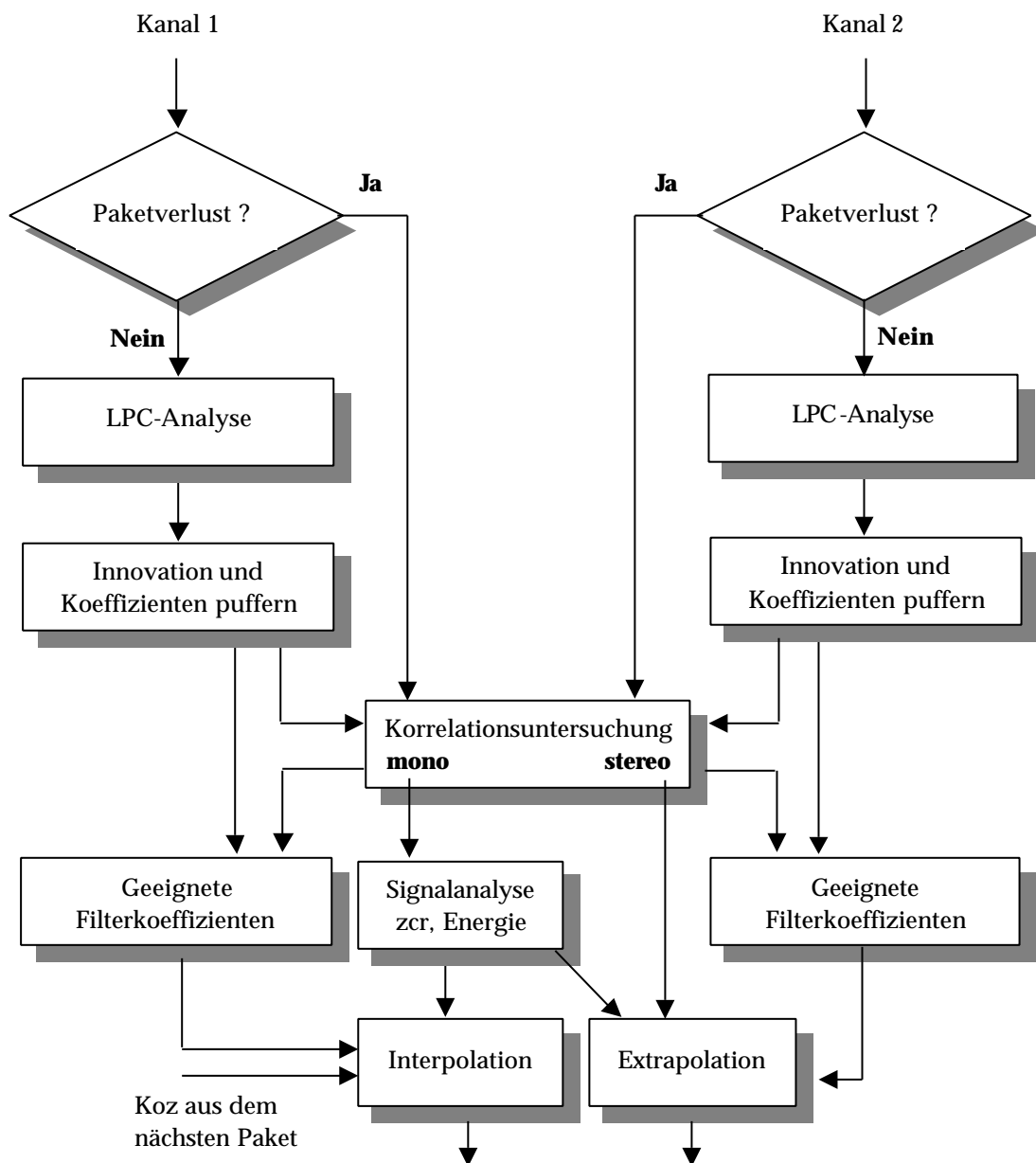


Bild 4.26: Prinzipschaltbild des LPC-GI-Substitutionsverfahrens

Signalsubstitution bei Bündelfehlern

Bündelfehler werden wie Einzelfehler behandelt, jedoch ist hier für die Rekonstruktion der durch den Verlust des ersten Paketes des Bündels entstandenen Lücke keine Interpolation mehr möglich. Aufgrund der Empfindlichkeit des Verfahrens gegenüber dicht aufeinanderfolgenden Paketverlusten (siehe dazu das Beispiel aus Bild 4.23-c) nimmt die Qualität des rekonstruierten Audiosignals bei Bündelfehlern schnell ab.

Es ist insgesamt angebrachter, die Lücke eines Bündels als eine große Lücke zu betrachten und sie dann wie einen Einzelpaketverlust zu behandeln. Dies ist leider mit diesem Verfahren nicht möglich, da es von einer konstanten Paketlänge ausgeht und für diese Länge fortlaufend Filterkoeffizienten und Innovationen im fehlerfreien Fall berechnet und puffert.

4.5 Stereoprädiktion zur Signalsubstitution bei Paketverlusten

Die Stereoprädiktion (SLPC: *Stereo Linear Predictive Coding*) stellt eine Methode zur Ausnutzung der Interkanal-Redundanz von Stereosignalen dar. Eine Stereoprädiktion wird beispielsweise bei der Datenkompression von Audiosignalen verwendet (MPEG, *Joint Stereo Audio Coding* [Fuc93]).

Die Stereoprädiktion betrachtet die statistischen Abhängigkeiten (Korrelationen) zwischen zwei zeitgleichen Ausschnitten (Segmenten) aus dem rechten und linken Kanal. Der bei der Stereoprädiktion erreichbare Prädiktionsgewinn wurde mit dem Gewinn konventioneller linearer Prädiktion (Monoprädiktion) verglichen [Cam93]. Die Untersuchungsergebnisse haben gezeigt, dass mit der Stereoprädiktion ein höherer Prädiktionsgewinn (und somit Codinggewinn) als mit der Monoprädiktion erreicht werden kann [Cam93].

In dieser Arbeit wird untersucht, ob mit dem Einsetzen der Stereoprädiktion für die Rekonstruktion von Lücken in Audiosignalen eine bessere Audioqualität von Substitutionssignalen erreicht werden können.

4.5.1 Synthese eines Substitutionssignals aus Innovation

In diesem Abschnitt wird erläutert, wie die Stereoprädiktion für die Synthese von Substitutionssignalen bei Paketverlusten verwendet werden kann (s. Bild 4.27). Die Stereoprädiktion nutzt zusätzlich zu den Autokorrelationen der Abtastwerte innerhalb eines Kanals auch die Kreuzkorrelation zwischen den Abtastwerten des linken Kanals und rechten Kanals aus. Die Stereoprädiktion wird hier direkt auf Stereo-PCM-Signale angewendet.

Nach der Übertragung und dem empfängerseitigen Decodieren der Stereodaten werden diese in Segmenten festgelegter Länge gepuffert. Mit der SLPC (Stereo LPC)-Analyse werden im fehlerfreien Fall aus diesen Daten Filterkoeffizienten¹ adaptiv berechnet (s. Kap 4.5.2). Diese werden zur Adaption von zwei zueinander inversen Filtern (dem SLPC-Fehlerfilter und dem SLPC-Synthesefilter) herangezogen. Im SLPC-Fehlerfilter werden Schätzsignale für den linken und den rechten Kanal berechnet und anschließend von den Eingangssignalen subtrahiert (Gl. 4.20). Die Innovationen (Differenzsignale) beider Kanäle werden im fehlerfreien Fall gepuffert. Die Eingangssignale $L(n)$ und $R(n)$ werden dann nach der Bestimmung der Filterkoeffizienten zum Ausgang weitergeführt. In diesem Fall gilt $L_{\text{Aus}}(n) = L(n)$ und $R_{\text{Aus}}(n) = R(n)$.

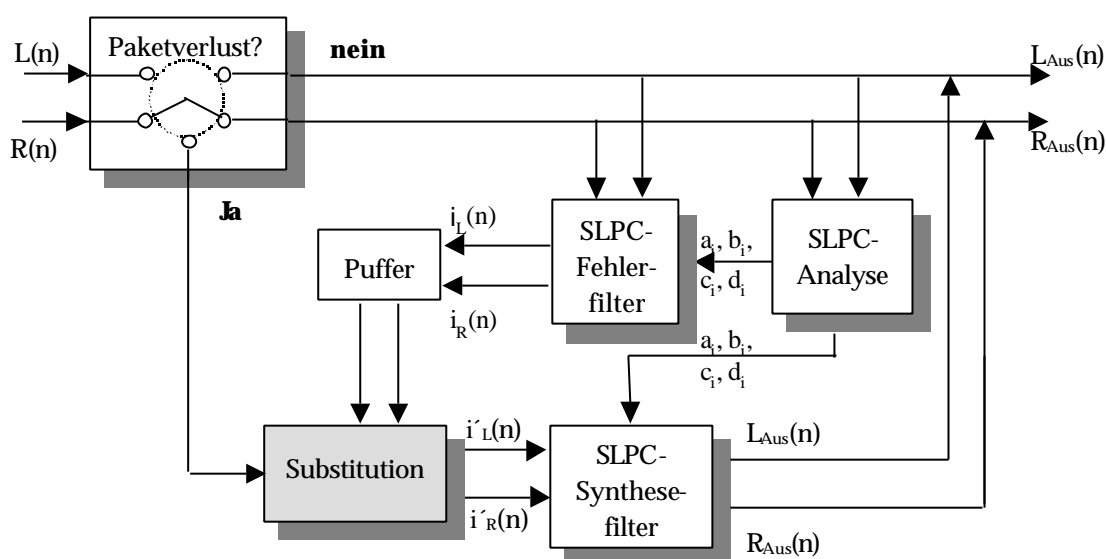


Bild 4.27: Paketsubstitution mit der Stereoprädiktion.

Im Fall eines Paketverlustes in einem Kanal werden keine neuen Koeffizienten berechnet. Die zuletzt berechneten Koeffizienten des SLPC-Synthesefilters werden beibehalten. Am Eingang dieses Filters liegen in diesem Fall die Innovationen $i'_L(n)$ und $i'_R(n)$ vor.

Analog zu der Substitution auf der Ebene der LPC-Innovation können hier auch die Innovationen am Eingang des Synthesefilters aus Nullwerten, aus gleichverteiltem Rauschen mit der Varianz der letzten gepufferten Innovation oder aus der letzten gepufferten Innovationen der zuletzt korrekt empfangenen Paketen bestehen. Ein direkter Links-/Rechts-Ersatz der Innovation ist mit diesem Substitutionsverfahren (auch bei vorhandenem Paket aus dem Nachbarkanal) nicht möglich, da weder Filterkoeffizienten noch Schätz- oder Innovationen im Fall eines Paketverlustes in einem der Kanäle berechnet werden können.

¹ Mit der Stereoprädiktion werden aus beiden Kanälen insgesamt vier Koeffizientensätze a_i , b_i , c_i und d_i berechnet.

4.5.2 Bestimmung der optimalen Filterkoeffizienten eines Stereoprädiktors

Ähnlich wie die Monoprädiktion lässt sich die Stereoprädiktion in fester bzw. adaptiver Form anwenden. Im Gegensatz zu der festen Prädiktion, bei der die Filterkoeffizienten des Stereoprädiktors fest angegeben sind, ändern sich die Koeffizienten des adaptiven Prädiktors blockweise. Die Koeffizienten werden in jedem Zyklus für zwei neue Segmente des Stereosignals (ein Segment aus jedem Kanal) neu berechnet und somit an die aktuellen Eigenschaften des Audiosignals angepasst. Die Berechnung der adaptiven Filterkoeffizienten des Stereoprädiktors ist komplexer als die des Monoprädiktors. Im diesem Abschnitt wird die Theorie zur Berechnung der Koeffizienten eines adaptiven Stereoprädiktors vorgestellt.

Bei der Stereoprädiktion werden die Prädiktorstruktur und das Adaptionsverfahren im Vergleich zum Monoprädiktor so erweitert, dass zusätzlich zur Autokorrelation auch die Kreuzkorrelation zwischen den korrespondierenden Teilsignalen (Segmente) des linken und rechten Kanals eines Stereosignals ausgenutzt wird.

Die Stereoprädiktion berechnet für jeden Wert $L(n)$ des linken Kanals (bzw. $R(n)$ des rechten Kanals) einen Schätzwert $\hat{L}(n)$ (bzw. $\hat{R}(n)$) aus vorangegangenen Signalwerten $L(n-k)$ (bzw. $R(n-k)$) der beiden Kanäle.

$$\hat{L}(n) = \sum_{k=1}^{k_a} a_k L(n-k) + \sum_{k=1}^{k_b} b_k R(n-k). \quad (4.18)$$

$$\hat{R}(n) = \sum_{k=1}^{k_c} c_k R(n-k) + \sum_{k=0}^{k_d} d_k L(n-k). \quad (4.19)$$

- k_a : Filterordnung des Autoprädiktors des linken Kanals
- k_b : Filterordnung des Kreuzprädiktors des linken Kanals
- k_c : Filterordnung des Autoprädiktors des rechten Kanals
- k_d : Filterordnung des Kreuzprädiktors des rechten Kanals
- a_k : Filterkoeffizienten des Autoprädiktors des linken Kanals
- b_k : Filterkoeffizienten des Kreuzprädiktors des linken Kanals
- c_k : Filterkoeffizienten des Autoprädiktors des rechten Kanals
- d_k : Filterkoeffizienten des Kreuzprädiktors des rechten Kanals

Jeder Schätzwert lässt sich als Summe von einem "Autoprädiktions"- und einem "Kreuzprädiktions"-Teil darstellen. So liefert z.B. der Autoprädiktor des linken Kanals eine Summe aus den vorangegangenen k_a Werten desselben Kanals, gewichtet mit den adaptiv berechneten Koeffizienten a_k . Der Kreuzprädiktor des linken Kanals liefert hingegen eine Summe aus den

vorangegangenen k_b Werten des Nachbarkanals, gewichtet mit den Koeffizienten b_k (s. Bild 4.28).

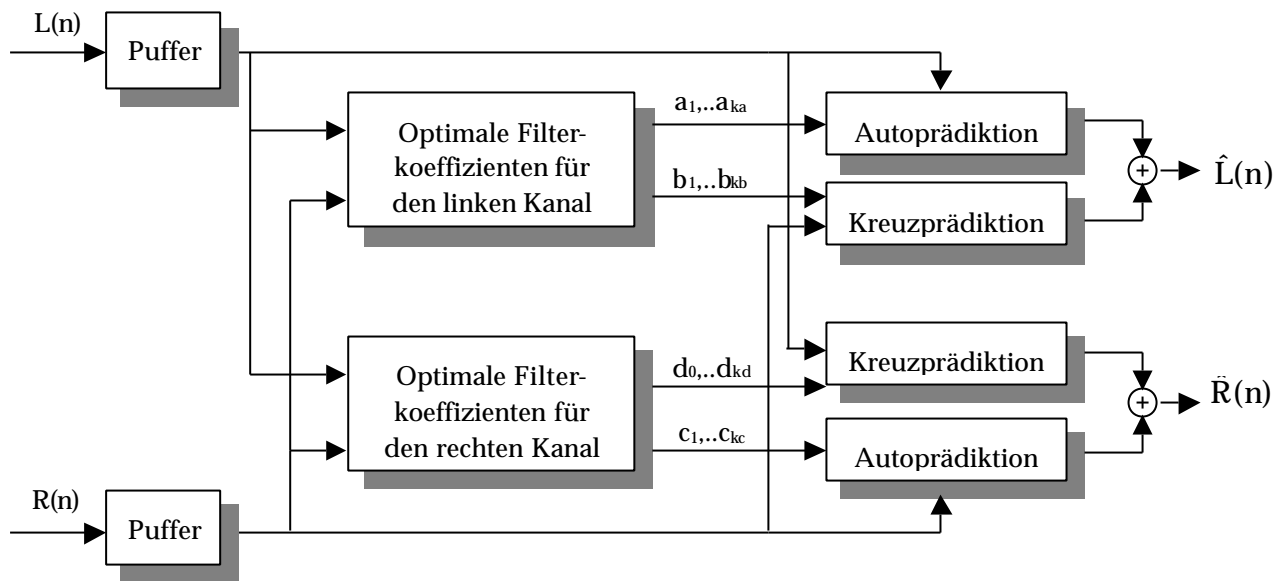


Bild 4.28: Das Prinzip der Stereoprädiktion (SLPC). Berechnung von vier Koeffizientensätze und Dekorrelation von beiden Kanälen des Stereosignals mittels Auto- bzw. Kreuzprädiktion.

Ziel der adaptiven Prädiktion ist es, eine möglichst gute Signalschätzung zu erhalten. Das wird dann erreicht, wenn die Differenz zwischen Original- und Schätzsignal minimiert wird. Für eine optimale Vorhersage müssen die Prädiktoren optimiert werden. Dazu wird der mittlere quadratische Fehler zwischen Original- und Schätzsignal und damit die Varianz des Differenzsignals minimiert.

Für das Differenzsignal des linken Kanals gilt:

$$D_L(n) = L(n) - \hat{L}(n) = L(n) - \sum_{k=1}^{k_a} a_k L(n-k) - \sum_{k=1}^{k_b} b_k R(n-k), \quad n \in \{0..N-1\}. \quad (4.20)$$

N ¹: Paketlänge

$L(n)$: Originalsignal aus dem linken Kanal

$\hat{L}(n)$: Schätzsignal für den linken Kanal

Das Differenzsignal $D_L(n)$ des linken Kanals hat dann die folgende Varianz:

$$\sigma_{D_L}^2 = \frac{1}{N} \cdot \sum_{n=0}^{N-1} [L(n) - \hat{L}(n)]^2. \quad (4.21)$$

¹ In diesem Kapitel Stereoprädiktion wird die Paketlänge mit N statt L bezeichnet, um Verwechslung mit der Bezeichnung des linken Kanals $L(n)$ zu vermeiden.

Eine Minimierung des Differenzsignals bzw. dessen Varianz bezüglich der Koeffizienten a_k bedeutet:

$$\begin{aligned}
\frac{\partial \sigma_{D_L}^2}{\partial a_k} = 0 &= \frac{\partial}{\partial a_k} \left\{ \sum_{n=0}^{N-1} [L(n) - \hat{L}(n)]^2 \right\} \\
&= \sum_{n=0}^{N-1} \left[2 \cdot (L(n) - \hat{L}(n)) \cdot \frac{\partial}{\partial a_k} (-\hat{L}(n)) \right] \\
&= -2 \cdot \sum_{n=0}^{N-1} (L(n) - \hat{L}(n)) \cdot \frac{\partial}{\partial a_k} \left(\sum_{k=1}^{k_a} a_k L(n-k) + \sum_{k=1}^{k_b} b_k R(n-k) \right) \\
&= -2 \cdot \sum_{n=0}^{N-1} (L(n) - \hat{L}(n)) \cdot L(n-k) \quad \text{für } k = 1, \dots, k_a
\end{aligned} \tag{4.22}$$

Für $k = j$ bedeutet das:

$$\begin{aligned}
\sum_{n=0}^{N-1} L(n-j) \cdot L(n) &= \sum_{n=0}^{N-1} \hat{L}(n) \cdot L(n-j) \\
&= \sum_{n=0}^{N-1} \left(\sum_{k=1}^{k_a} a_k L(n-k) + \sum_{k=1}^{k_b} b_k R(n-k) \right) \cdot L(n-j) \\
&= \sum_{n=0}^{N-1} \left(L(n-j) \cdot \sum_{k=1}^{k_a} a_k L(n-k) \right) + \sum_{n=0}^{N-1} \left(L(n-j) \cdot \sum_{k=1}^{k_b} b_k R(n-k) \right)
\end{aligned} \tag{4.23}$$

Im Allgemeinen gilt für eine Korrelationsfunktion:

$$R_{pq}(i, j) = \sum_{n=0}^{N-1} p(n-i)q(n-j) \tag{4.24}$$

Also liefert Gleichung 4.23 folgendes Ergebnis:

$$R_{LL}(j, 0) = \sum_{k=1}^{k_a} a_k R_{LL}(j, k) + \sum_{k=1}^{k_b} b_k R_{LR}(j, k) \quad \text{für } j = 1, \dots, k_a. \tag{4.25}$$

Analog zu der oben ausgeführten Berechnung führt eine Minimierung bezüglich der Koeffizienten b_k zu der folgenden Gleichung:

$$R_{LR}(0, j) = \sum_{k=1}^{k_a} a_k R_{LR}(k, j) + \sum_{k=1}^{k_b} b_k R_{RR}(k, j) \quad \text{für } j = 1, \dots, k_b. \tag{4.26}$$

Beide Gleichungen 4.25 und 4.26 lassen sich in dem folgenden Gleichungssystem zusammenfassen:

$$\begin{pmatrix} \mathbf{R}_{LL}(1,0) \\ \vdots \\ \mathbf{R}_{LL}(k_a,0) \\ \hline \mathbf{R}_{LR}(0,1) \\ \vdots \\ \mathbf{R}_{LR}(0,k_b) \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{LL}(1,1) & \cdots & \mathbf{R}_{LL}(1,k_a) & | & \mathbf{R}_{LR}(1,1) & \cdots & \mathbf{R}_{LR}(1,k_b) \\ \vdots & \vdots & \vdots & | & \vdots & \vdots & \vdots \\ \mathbf{R}_{LL}(k_a,1) & \cdots & \mathbf{R}_{LL}(k_a,k_a) & | & \mathbf{R}_{LR}(k_a,1) & \cdots & \mathbf{R}_{LR}(k_a,k_b) \\ \hline \mathbf{R}_{LR}(1,1) & \cdots & \mathbf{R}_{LR}(k_a,1) & | & \mathbf{R}_{RR}(1,1) & \cdots & \mathbf{R}_{RR}(k_b,1) \\ \vdots & \cdots & \vdots & | & \vdots & \vdots & \vdots \\ \mathbf{R}_{LR}(1,k_b) & \cdots & \mathbf{R}_{LR}(k_a,k_b) & | & \mathbf{R}_{RR}(1,k_b) & \cdots & \mathbf{R}_{RR}(k_b,k_b) \end{pmatrix} \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_{k_a} \\ \hline \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_{k_b} \end{pmatrix} \quad (4.27)$$

In diesem Gleichungssystem sind $\mathbf{R}_{LL}(i,j)$ und $\mathbf{R}_{RR}(i,j)$ Autokorrelationswerte und $\mathbf{R}_{LR}(i,j)$ Kreuzkorrelationswerte. Die Stereo-Kovarianzmatrix aus Gleichung 4.27 hat die Dimension $(k_a+k_b) \times (k_a+k_b)$. Für die Ermittlung der Filterkoeffizienten muss das Gleichungssystem aus Gleichung 4.27 nach den Koeffizienten \mathbf{a} und \mathbf{b} gelöst werden. Die Lösung dieses Gleichungssystems erfolgt mittels *LU-Decomposition* [Pre96].

Für die Berechnung der Koeffizienten c_k und d_k wird die Varianz des Differenzsignals des rechten Kanals minimiert. Es ergeben sich folgende Gleichungen 4.28 und 4.29, die in dem Gleichungssystem aus 4.30 zusammengefasst werden.

$$\mathbf{R}_{RR}(j,0) = \sum_{k=1}^{k_c} c_k \mathbf{R}_{RR}(j,k) + \sum_{k=0}^{k_d} d_k \mathbf{R}_{RL}(j,k). \quad (4.28)$$

$$\mathbf{R}_{RL}(0,j) = \sum_{k=1}^{k_c} c_k \mathbf{R}_{RL}(k,j) + \sum_{k=0}^{k_d} d_k \mathbf{R}_{LL}(k,j). \quad (4.29)$$

$$\begin{pmatrix} \mathbf{R}_{RR}(1,0) \\ \vdots \\ \mathbf{R}_{RR}(k_c,0) \\ \hline \mathbf{R}_{RL}(0,0) \\ \vdots \\ \mathbf{R}_{RL}(0,k_d) \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{RR}(1,1) & \cdots & \mathbf{R}_{RR}(1,k_c) & | & \mathbf{R}_{RL}(1,0) & \cdots & \mathbf{R}_{RL}(1,k_d) \\ \vdots & \vdots & \vdots & | & \vdots & \vdots & \vdots \\ \mathbf{R}_{RR}(k_c,1) & \cdots & \mathbf{R}_{RR}(k_c,k_c) & | & \mathbf{R}_{RL}(k_c,0) & \cdots & \mathbf{R}_{RL}(k_c,k_d) \\ \hline \mathbf{R}_{RL}(1,0) & \cdots & \mathbf{R}_{RL}(k_c,0) & | & \mathbf{R}_{LL}(0,0) & \cdots & \mathbf{R}_{LL}(k_d,0) \\ \vdots & \vdots & \vdots & | & \vdots & \vdots & \vdots \\ \mathbf{R}_{RL}(1,k_d) & \cdots & \mathbf{R}_{RL}(k_c,k_d) & | & \mathbf{R}_{LL}(0,k_d) & \cdots & \mathbf{R}_{LL}(k_d,k_d) \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_{k_c} \\ \hline d_0 \\ \vdots \\ d_{k_d} \end{pmatrix} \quad (4.30)$$

Diese Stereo-Kovarianzmatrix aus Gl. 4.30 hat die Dimension $(k_c+k_d+1) \times (k_c+k_d+1)$.

4.5.3 Instabilitätsproblem des SLPC-Synthesefilters

Eine erste Untersuchung der Paketsubstitution mit der Stereoprädiktion zeigt, dass das SLPC-Synthesefilter oft instabil wird, was eine drastische Verschlechterung der Audioqualität bewirkt. Die Stereoprädiktion stellt nicht sicher, dass das SLPC-Synthesefilter immer stabil bleibt. In Bild 4.29 ist ein Beispiel für eine Paketsubstitution ohne (Paket 1) und mit (Paket 2) Instabilität des SLPC-Synthesefilters dargestellt. Das hierbei verwendete Substitutionsverfahren ist die Nullsubstitution auf der Ebene der Innovation nach einer Stereoprädiktion (SLPC-Null).

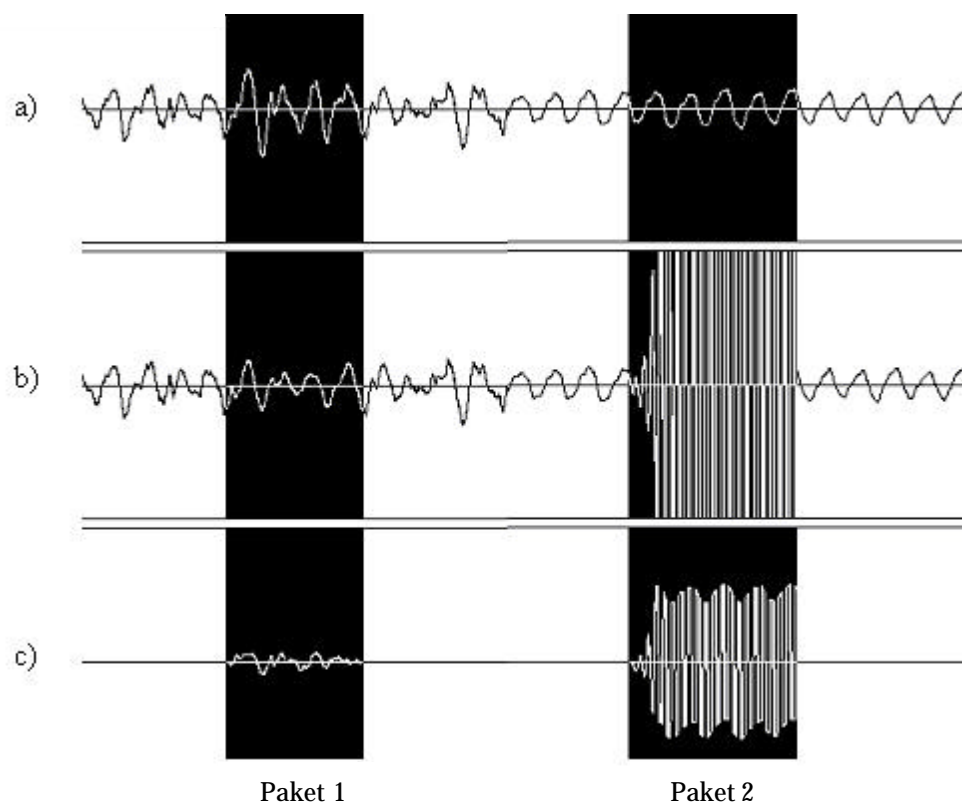


Bild 4.29: Instabilitätsproblem bei der Stereoprädiktion. a) Originalsignal, b) Signal nach der Substitution enthält zwei substituierte Pakete. Das hierbei verwendete Substitutionsverfahren ist die Nullsubstitution der Innovation und anschließende Synthese des Substitutionssignals mit dem SLPC-Synthesefilter. Das Synthesefilter ist bei Paket 1 stabil und bei Paket 2 instabil. In c) ist die Differenz der Signale aus a) und b) dargestellt.

Aus Bild 4.29 wird ersichtlich, dass Paket 1 "optisch" gut verschleiert wurde. Das Verfahren erkennt die zwischen beiden Kanälen vorhandene Korrelation und nutzt sie zur Verdeckung der Lücke in einem Kanal aus. Dabei ist es wichtig, dass bei fehlendem Paket in einem Kanal das zeitgleiche Paket aus dem Nachbarkanal zur Verfügung steht.

Die Rekonstruktion von Paket 2 ist aufgrund der Instabilität des Synthesefilters an dieser Stelle fehlgeschlagen. Diese Instabilität bewirkt akustisch eine drastische Verschlechterung der wahrgenommenen Audioqualität. Im Fall der Instabilität des Synthesefilters sollen dann andere Maßnahmen ergriffen werden. Man kann das Problem beheben, indem man die Filterkoeffizienten, die hier für die Instabilität verantwortlich sind, so ändert, dass sich damit ein stabiles Synthesefilter ergibt. Als Abhilfe für die Stabilitätsuntersuchung dient die Berechnung der aus den optimalen Prädiktorkoeffizienten abgeleiteten Koeffizientensätze z.B. PARCOR- (*Partial Correlation*) oder auch Reflexionskoeffizienten genannt, LOG-Area oder LSP-Koeffizienten (Linien Spektrum Parameter).

4.5.3.1 Stabilitätskontrolle mit PARCOR-Koeffizienten

Von großer Bedeutung sind die PARCOR-Koeffizienten p . Für sie lässt sich ein zulässiger Wertebereich angeben, der ein notwendiges und hinreichendes Kriterium für die Stabilität eines rekursiven Filters darstellt. Für ein stabiles Filter gilt:

$$|p_i| \leq 1 \quad i = 1 \dots k, \quad (4.31)$$

wobei k die Ordnung des Prädiktorfilters entspricht.

In [Rab78] wurde gezeigt, wie sich die PARCOR-Koeffizienten p aus den Prädiktorkoeffizienten a_i rekursiv berechnen lassen:

$$p_i = a_i \quad i = k \dots 1. \quad (4.32)$$

$$a_j^{(i-1)} = \frac{a_j^{(i)} + a_i^{(i)} \cdot a_{i-j}^{(i)}}{1 - p_i^2}, \quad 1 \leq j \leq i-1; \quad k \leq i \leq 1 \quad (4.33)$$

Die tiefgestellten Indizes bedeuten den Verweis auf einen Feldindex. Die hochgestellten Indizes in Klammern sind die Rekursionsindizes.

Die Methode der Instabilitätskontrolle ist mit Hilfe der PARCOR-Koeffizienten p aus Gleichung 4.33 nur bei rein rekursiven Filtern anwendbar, wie es bei der linearen Prädiktion (LPC) der Fall ist. Es stellt sich also zuerst die Frage, ob das SLPC-Synthesefilter im Fall eines Paketverlustes rein rekursiv ist.

- **Beispiel eines alleinigen Paketverlustes im linken Kanal**

In diesem Beispiel wird davon ausgegangen, dass ein Paket aus dem linken Kanal fehlt. Das zeitgleiche Paket aus dem rechten Kanal ist vorhanden. Die Aufgabe des SLPC-

Synthesefilters besteht darin, ein Substitutionssignal $L_{\text{Aus}}(n)$ für den linken Kanal am Ausgang des SLPC-Synthesefilters zu schätzen:

$$L_{\text{Aus}}(n) = i'_L(n) + \sum_{k=1}^{k_a} a_k L_{\text{Aus}}(n-k) + \sum_{k=1}^{k_b} b_k R_{\text{Aus}}(n-k), \quad n \in \{0..N-1\}. \quad (4.34)$$

$R_{\text{Aus}}(n-k)$ entspricht in diesem Fall $R(n-k)$, da das Paket aus dem rechten Kanal vorhanden ist und deshalb zum Ausgang weitergeleitet wird. $i'_L(n)$ ist die Innovation.

Im Fall einer Nullsubstitution der Innovation des linken Kanals gilt:

$$i'_L(n) = 0, \quad n \in \{0..N-1\}. \quad (4.35)$$

Dies bedeutet für $L_{\text{Aus}}(n)$:

$$L_{\text{Aus}}(n) = \sum_{k=1}^{k_a} a_k L_{\text{Aus}}(n-k) + \sum_{k=1}^{k_b} b_k R(n-k), \quad n \in \{0..N-1\} \quad (4.36)$$

Es handelt sich hierbei nicht um ein rein rekursives Filter. Die Stabilität des Filters kann hier also nicht mit Hilfe von PARCOR-Koeffizienten überprüft werden.

In der Literatur existiert eine Reihe von Adaptionalgorithmen, welche die Instabilität von IIR-Filtern kontrollieren und gegebenenfalls aufheben [Camp00], [Shy89], [Reg95], [Rod90].

Effiziente Kriterien sind in der Lage, die Position der Polstellen der Übertragungsfunktion zu überwachen, wie z.B. das Jury-Marden-Kriterium [Ant93], das auf die explizite Berechnung der Polstellen verzichtet.

Methoden, die im Instabilitätsfall eine Pol-Projektion in den Einheitskreis durchführen ([Shy89], [Ant93]) sind komplexer aber meistens schneller und robuster als andere Verfahren [Camp00].

Um die Stabilitätskontrolle zu vereinfachen, kann das IIR-Filter in Kaskade, paralleler Form von Filtern zweiter Ordnung oder in Lattice-Form [Shy89] [Reg95] implementiert werden.

In dieser Arbeit wird die Stabilitätskontrolle des SLPC-Synthesefilters mit Hilfe einer Polstellen-Bestimmung aus den berechneten Filterkoeffizienten und im instabilen Fall mittels einer Polstellen-Projektion in den Einheitskreis durchgeführt.

4.5.3.2 Stabilitätskontrolle mit Polstellen-Untersuchung

Eine Möglichkeit der Stabilitätskontrolle von rekursiven Filtern besteht in der Berechnung und Untersuchung der Polstellen der Übertragungsfunktion des Synthesefilters. Stabile Filter erfüllen folgende Bedingung:

Ein IIR-Filter ist nur dann stabil, wenn alle Pole seiner Übertragungsfunktion in dem Einheitskreis der komplexen z -Ebene liegen.

In dem zeitvarianten Fall (bei nicht stationären Signalen) ist die oben genannte Bedingung zwar notwendig, aber nicht hinreichend. In diesem Fall soll noch die folgende Voraussetzung erfüllt werden.

Die Filterkoeffizienten des IIR-Filters dürfen sich nur langsam ändern (slow coefficient variation [Camp00]).

In der Praxis wird die zweite Bedingung oft vernachlässigt, wenn die erste erfüllt ist. Die zweite Bedingung kann bei instationären Signalen von großer Bedeutung sein, wenn das Filter nahe der Instabilitätsgrenze betrieben wird [Reg95].

In diesem Abschnitt soll gezeigt werden, wie ein instabiles SLPC-Synthesefilter zu einem stabilen Verhalten gezwungen wird, nachdem eine Manipulation an seinen Koeffizienten durchgeführt wird. Als erstes soll die Übertragungsfunktion des SLPC-Synthesefilters bestimmt werden, um anschließend die Polstellen berechnen zu können.

Das SLPC-Synthesefilter verwendet bei der Synthese die bei der Analyse berechneten Filterkoeffizienten. Im Fall eines Paketverlustes liegen am Eingang dieses Filters die Innovationen $i_L(n)$ und $i_R(n)$ und am Ausgang $L_{Aus}(n)$ und $R_{Aus}(n)$ vor (s. Bild 4.27). Die Synthese hängt davon ab, wo ein Paketverlust aufgetreten ist und was rekonstruiert werden soll.

Es werden insgesamt drei Fälle unterschieden:

- 1) Es tritt ein Paketverlust nur im linken Kanal auf.
- 2) Es tritt ein Paketverlust nur im rechten Kanal auf.
- 3) Es treten Paketverluste in beiden Kanälen auf.

1) Fall eines alleinigen Paketverlustes im linken Kanal

Allgemein gilt für die Synthese des linken Kanals:

$$L_{\text{Aus}}(n) = i'_L(n) + \sum_{k=1}^{k_a} a_k L_{\text{Aus}}(n-k) + \sum_{k=1}^{k_b} b_k R_{\text{Aus}}(n-k), \quad n \in \{0..N-1\}$$

$$\Leftrightarrow L_{\text{Aus}}(n) - \sum_{k=1}^{k_a} a_k L_{\text{Aus}}(n-k) = i'_L(n) + \sum_{k=1}^{k_b} b_k R_{\text{Aus}}(n-k).$$
(4.37)

Die z-Transformierte dieser Gleichung lautet:

$$L_{\text{Aus}}(z) \cdot \left(1 - \sum_{k=1}^{k_a} a_k z^{-k}\right) = i'_L(z) + R_{\text{Aus}}(z) \cdot \sum_{k=1}^{k_b} b_k z^{-k}$$
(4.38)

Im Fall eines alleinigen Paketverlustes im linken Kanal gilt: $R_{\text{Aus}}(z) = R(z)$.

$L_{\text{Aus}}(z)$ liegt am Ausgang des Filters vor und wird folgendermaßen berechnet:

$$L_{\text{Aus}}(z) = i'_L(z) \cdot \frac{1}{1 - \sum_{k=1}^{k_a} a_k z^{-k}} + R(z) \cdot \frac{\sum_{k=1}^{k_b} b_k z^{-k}}{1 - \sum_{k=1}^{k_a} a_k z^{-k}}$$
(4.39)

Das Synthesefilter entspricht dann dem in Bild 4.30 dargestellten System, wobei

$$H_D(z) = \frac{1}{1 - \sum_{k=1}^{k_a} a_k z^{-k}} \quad \text{und} \quad H_R(z) = \frac{\sum_{k=1}^{k_b} b_k z^{-k}}{1 - \sum_{k=1}^{k_a} a_k z^{-k}}$$
(4.40)

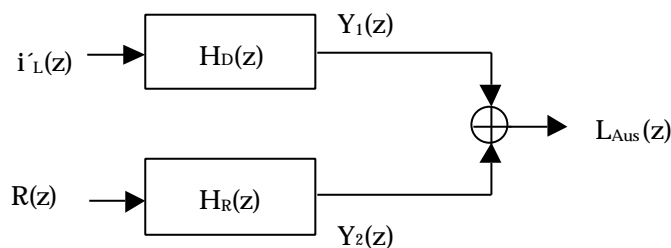


Bild 4.30: Synthesefilter des linken Kanals besteht aus der Parallelschaltung von zwei IIR-Filtern mit den Übertragungsfunktionen $H_D(z)$ und $H_R(z)$.

Die Stabilität des gesamten Systems hängt von der Stabilität beider Filter $H_D(z)$ und $H_R(z)$ ab. Sie sind nur dann stabil, wenn alle Pole innerhalb des Einheitskreises liegen.

Für die Berechnung der Polstellen beider Übertragungsfunktionen $H_D(z)$ und $H_R(z)$ soll folgende Gleichung nach z gelöst werden.

$$\begin{aligned}
1 - \sum_{k=1}^{k_a} a_k \cdot z^{-k} = 0 &\Leftrightarrow 1 - a_1 \cdot z^{-1} - a_2 \cdot z^{-2} - \dots - a_{k_a} \cdot z^{-k_a} = 0 \\
&\Leftrightarrow z^{k_a} - a_1 \cdot z^{k_a-1} - a_2 \cdot z^{k_a-2} - \dots - a_{k_a} = 0 \\
&\Leftrightarrow \sum_{k=0}^{k_a} \alpha_k \cdot z^k = 0,
\end{aligned} \tag{4.41}$$

wobei $\alpha_{k_a} = 1$ und $\alpha_j = -a_{(k_a-j)}$ $j \in \{0, \dots, k_a - 1\}$.

Die Lösung des Polynoms der Ordnung k_a aus Gleichung 4.41 erfolgt mit der Eigenwertmethode [Pre96].

2) Fall eines alleinigen Paketverlustes im rechten Kanal

Sollte beispielsweise ein Paket nur aus dem rechten Kanal fehlen, so muss der Empfänger ein Substitutionssignal $R_{\text{Aus}}(n)$ für diesen Kanal rekonstruieren. $L_{\text{Aus}}(n)$ entspricht in diesem Fall $L(n)$, da das Paket des linken Kanals fehlerfrei empfangen wurde. Für die Synthese des rechten Kanals gilt allgemein:

$$\begin{aligned}
R_{\text{Aus}}(n) &= i'_R(n) + \sum_{k=1}^{k_c} c_k R_{\text{Aus}}(n-k) + \sum_{k=0}^{k_d} d_k L_{\text{Aus}}(n-k), \quad n \in \{0..N-1\} \\
&\Leftrightarrow R_{\text{Aus}}(n) - \sum_{k=1}^{k_c} c_k R_{\text{Aus}}(n-k) = i'_R(n) + \sum_{k=0}^{k_d} d_k L(n-k)
\end{aligned} \tag{4.42}$$

Die z-Transformierte der obigen Gleichung ergibt

$$R_{\text{Aus}}(z) = i'_R(z) \cdot \frac{1}{1 - \sum_{k=1}^{k_c} c_k z^{-k}} + L(z) \cdot \frac{\sum_{k=0}^{k_d} d_k z^{-k}}{1 - \sum_{k=1}^{k_c} c_k z^{-k}} \tag{4.43}$$

Analog zu der theoretischen Betrachtung im ersten Fall ergibt sich folgende notwendige Bedingung: Die Synthese des Substitutionssignals für den rechten Kanal kann nur dann erfolgen, wenn auch das gesamte Signal aus Gleichung 4.43 stabil ist. Es soll also gelten, dass der Betrag aller Nullstellen der folgenden Gleichung jeweils kleiner als Eins ist.

$$\sum_{k=0}^{k_c} \gamma_k z^k = 0 \quad |z_i| \leq 1 \quad \forall z_i, \tag{4.44}$$

wobei

$$\gamma_{k_c} = 1 \quad \text{und} \quad \gamma_j = -c_{(k_c-j)} \quad j \in \{0, \dots, k_c - 1\} \tag{4.45}$$

3) Fall eines Paketverlustes in beiden Kanälen

Hier wird angenommen, dass beide Pakete aus dem linken und aus dem rechten Kanal fehlen. In diesem Fall stehen nur die Koeffizienten des Synthesefilters, die in dem letzten Zyklus berechnet wurden, und die Innovationen $i'_L(n)$ und $i'_R(n)$ zur Verfügung.

Für den rechten Kanal gilt:

$$\begin{aligned} R_{Aus}(n) &= i'_R(n) + \sum_{k=1}^{k_c} c_k R_{Aus}(n-k) + \sum_{k=0}^{k_d} d_k L_{Aus}(n-k), \quad n \in \{0..N-1\} \\ &= i'_R(n) + \sum_{k=1}^{k_c} c_k R_{Aus}(n-k) + d_0 L_{Aus}(n) + \sum_{k=1}^{k_d} d_k L_{Aus}(n-k), \end{aligned} \quad (4.46)$$

In dieser Gleichung wird nun $L_{Aus}(n)$ aus Gleichung 4.34 eingesetzt:

$$\begin{aligned} R_{Aus}(n) &= i'_R(n) + \sum_{k=1}^{k_c} c_k R_{Aus}(n-k) + \sum_{k=1}^{k_d} d_k L_{Aus}(n-k) \\ &\quad + d_0 \left[\underbrace{i'_L(n) + \sum_{k=1}^{k_a} a_k L_{Aus}(n-k) + \sum_{k=1}^{k_b} b_k R_{Aus}(n-k)}_{L_{Aus}(n)} \right] \\ &= i'_R(n) + d_0 i'_L(n) \\ &\quad + \sum_{k=1}^{k_c} c_k R_{Aus}(n-k) + d_0 \sum_{k=1}^{k_b} b_k R_{Aus}(n-k) \\ &\quad + \sum_{k=1}^{k_d} d_k L_{Aus}(n-k) + d_0 \sum_{k=1}^{k_a} a_k L_{Aus}(n-k) \end{aligned} \quad (4.47)$$

Die z-Transformierte dieser Gleichung lautet:

$$\begin{aligned} R_{Aus}(z) &= i'_R(z) + d_0 i'_L(z) \\ &\quad + R_{Aus}(z) \cdot \left(\sum_{k=1}^{k_c} c_k z^{-k} + d_0 \sum_{k=1}^{k_b} b_k z^{-k} \right) \\ &\quad + L_{Aus}(z) \cdot \left(\sum_{k=1}^{k_d} d_k z^{-k} + d_0 \sum_{k=1}^{k_a} a_k z^{-k} \right) \end{aligned} \quad (4.48)$$

Es gelten also für $R_{Aus}(z)$ und $L_{Aus}(z)$ folgende Gleichungen:

$$R_{Aus}(z) = \frac{i'_R(z) + d_0 i'_L(z)}{1 - \sum_{k=1}^{k_c} c_k z^{-k} - d_0 \sum_{k=1}^{k_b} b_k z^{-k}} + L_{Aus}(z) \frac{\sum_{k=1}^{k_d} d_k z^{-k} + d_0 \sum_{k=1}^{k_a} a_k z^{-k}}{1 - \sum_{k=1}^{k_c} c_k z^{-k} - d_0 \sum_{k=1}^{k_b} b_k z^{-k}} \quad (4.49)$$

$$L_{\text{Aus}}(z) = R_{\text{Aus}}(z) \frac{1 - \sum_{k=1}^{k_c} c_k z^{-k} - d_0 \sum_{k=1}^{k_b} b_k z^{-k}}{\sum_{k=1}^{k_d} d_k z^{-k} + d_0 \sum_{k=1}^{k_a} a_k z^{-k}} - \frac{i'_R(z) + d_0 i'_L(z)}{\sum_{k=1}^{k_d} d_k z^{-k} + d_0 \sum_{k=1}^{k_a} a_k z^{-k}} \quad (4.50)$$

Seien $k_1 = \max(k_a, k_d)$ und $k_2 = \max(k_c, k_b)$.

$R_{\text{Aus}}(z)$ und $L_{\text{Aus}}(z)$ lassen sich dann wie folgt vereinfachen:

$$R_{\text{Aus}}(z) = \frac{i'_R(z) + d_0 i'_L(z)}{\sum_{k=0}^{k_2} \beta_k z^{-k}} + L_{\text{Aus}}(z) \frac{\sum_{k=1}^{k_1} \alpha_k z^{-k}}{\sum_{k=0}^{k_2} \beta_k z^{-k}} \quad (4.51)$$

und

$$L_{\text{Aus}}(z) = R_{\text{Aus}}(z) \frac{\sum_{k=0}^{k_2} \beta_k z^{-k}}{\sum_{k=1}^{k_1} \alpha_k z^{-k}} - \frac{i'_R(z) + d_0 i'_L(z)}{\sum_{k=1}^{k_1} \alpha_k z^{-k}}, \quad (4.52)$$

wobei

$$\alpha_k = d_k + d_0 \cdot a_k, \quad (d_k = 0 \text{ für } k > k_d; a_k = 0 \text{ für } k > k_a). \quad (4.53)$$

$$\beta_k = -(c_k + d_0 \cdot b_k), \quad (c_k = 0 \text{ für } k > k_c; b_k = 0 \text{ für } k > k_b). \quad (4.54)$$

$$\beta_0 = 1$$

Das Synthesefilter für die Berechnung von $L_{\text{Aus}}(n)$ und $R_{\text{Aus}}(n)$ ist nur dann stabil, wenn für die Lösung z_i der folgenden Gleichung die Voraussetzung $|z_i| < 1$ erfüllt ist.

$$\sum_{k=1}^{k_1} \alpha_k z^{-k} = \sum_{k=1}^{k_2} \beta_k z^{-k} = 0 \quad (4.55)$$

4.5.3.3 Behandlung der Instabilität des Synthesefilters

Das Problem der Instabilität des SLPC-Synthesefilters kann mit Hilfe einer Kontrolle in einem Adaptionsalgorithmus bewältigt werden. Dieser Algorithmus soll ausgehend von den aktuellen Filterkoeffizienten des Synthesefilters die Instabilitäten erfassen und eine stabile Lösung der Filterkoeffizienten erzwingen. Die Filterkoeffizienten können z.B. so geändert werden, dass die Pole, die außerhalb des Einheitskreises liegen und somit die Instabilität des Filters verursachen, in den Einheitskreis verschoben werden. In [Ant93] wurde gezeigt, dass solche Pole durch ihre Reziproken ersetzt werden können, ohne dass sich der Amplitudengang ändert. Allerdings soll hier die Übertragungsfunktion zusätzlich mit einer Konstanten

(Verstärkungsfaktor¹) multipliziert werden (Gl. 4.56 und Gl. 4.57). Im Folgenden wird der theoretische Beweis dieser Annahme aufgeführt.

Für eine Polstelle z_i der Übertragungsfunktion $H(z)$ gilt:

$$\begin{aligned} |(e^{j\omega T} - z_i)| &= \left| e^{j\omega T} \cdot z_i \cdot \left(e^{-j\omega T} - \frac{1}{z_i} \right) \right| \\ &= |z_i| \cdot \left| \left(e^{j\omega T} - \frac{1}{z_i^*} \right)^* \right| \\ &= |z_i| \cdot \left| \left(e^{j\omega T} - \frac{1}{z_i} \right) \right| \end{aligned} \quad (4.56)$$

z_i^* ist die zu z_i konjugiert komplexe Polstelle der Übertragungsfunktion $H(z)$.

Für z_i reell gilt:

$$|(e^{j\omega T} - z_i)| = |z_i| \cdot \left| \left(e^{j\omega T} - \frac{1}{z_i} \right) \right| \quad (4.57)$$

Für komplexes z_i gilt:

$$|(e^{j\omega T} - z_i) \cdot (e^{j\omega T} - z_i^*)| = |z_i|^2 \cdot \left| \left(e^{j\omega T} - \frac{1}{z_i} \right) \cdot \left(e^{j\omega T} - \frac{1}{z_i^*} \right) \right| \quad (4.58)$$

Im Anhang B.1 ist ein zusätzlicher trigonometrischer Beweis zur Bestätigung dieser Annahme aufgeführt.

In den folgenden Bildern wird der Einfluss der Polstellenverschiebung nach [Ant93] auf die Impulsantwort, den Amplituden- und Phasengang eines instabilen Filters der Ordnung 5 dargestellt.

Bild 4.31 zeigt die Pole dieser Übertragungsfunktion vor und nach der Verschiebung. Die Übertragungsfunktion hat fünf Pole, wobei ein Pol außerhalb des Einheitskreises liegt. Nach der Verschiebung liegen alle Pole innerhalb des Einheitskreises.

¹ Der Verstärkungsfaktor beträgt $1/|z_i|$, wobei z_i die instabile Polstelle bezeichnet.

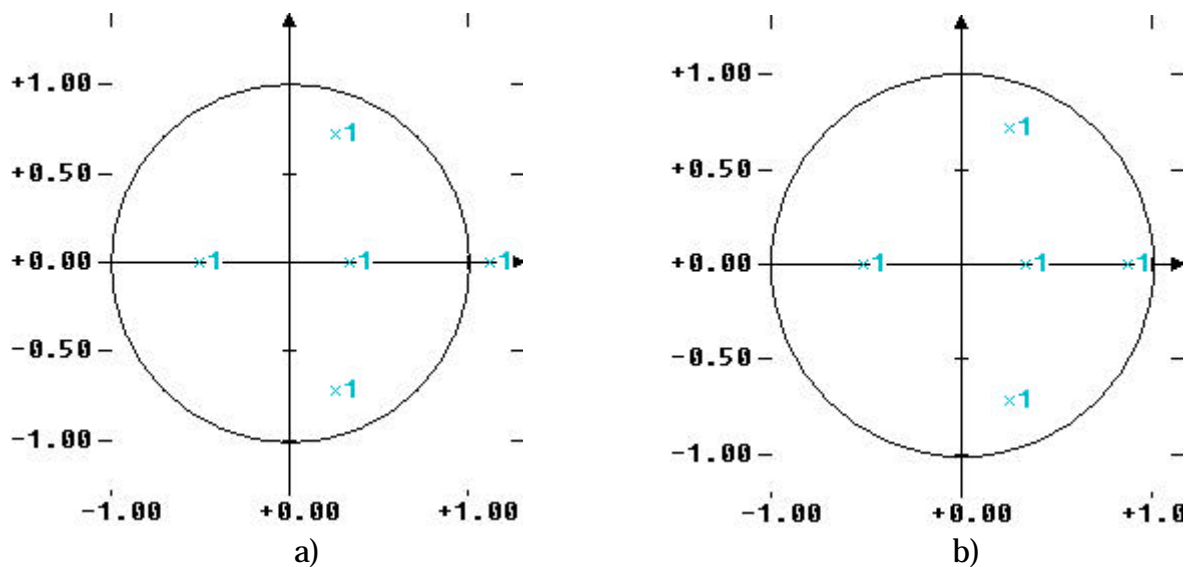


Bild 4.31: Lage der Polstellen der Übertragungsfunktion des Synthesefilters; a) vor der Polstellenverschiebung. Hier liegt eine Polstelle außerhalb des Einheitskreises; b) nachher: alle Polstellen liegen innerhalb des Einheitskreises und definieren damit ein stabiles Synthesefilter.

In Bild 4.32 sind die Impulsantworten der Übertragungsfunktion vor (Bild a) und nach (Bild b) der Polstellenverschiebung dargestellt. Es ist nicht zu übersehen, dass die Impulsantwort ursprünglich instabil war.

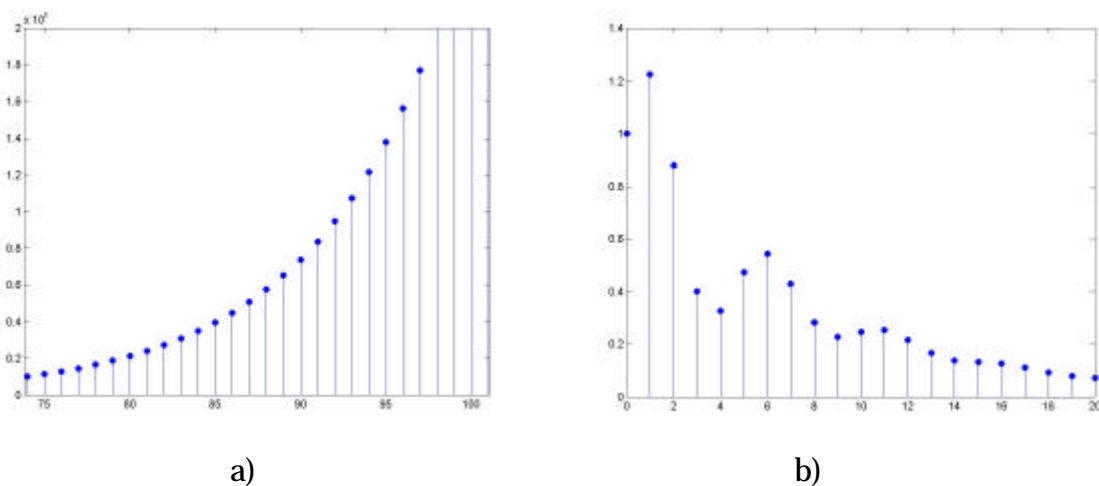


Bild 4.32: Impulsantwort des Synthesefilters. a) vor der Polstellenverschiebung, hier befand sich eine Polstelle außerhalb des Einheitskreises. Das Synthesefilter ist hier eindeutig instabil; b) danach: Nach der Polstellenverschiebung ergibt sich ein stabiles rekursives Filter

Die Behauptung, dass die Polstellenverschiebung keinen Einfluss auf den Amplitudengang hat, wird in Bild 4.33 bestätigt. Der Amplitudengang hat sich trotz der Änderung der Über-

tragungsfunktion nicht verändert (a und c). Im Gegensatz dazu hat sich der Phasengang verändert (b und d).

In [Hut94] wurde untersucht, ob die Phasenbeziehungen der Teiltöne eines Audiosignals für das Hören von Belang sind oder nicht. Die Untersuchungen zeigten, dass Phasenänderungen eines Teiltones nur dann hörbar sind, wenn dieser in eine Frequenzgruppe mit einem anderen Teilton fällt. Es zeigt sich aber, dass neben der Lage der Teiltöne der Betrag der Phasenänderung für deren Hörbarkeit von Bedeutung ist. Wird die Phase eines Teiltones nämlich um weniger als 45° gegenüber der Grundschiwingung des Klanges geändert, so ist es nicht wahrscheinlich, dass diese Änderung hörbar ist, obwohl der phasengeänderte Teilton in eine Frequenzgruppe mit einem anderen Teilton fällt. Änderungen der Teiltonphase sind jedoch hörbar, wenn der Betrag der Phasenänderung größer oder gleich 90° ist.

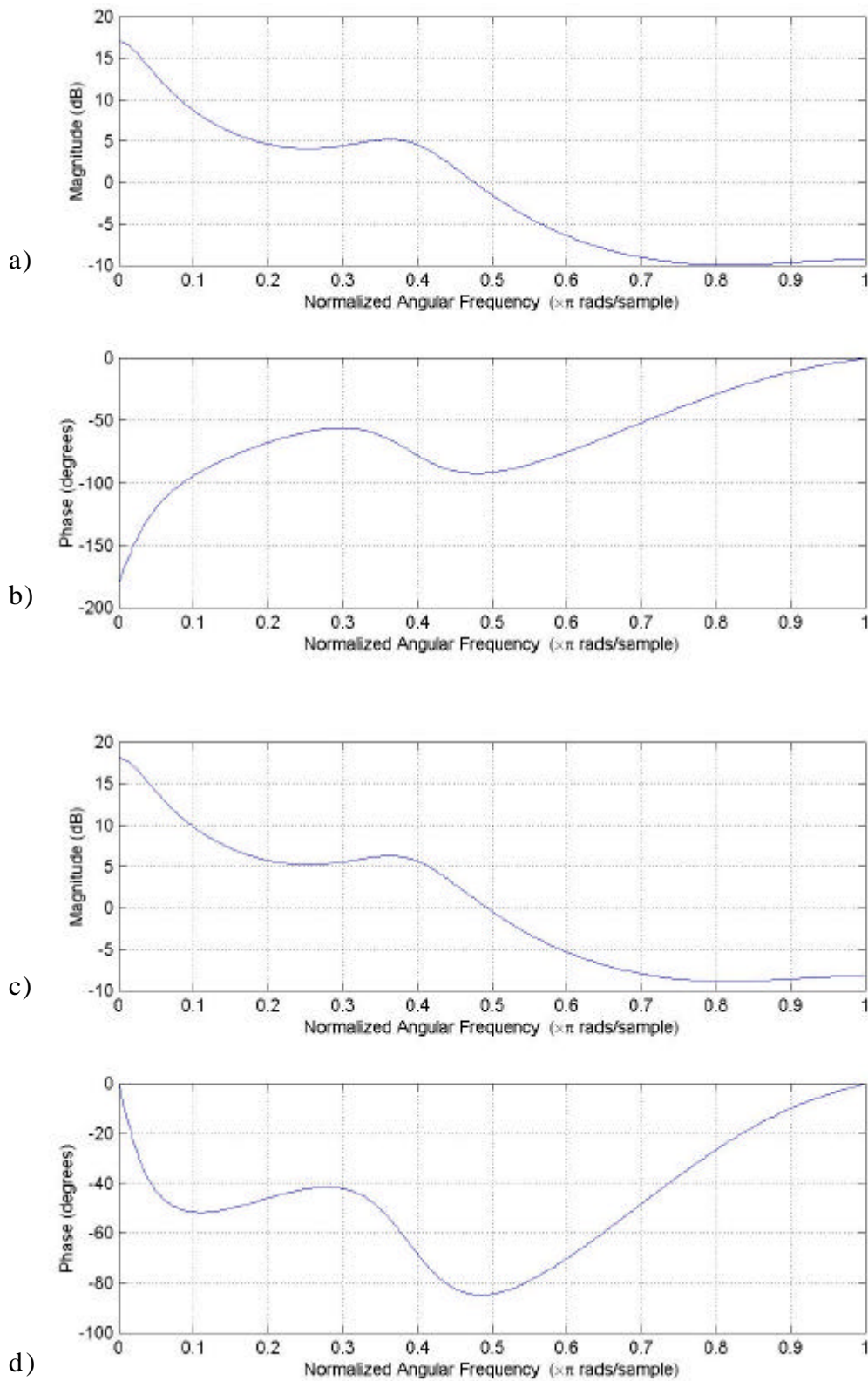


Bild 4.33: Amplituden- und Phasengang der Übertragungsfunktion des SLPC-Synthesefilters vor (a und b) und nach (c und d) der Polstellenverschiebung. Der Amplitudengang hat sich trotz Polstellenverschiebung - Im Gegensatz zum Phasengang - nicht verändert.

Der aus der Polstellenverschiebung entstandene Verstärkungsfaktor v wird für die Synthese des Substitutionssignals herangezogen. Im folgenden wird dieses anhand eines Beispiels erläutert:

Im Fall eines alleinigen Paketverlustes im linken Kanal werden beide Übertragungsfunktionen $H_D(z)$ und $H_R(z)$ aus Gleichung 4.40 mit dem Faktor v multipliziert.

Das Ausgangssignal $L_{Aus}(n)$ aus Gleichung 4.36 wird wie folgt modifiziert:

$$L_{Aus}(n) = \frac{1}{v} \cdot i'_L(n) + \sum_{k=1}^{k_a} a_k L_{Aus}(n-k) + \frac{1}{v} \cdot \sum_{k=1}^{k_b} b_k R_{Aus}(n-k) \quad (4.59)$$

Bedingte Stabilität des Synthesefilters

Mit dem oben erwähnten Adaptionsalgorithmus werden alle Polstellen des Synthesefilters, die außerhalb des Einheitskreises liegen, ersetzt. Es ergibt sich also somit ein stabiles Synthesefilter. Nun reicht die Stabilität dieses Filters für die Synthese nicht völlig aus, da es sich dabei auch um nur eine bedingte Stabilität handeln kann. Bedingte Stabilität bedeutet, dass die Pole des Filters dicht an dem Einheitskreises liegen. In diesem Fall zeigt das Filter eine „Überreaktion“ und verhält sich wie ein instabiles Filter.

In Tabelle 4.1 ist dargestellt, wie oft das SLPC-Synthesefilter, bei einer Simulation der Rekonstruktion von Paketverlusten mit der Stereoprädiktion, ein instabiles Verhalten aufweist. Die Synthese des Substitutionssignals erfolgt im ersten Fall ohne Stabilitätskontrolle und Verschiebung der Polstellen (a). In den Fällen b) bis e) wird eine Stabilitätskontrolle durchgeführt. Es werden alle Polstellen, die eine definierten Stabilitätsgrenze übersteigen, als kritisch bezeichnet. Die Stabilitätsgrenze wird bei 1 (b), bei 0.95 (c), bei 0.92 (d) oder bei 0.9 (e) gesetzt. Das dabei verwendete Substitutionsverfahren ist die Paketwiederholung auf der Ebene der Innovation und anschließende Synthese mit dem SLPC-Synthesefilter.

	$K_I = 5$	$K_I = 10$	$K_I = 15$
a) Ohne Stabilitätskontrolle	7	6	7
b) Stabilitätsgrenze bei 1	4	3	4
c) Stabilitätsgrenze bei 0.95	3	1	1
d) Stabilitätsgrenze bei 0.92	2	1	1
e) Stabilitätsgrenze bei 0.90	0	0	0

Tabelle 4.1: Häufigkeit der vorhandenen Instabilitäten bei der Synthese eines Substitutionssignals. Die Filterkoeffizienten bleiben bei a) unverändert; In den Fällen b) bis e) werden die Polstellen, welche die dabei definierte Instabilitätsgrenze überstreifen als instabil interpretiert und modifiziert. Die Filterordnung des

Prädiktorfilters $k_i=k_a=k_b=k_c=k_d$.; Paketverlustwahrscheinlichkeit $P_v = 1\%$ und Paketlänge N beträgt 512 Abtastwerte. Signaldauer = 20 sec.

Ohne Stabilitätskontrolle wird das SLPC-Synthesefilter z.B. für $k_i = 5$ an sieben Stellen instabil. Die Paketsubstitution mit der Stereoprädiktion führt an dieser Stelle zu einer Verschlechterung der wahrgenommenen Audioqualität. Mit der Stabilitätskontrolle werden alle Polstellen, die außerhalb des Einheitskreises liegen, in den Einheitskreis verschoben. Im Fall b) liegt die Stabilitätsgrenze bei 1 (Radius des Einheitskreises). Das SLPC-Synthesefilter weist trotzdem noch an vier Stellen ein instabiles Verhalten auf, obwohl alle Polstellen innerhalb des Einheitskreises liegen. Das ist damit zu begründen, dass einige Polstellen dicht bei 1 liegen. Das Synthesefilter ist somit nur bedingt stabil. Setzt man die Instabilitätsgrenze auf 0,95; 0,92 oder 0,9, so erzielt man bessere Ergebnisse (weniger Instabilitäten).

Fall der bedingten Stabilität

Liegt die Polstelle des SLPC-Synthesefilter zwischen der festgelegten Stabilitätsgrenze und dem Rand des Einheitskreises, so ist das Filter nur bedingt stabil. In diesem Fall sollte die Polstelle weiter ins Innere des Einheitskreises verschoben werden.

Der Betrag der neuen Polstelle entspricht der festgelegten Stabilitätsgrenze. Real- und Imaginär-Teil der alten Polstelle werden so verändert, dass die Phase erhalten bleibt. Für die Komponenten der neuen Polstelle nach der Verschiebung gilt:

$$x_n = \frac{0,9}{|z_a|} \cdot x_a \quad (4.60)$$

$$y_n = \frac{0,9}{|z_a|} \cdot y_a, \text{ wobei} \quad (4.61)$$

x_n und y_n : Koordinaten der neuen Polstelle z_n .

x_a und y_a : Koordinaten der alten Polstelle z_a .

Der Amplitudengang der Übertragungsfunktion wird in diesem Fall verändert. Es besteht im Gegensatz zu der Polstellenprojektion kein direkter Proportionalitätsfaktor zwischen der ursprünglichen und der nach der Polstellenverschiebung entstandenen Übertragungsfunktion des Synthesefilters.

Für die Rekonstruktion von Paketverlusten mit der Stereoprädiktion wird die Instabilitätsgrenze von 0,9 verwendet.

Rekonstruktion von Bündelfehlern

Bündelfehler lassen sich prinzipiell wie Einzelfehler behandeln. Bei einem Verlust von zwei oder mehreren aufeinanderfolgenden Paketen unterscheiden sich die Längen der zu rekonstruierenden Signale. Für das erste Paket wird ein Substitutionssignal der Länge $N + 2\Delta N$ erzeugt, wobei N die Paketlänge und ΔN die Länge des Merging-Bereichs bezeichnen. Ab dem zweiten Paket ist für die Rekonstruktion nur noch die Länge $N + \Delta N$ notwendig.

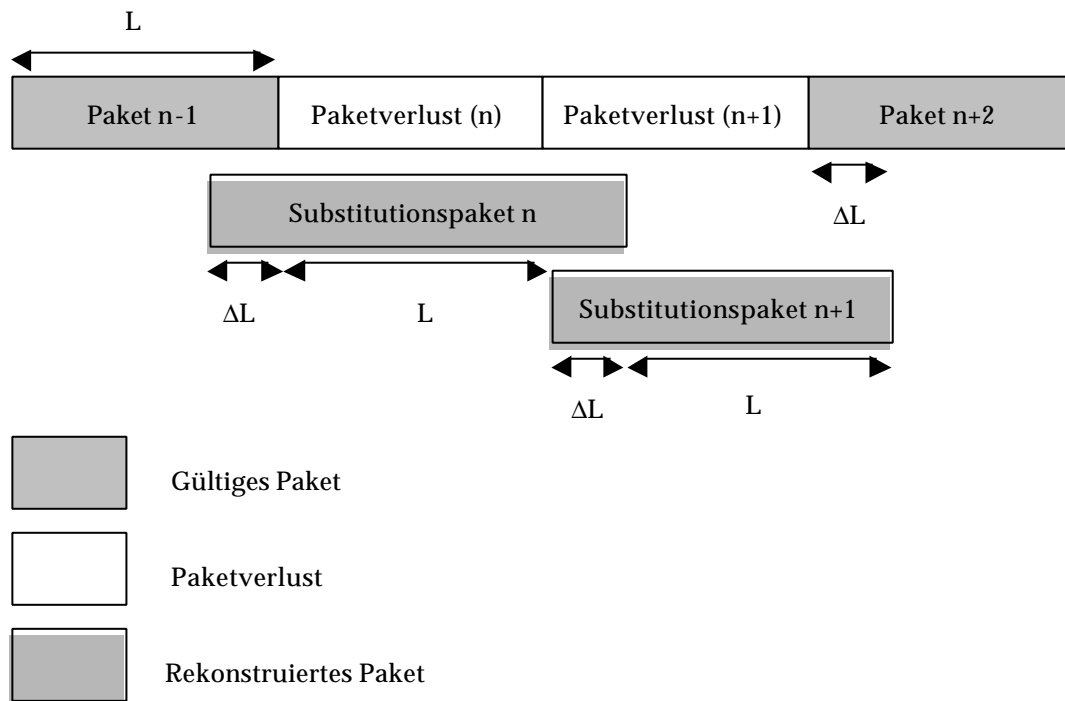


Bild 4.34: Längen der Substitutionssignale bei Bündelfehler. Für die Rekonstruktion der ersten Lücke Bündel ist ein Substitutionssignal der Länge $N + 2\Delta N$ erforderlich. Für die zweite Lücke wird ein Substitutionssignal der Länge $N + \Delta N$ benötigt.

Hinweis

Eine wesentliche Verbesserung der Paketsubstitution bei Paketverlusten mittels Stereoprädiktion wird dadurch erreicht, dass man für die Substitution die tatsächlichen Filterkoeffizienten des verlorenen Paketes verwendet. Diese stehen natürlich nicht zur Verfügung. Man könnte beispielsweise die Koeffizienten senderseitig berechnen und sie mit dem nächsten Paket übertragen, so dass bei Paketverlust die Koeffizienten nicht verloren gehen und zur Rekonstruktion verwendet werden können. Dies würde aber eine Erhöhung der zu übertragenden Datenmenge bewirken.

Signalverläufe

Bild 4.35 zeigt die Signalverläufe und die dazugehörigen Spektrogramme eines Stereo-Audiosignals (Pop-Musik-Titel). Es handelt sich hier um einen Ausschnitt der Länge 10000 Abtastwerte (227 ms) aus dem Originalsignal. Die Kanäle links und rechts sind in Bild 4.35-a und Bild 4.35-b dargestellt. Beide Kanäle weisen eine hohe Ähnlichkeit des Signalverlaufs auf. Bild 4.35-c stellt das Signal des linken Kanals mit Paketverlusten dar. Die Paketlänge beträgt hierbei 1024 Abtastwerte (22 ms). Der rechte Kanal bleibt hier fehlerfrei. Für die Synthese eines Substitutionssignals verwendet das SLPC-Synthesefilter die Filterkoeffizienten aus dem zuletzt korrekt empfangenen Paket und eine Innovation $i'(n)$, die in Bild 4.35-d) aus Nullabtastwerten und in Bild 4.35-e) aus der Innovation des letzten Paketes besteht. Für die Rekonstruktion der zweiten Lücke aus Bild 4.35-c) nutzt die Stereoprädiktion die Korrelation zwischen beiden Kanälen aus. Das Substitutionssignal aus Bild 4.35-e) weist sowohl im Zeit- als auch im Spektralbereich eine hohe Ähnlichkeit mit dem Originalsignal auf. Sollte das Paket aus dem rechten Kanal fehlen, so würde das Verfahren keine Möglichkeit mehr haben, diese Lücke genauso gut zu rekonstruieren.

Die erste Lücke befindet sich in einem rauschähnlichen Signalabschnitt. Hier ist es mit dem Verfahren nicht möglich, den ursprünglichen Signalverlauf wiederherzustellen. Auch, wenn die Spektraldarstellung in diesem Fall (Bild 4.35-e) auf eine gute Substitution deutet, bleibt die empfundenen subjektive Audioqualität an dieser Stelle störend.

Die Paketsubstitution mittels Stereoprädiktion eignet sich nur bei Signalen, die eine hohe Stereokorrelation besitzen (siehe dazu die Bewertung aus Kapitel 5). Es ist hier auch notwendig, dass ein Paketverlust nur in einem Kanal auftritt, damit das Verfahren die Möglichkeit bekommt, das Audiosignal aus dem Nachbarkanal für die Rekonstruktion heranzuziehen.

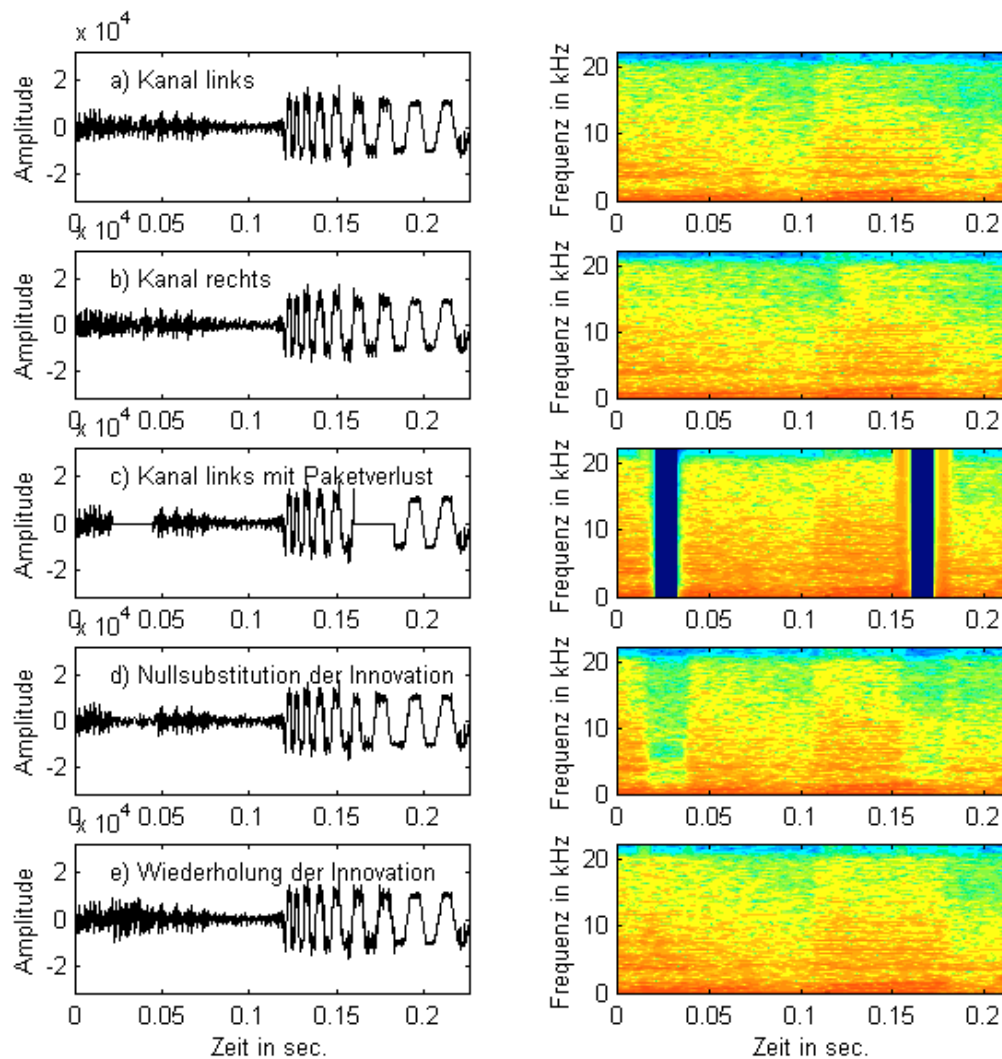


Bild 4.35: Paketsubstitution mit der Stereoprädiktion; Signalverläufe links und die dazugehörigen Spektrogramme rechts. Originalsignal aus einem Pop-Musik-Titel (10000 AW: 227 ms); a) Linker Kanal des Stereosignals; b) Rechter Kanal des Stereosignals; c) Linker Kanal mit Paketverlusten. Die Paketlänge beträgt hierbei 1024 Abtastwerte (22 ms); d) Mit der Stereoprädiktion substituiertes Signal, wobei hier die Innovation durch ein Nullsignal ersetzt wird. e) Mit der Stereoprädiktion substituiertes Signal, wobei hier die Innovation aus dem letzten Paket wiederholt wird.

4.6 Substitution im Frequenzbereich

4.6.1 Motivation für die Rekonstruktion von Paketverlusten im Frequenzbereich

Im Frequenzbereich werden Eigenschaften von Signalen sichtbar, die im Zeitbereich augenscheinlich verborgen bleiben. Sollten diese Eigenschaften mit Hilfe mathematischer Größen beschrieben werden, so besteht die Möglichkeit, diese Größen bei einem Paketverlust als quasistationär anzusehen und demzufolge zu interpolieren. Die Signalcharakteristik der Lücke kann somit geschätzt werden.

In diesem Abschnitt werden zwei Substitutionsverfahren vorgeschlagen, die mit Hilfe einer spektralen Analyse versuchen, Lücken im Audiosignal durch Parameter-Interpolation zu verdecken. Die spektrale Transformation erfolgt im ersten Verfahren (Spektrales Morphing), in welchem ein Spektrum für die Lücke anhand der Spektren vor und nach der Lücke geschätzt wird, mit der diskreten Cosinustransformation (DCT, s. Kap. 2.4.2). In dem zweiten Verfahren (Frequenz-Tracking) erfolgt die spektrale Analyse mit der diskreten Fourier-Transformation (DFT, s. Kap. 2.4.1).

4.6.2 Spektrales Morphing

Die Grundidee für diesen Algorithmus bestand in der Überlegung, die zum Ausfüllen der Lücke erforderlichen Segmente aus den spektralen Eigenschaften der zur Verfügung stehenden Segmente zu bestimmen. Zur Frequenzanalyse wird hier die DCT verwendet.

Die DCT hat den Vorteil, dass sie nur reelle Transformationskoeffizienten liefert und zudem weniger Fensterungseffekte als die DFT hat.

Das Grundprinzip des Verfahrens, das hier mit SM (Spektrales Morphing) bezeichnet wird, ist in Bild 4.37 dargestellt.

Beschreibung des Algorithmus

Für die Verdeckung der Lücke, die durch den Verlust des n -ten Paketes entsteht, werden die Spektren unmittelbar vor und nach dieser Lücke berechnet und gemittelt. Aus der Rücktransformation des neuen Spektrums in dem Zeitbereich ergibt sich ein Audiosignal, das

die Lücke ausfüllen soll. Der genaue Ablauf dieses Rekonstruktionsverfahrens wird im folgenden erläutert:

Schritt 1

Zuerst wird für den Bereich vor dem Paketverlust ein akkumuliertes Spektrum S_{akk} berechnet (2), das sich aus der Mittelwertbildung über die Spektren ($S_1..S_{P_n}$) aller gepufferten Segmente vor der Lücke ergibt (1), wobei P_n die Anzahl der vor der Lücke gepufferten Pakete darstellt (11). L_1 stellt die Pufferlänge dar, die bei unveränderter Paketlänge L gerade $L \cdot P_n$ entspricht. Die Mittelung der Spektren ergibt eine grobe Aussage über die Signalstatistik vor dem Verlust.

Um Diskontinuitäten in dem gepufferten Bereich zu vermeiden, werden die Ausgangsdaten gepuffert.

Schritt 2

Als nächstes wird das Zielspektrum S_{next} des Signalausschnittes, der der Lücke unmittelbar folgt, berechnet (4). Sollte es sich um einen Bündelfehler handeln, so wird ein aus Nullwerten bestehendes Spektrum verwendet.

Schritt 3

Nun wird das akkumulierte Spektrum S_{akk} mit dem Zielspektrum S_{next} gemittelt (5) und das erhaltene Spektrum noch zwei weitere Male mit dem Zielspektrum gemittelt (6 und 7). Man erhält dann drei Spektren S_{rek1} , S_{rek2} und S_{rek3} , die ausgehend von dem akkumulierten Spektrum dem Zielspektrum immer ähnlicher werden.

Schritt 4

Die Spektren S_{rek1} , S_{rek2} und S_{rek3} werden durch inverse DCT in Zeitsignalsegmente umgewandelt (8), welche gemäß Bild 4.36 mit der Hanning-Funktion gefenstert und zu einem Segment Y der doppelten Länge versetzt überlagert werden.

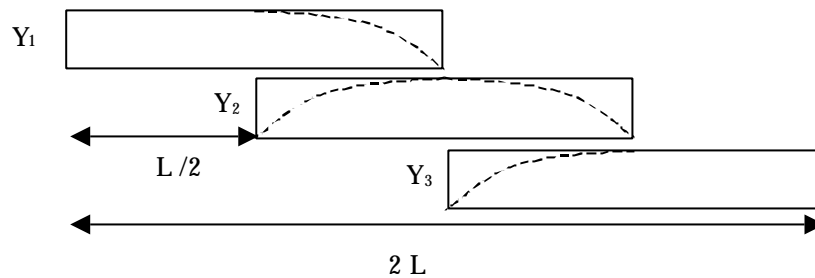


Bild 4.36: Gewichtung und Überlagerung der Zeitsignal-Segmente Y_1 , Y_2 und Y_3 .

Die Fensterung der ersten Hälfte von Y_1 und zweiten Hälfte von Y_3 wird weggelassen, da dies zu einer zweifachen Gewichtung führt. Wenn das Rekonstruktionssignal in der Lücke eingefügt wird, müssen die Ränder der Lücke und die des Rekonstruktionssignal gefenstert und überlagert werden (*merging*).

Durch die Mittelung aus Schritt 3 wird sich die Amplitude der Segmente vom Niveau vor der Lücke auf das Niveau nachher entwickeln, damit würde diese Lücke gut ausgefüllt werden.

Schritt 5

Dieser Schritt wird nur für Paketlängen kleiner als 1024 AW ausgeführt. Der Grund dafür wird in dem nächsten Abschnitt erklärt (Anpassung für kleine Paketlängen).

Aus dem rekonstruierten Signal Y , das für kleine Paketlängen 2048 AW enthält, wird zum Ausfüllen der Lücke ein Signalausschnitt Y_{opt} der Länge $L+2\Delta L$ benötigt, der die maximale Ähnlichkeit mit dem verlorenen Signalausschnitt haben sollte. Bei 50% Merging, was bei diesem Rekonstruktionsverfahren zu empfehlen ist, hat Y_{opt} die Länge $2L$. Dies erfolgt mit Hilfe einer Korrelations-Untersuchung zwischen einem temporären Feld und dem Puffer, wie sie in Kapitel 4.3.6 beschrieben ist.

Y_{opt} wird anschließend in den Ringpuffer (11) geschrieben und als Substitutionssignal ausgegeben.

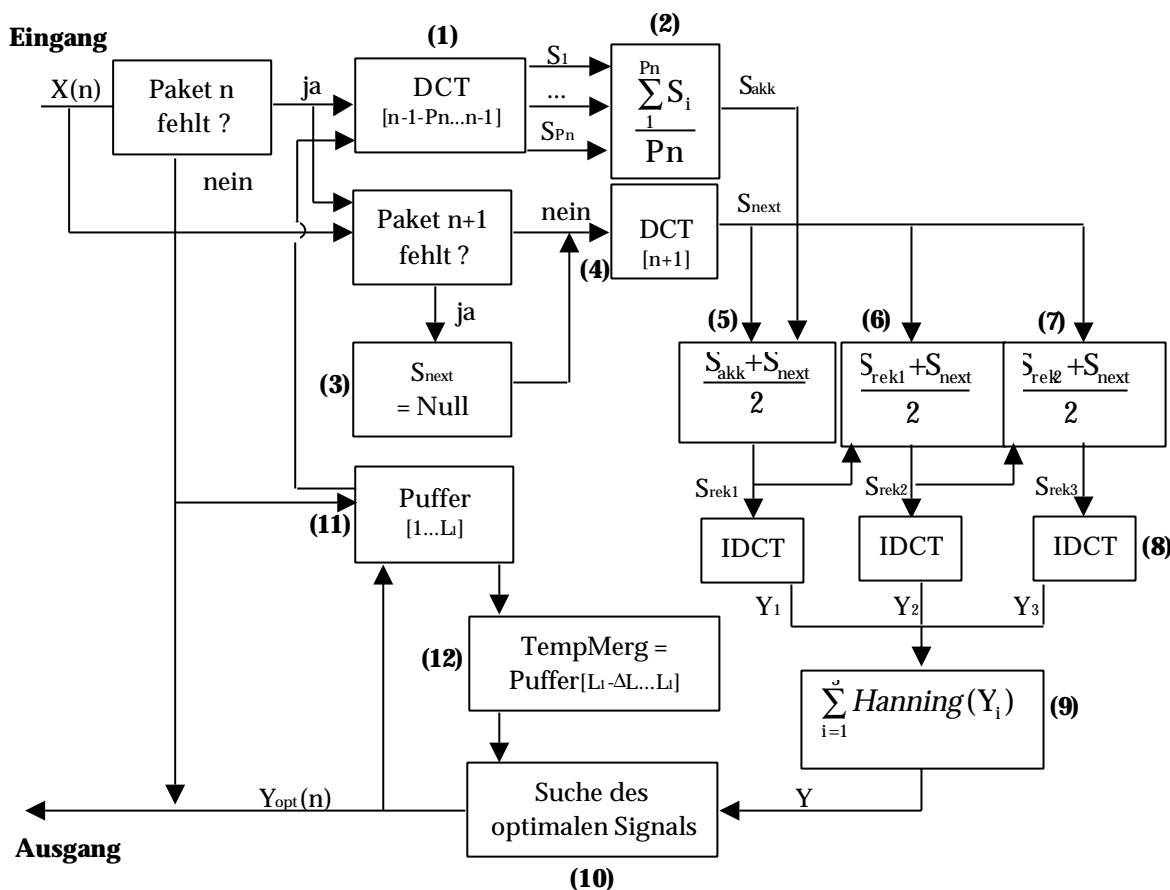


Bild 4.37: Funktionsprinzip des Spektrales-Morphing-Rekonstruktionsverfahrens

Anpassung an kleine Paketlängen

Der oben beschriebene Algorithmus wurde für unterschiedliche Paketlängen untersucht. Große Lücken, z.B. 1024¹ Abtastwerte, lassen sich mit diesem Verfahren gut verdecken. Die durch die Lücke entstandenen Störungen im Audiosignal sind nach der Verdeckung entweder meistens wahrnehmbar aber nicht störend. Dies gilt insbesondere für Paketverlustwahrscheinlichkeiten bis maximal 5%. Bei höheren Verlustwahrscheinlichkeiten bzw. kleineren Paketlängen wird die Audioqualität sogar stören. Letztes ist damit zu erklären, dass mit der DCT nur Kurzzeitspektren erstellt werden können. Die DCT errechnet aus N Abtastwerten N Transformationskoeffizienten, somit ergibt sich das Problem einer relativ geringen spektralen Auflösung bei kleinen Paketlängen. Aus diesem Grund war eine Modifikation des Algorithmus für kleine Paketlängen zwecks Verbesserung der subjektiven Audioqualität unentbehrlich.

Am Algorithmus wurden für Paketlängen unterhalb 1024 AW folgende Modifikationen vorgenommen:

¹ Entspricht 23,22 ms für 44,1 kHz Audio

- Die Pufferlänge, die den Analysebereich unmittelbar vor der Lücke festlegt, beträgt unabhängig von der Paketlänge 2048¹ Abtastwerte.
- Der Analysebereich unmittelbar nach der Lücke umfasst 1024 Abtastwerte. Sollten die entsprechenden Pakete fehlen, so wird dieser Bereich mit Nullen aufgefüllt.
- Bündelfehler werden zusammengefasst und als Einzelfehler behandelt, soweit die durch den Bündel entstandenen Lücken im Audiosignal 1024 Abtastwerte nicht überschreiten. Andernfalls werden die Paketverluste einzeln rekonstruiert.
- Das Rekonstruktionssignal besteht, unabhängig von der Paketlänge, aus 2048 Abtastwerten. Der zur Verdeckung der Lücke verwendete Signalausschnitt entsteht aus einer Suche des zur Lücke passenden Signals aus dem kompletten Rekonstruktionssignal (Schritt 5).

Rekonstruktion bei Stereosignalen

Mit dem vorgestellten SM-Rekonstruktionsverfahren werden Paketverluste in einem Kanal des Stereosignals ohne Zuhilfenahme der Information aus dem Nachbarkanal rekonstruiert. Manchmal ist diese Information für die Rekonstruktion jedoch hilfreich, insbesondere wenn es sich um Stereosignale handelt, die eine hohe Korrelation zwischen beiden Kanälen aufweisen wie z.B. Pop-Musik.

Für die Rekonstruktion von Paketverlusten in Stereosignalen mit dem SM-Verfahren wurden folgende Anpassungen vorgenommen:

- Zuerst wird untersucht, ob überhaupt eine Korrelation zwischen beiden Kanälen des Audiosignals vorhanden ist. Diese Untersuchung wird nur dann durchgeführt, wenn im Fall eines Paketverlustes in einem Kanal das zeitsynchrone Paket in dem Nachbarkanal vorhanden ist, andernfalls wird eine Mono-Rekonstruktion (Bild 4.37) vorgenommen.
- Die Untersuchung der Stereokorrelation erfolgt anhand des Vergleiches eines temporären Feldes² aus dem fehlerbehafteten Kanal mit einem Puffer aus dem zweiten Kanal. Es ergibt sich dann ein Feld aus dem zweiten Kanal, der die maximale Korrelation mit dem Temp-Feld des ersten Kanals ergibt. Sollte die Energie der Differenz beider Felder unterhalb 10 Prozent der Energie des Temp-Feldes liegen, so wird auf eine Stereokorrelation entschieden, andernfalls wird das Stereosignal als nicht korreliert angesehen und die Substitution im Mono-Kanal vorgenommen.
- Bei Stereo-Signalen mit Stereokorrelation zwischen beiden Kanälen werden die Module (5), (6) und (7) aus Bild 4.37 wie folgt modifiziert:

$$S_{\text{rek1}} = \frac{S_{\text{akk}} + S_{\text{Nachbar}}}{2} \quad (4.62)$$

¹ ungefähr 47 ms bei 44,1 kHz Audio

² Der Merging-Bereich vor der Lücke

$$S_{\text{rek2}} = S_{\text{Nachbar}} \quad (4.63)$$

$$S_{\text{rek3}} = \frac{S_{\text{rek2}} + S_{\text{next}}}{2} \quad (4.64)$$

Mit S_{Nachbar} wird das Spektrum des Paketes aus dem fehlerfreien Nachbarkanal bezeichnet.

Signalverläufe

Bild 4.38 zeigt die Signalverläufe und die dazugehörigen Spektrogramme des Stereosignals eines Audiosignals aus einem Pop-Musik-Titel. Es handelt sich hier um einen Ausschnitt der Länge 8000 Abtastwerte (180 ms) aus dem Originalsignal. Die Kanäle links und rechts sind in Bild 4.38-a und Bild 4.38-b dargestellt. Beide Kanäle weisen eine hohe Ähnlichkeit des Signalverlaufs. Bild 4.38-c) stellt das Signal des linken Kanals mit einem Paketverlust dar. Die Paketlänge beträgt hierbei 1024 Abtastwerte (22 ms). Der rechte Kanal bleibt hier fehlerfrei.

Die Lücke findet am Anfang einer Attacke statt. Hier ist es mit einfachen Substitutionsverfahren, wie z.B. der Paketwiederholung, nicht möglich diese Attacke zu rekonstruieren. Das SM-Verfahren verfolgt diese Änderung innerhalb der Lücke. Es wird damit nicht der ursprüngliche Verlauf rekonstruiert, was auch nicht Ziel des Verfahrens ist, sondern eher ein „natürlicher“ Übergang der Signaleigenschaften von dem Zustand vor dem Verlust auf dem Zustand danach gewährleistet (Bild 4.38-d).

Mit diesem Verfahren werden zwar die durch den Verlust hervorgerufenen Störungen im Audiosignal (Knackgeräusche) nicht mehr wahrnehmbar, allerdings klingt das Substitutionssignal in den meisten Fällen anders als das Originalsignal. Besonders stark fällt das bei Attacken, z.B. Schlaginstrumenten, auf.

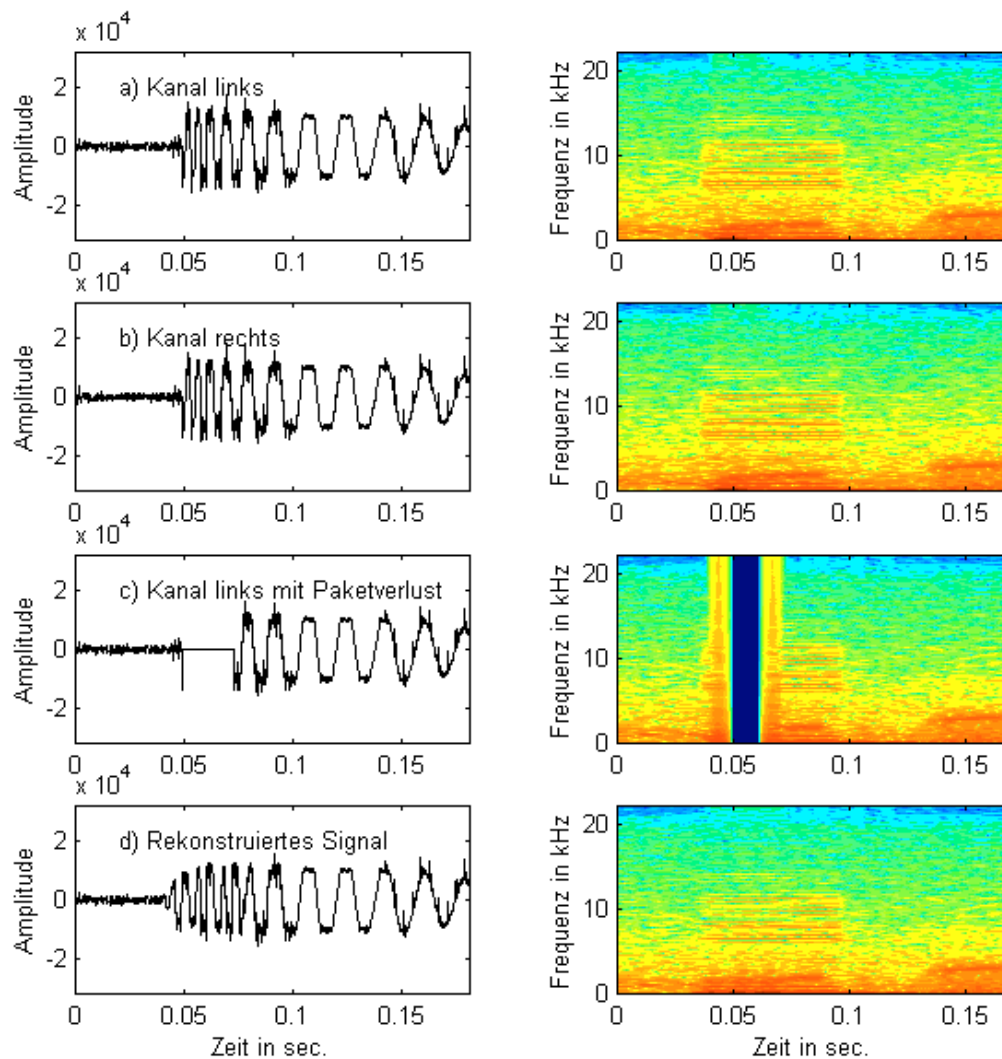


Bild 4.38: Paketsubstitution mit Spektralem Morphing; Signalverläufe links und die dazugehörigen Spektrogramme rechts. Originalsignal aus einem Pop-Musik-Titel (8000 AW: 180 ms); a) Linker Kanal des Stereosignals; b) Rechter Kanal des Stereosignals; c) Linker Kanal mit Paketverlust. Die Paketlänge beträgt hierbei 1024 Abtastwerte (22 ms); d) Mit Spektralem Morphing rekonstruiertes Signal.

Substitution bei Bündelfehlern

Mit dem Spektralen-Morphing-Verfahren erfolgt eine Interpolation der Spektren vor und nach dem Paketverlust. Dies setzt voraus, dass das Paket unmittelbar nach dem Verlust vorhanden ist. Ist das nicht der Fall (Bündelfehler), so wird die Mittelung mit einem Null-Spektrum durchgeführt, was ein langsames Abklingen des Signals gegen Null bedeutet. Am Ende des Bündels nimmt das Substitutionssignal langsam den Zustand des Audiosignals nach dem Verlust auf, da für diese Paketsubstitution ein Spektrum nach dem Verlust zur

Verfügung steht. Ein Beispiel dafür liefert Bild 4.39. Hier besteht die Lücke aus Bild 4.39-c aus einem Bündel bestehend aus zwei Paketverlusten.

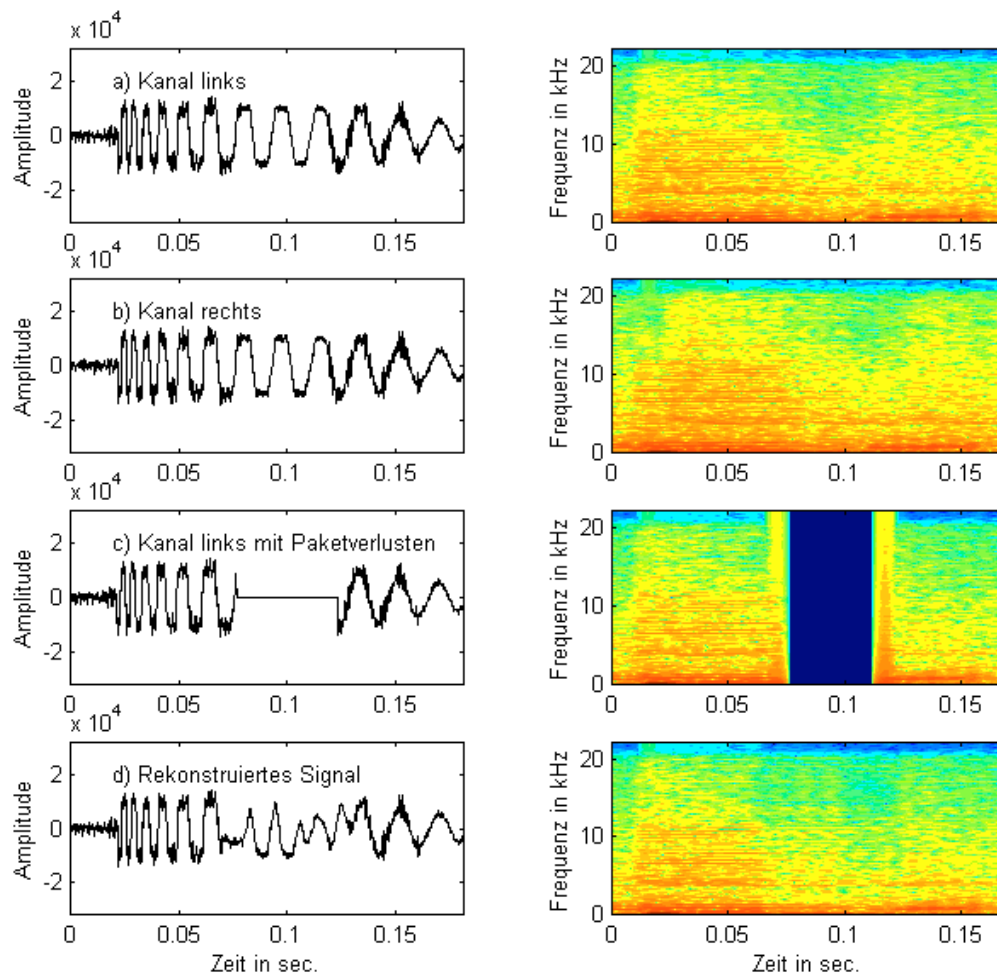


Bild 4.39: Paketsubstitution mit dem Spektralen-Morphing-Verfahren bei Bündelfehler; Signalverläufe links und die dazugehörigen Spektrogramme rechts. Originalsignal aus einem Pop-Musik-Titel (8000 AW: 180 ms); a) Linker Kanal des Stereosignals; b) Rechter Kanal des Stereosignals; c) Linker Kanal links mit Paketverlusten. Hier handelt es sich um zwei aufeinanderfolgende Paketverluste (Bündelfehler). Die Paketlänge beträgt 1024 Abtastwerten. d) Mit Spektralem-Morphing-Verfahren rekonstruiertes Signal.

Die Audioqualität des Substitutionssignals nimmt bei dem SM-Verfahren mit steigender Paketverlustwahrscheinlichkeit und bei dicht aufeinander folgenden Paketverlusten schnell ab. Das wird damit begründet, dass für die Paketsubstitution die Spektren gepufferter, zuvor empfangener oder auch rekonstruierter Signale, gemittelt werden. Eine schlechte Rekonstruktion kann die Rekonstruktion später auftretender Lücken beeinflussen (Fehlerschleppe) (Bild 4.13).

4.6.3 Frequenz-Tracking

Das Hauptproblem der Paketsubstitution bei Paketverlusten in Audiosignalen liegt meistens in der unzureichenden Parametrisierung des Signals. Die sinusoidale Modellierung von Audiosignalen (Kap. 2.5) liefert den wichtigsten Ansatz bezüglich dieser Problematik.

Mit der sinusoidalen Modellierung wird das Audiosignal in seine einzelnen Komponenten zerlegt unter der Ermittlung dessen Augenblicks-Frequenzen, -Amplituden und -Phasen. Das vorgeschlagene Frequenz-Tracking-Verfahren (FTR) zur Rekonstruktion von Paketverlusten in Audiosignalen beruht auf dem sinusoidalen Modell nach McAulay und Quatieri [McA86].

Für die Rekonstruktion der Lücke im Audiosignal, die durch einen Paketverlust während der Übertragung hervorgerufen wurde, müssen die Bereiche vor und nach der Lücke analysiert werden. Es werden also zwei Bereiche definiert; der vorangegangene und der nachfolgende Analysebereich. Rein intuitiv könnte die Länge der Analysebereiche gleich der Paketlänge gesetzt werden. Dieses würde bedeuten, dass sich die Analysebereiche auf die Pakete unmittelbar vor und nach der Lücke beschränken. Das hat einerseits den Vorteil, dass das Zentrum der Analyse dicht an der Fehlerstelle liegt und somit für die Rekonstruktion der Lücke relevante Informationen liefert. Andererseits ist eine kleine Paketlänge meistens nicht ausreichend, um eine vernünftige Frequenzauflösung zu erhalten, was zu einer ungenauen Analyse führt. Mit kleinen Paketlängen können die tiefen Frequenzen im Audiosignal nicht erfasst werden.

Beispiel

Eine Periode eines 50 Hz-Tones hat in einem mit 44,1 kHz abgetasteten Audiosignal eine Länge von 882 Abtastwerten. Um diesen Ton zu erfassen, müsste die Analyselänge mindestens das zweifache der Schwingungslänge haben, also ungefähr 1764 AW. Die Analyse von Paketen kleinerer Längen könnte also diese Frequenz nicht erfassen.

Es folgt eine Beschreibung des vorgeschlagenen Substitutionsverfahrens.

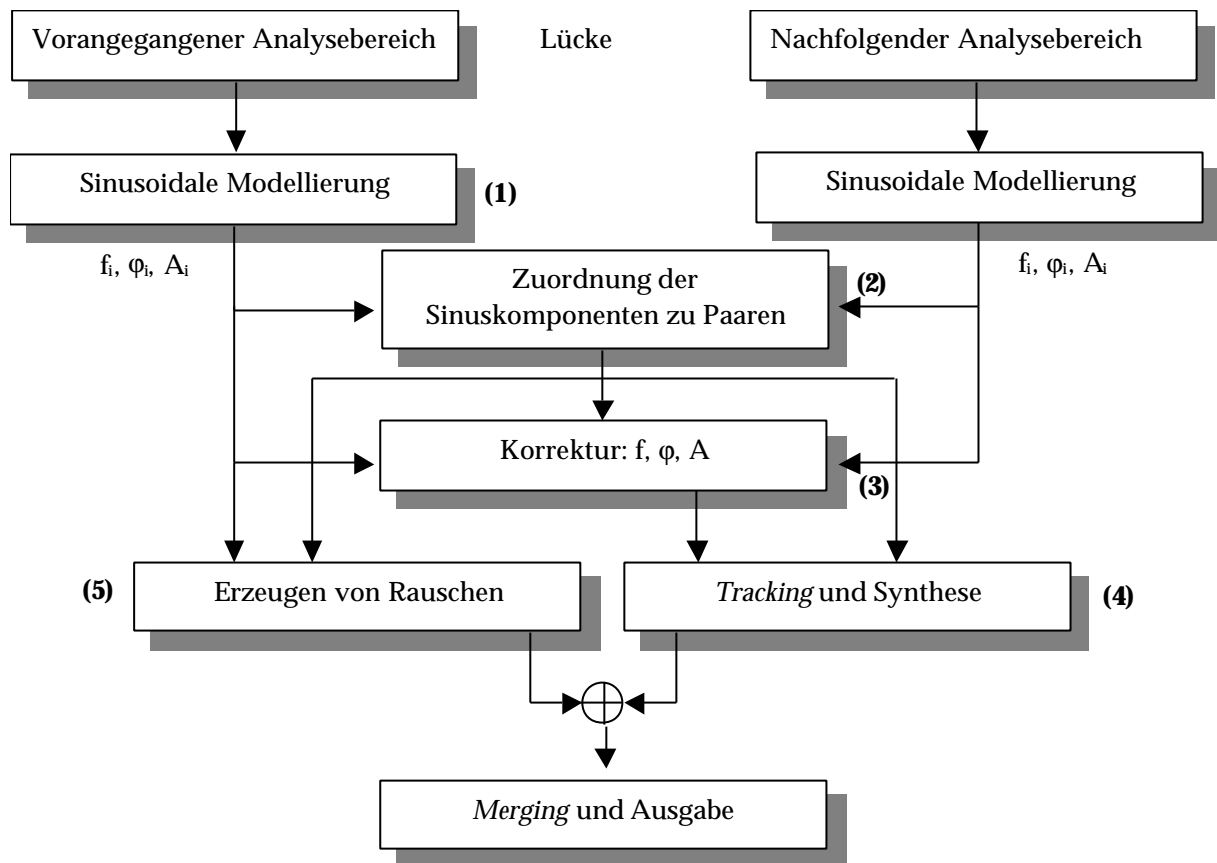


Bild 4.40: Grundprinzip des FTR-Substitutionsverfahrens

(1) Sinusoidale Modellierung

Bei einem Paketverlust werden die Bereiche vor und nach der Lücke analysiert. Das allgemeine Prinzip der sinusoidalen Modellierung eines Audiosignals wurde bereits in Kapitel 2.5 dargestellt.

- Vor der Analyse wird der gesamte Analysebereich gefenstert. In der Literatur wurden für die Analyse unterschiedliche Fenster verwendet z.B. ein Kaiser [Ser90] oder Hamming-Fenster [Mas94]. In dieser Arbeit wird das Hanning-Fenster verwendet, weil es in Bezug auf diese Anwendung keine Artefakte im hochfrequenten Bereich gezeigt hat, im Gegensatz zum Hamming-Fenster. Die Wahl des Tiefpass-Fensters ist für die spätere Analyse entscheidend. Es sollte ein Fenster verwendet werden, dessen Spektrum mit einer schmalen Hauptkeule und einer stark abfallenden Amplitude der Nebenkeulen gekennzeichnet ist. Dieses erleichtert die spätere Erfassung der Maxima im Betragsspektrum des analysierten Signals.
- Als nächstes wird das gefensterte Segment auf die gewünschte DFT-Länge¹ verlängert, indem die restlichen Abtastwerte mit Nullen aufgefüllt werden (*Zeropadding*). *Zeropad-*

¹ In diesem Verfahren wird für die Analyse eine DFT-Länge von 2048 verwendet.

ding bewirkt bei der Frequenzanalyse eine Erhöhung der Frequenzauflösung und somit eine genauere Frequenzermittlung.

- Aus den diskreten DFT-Koeffizienten wird nachfolgend der Betrag gebildet und alle lokalen Maxima im Betragsspektrum bestimmt.

Es ist hier zu beachten, dass sich einige Maxima nur aus der Gewichtung mit den Nebenkeulen im Spektrum des gewählten Fensters ergeben und somit keine wirkliche Frequenzen darstellen. Diese sollten bei der Analyse ausgeschlossen bleiben. Dies geschieht, indem eine minimale Schwelle¹ gesetzt wird, unter der alle Komponenten als irrelevante Maxima betrachtet werden. Alle DFT-Koeffizienten, die keine lokalen Maxima darstellen und auch nicht in deren unmittelbarer Umgebung liegen, werden als Rauschkomponenten angesehen.

- Nach der Ermittlung der relevanten Maxima im Spektrum wird eine genauere Bestimmung der Frequenzen dieser Maxima durchgeführt. Bei den im letzten Schritt ermittelten Frequenzen handelt es sich um diskrete Rasterfrequenzen gleicher Frequenzabstände. Durch Interpolation dieser Frequenzen mit den unmittelbar angrenzenden Frequenzwerten wird die wirkliche Frequenz für jedes lokale Maximum angenähert. Dies ist aufgrund der hohen Empfindlichkeit des Gehörs gegenüber Frequenzunterschieden notwendig.

Die Auflösung der DFT ist meist schlechter als die des Ohres. Eine typische DFT-Länge ist 1024 Punkte. Für 44,1-kHz-Audiosignale bedeutet das eine Auflösung von ungefähr 43 Hz. Das Ohr kann aber Töne, die einen Frequenzunterschied von 0,03 bis 0,08 Hz voneinander haben unterscheiden [Rak71].

Die Frequenzinterpolation erfolgt mit der quadratischen Interpolation [Mas96] [McI92]. Im Anhang B.2 wird die Theorie zur quadratischen Interpolation, oft mit *parabolic regression* bezeichnet, aufgeführt.

- Nach der Frequenzinterpolation werden für die ermittelten Frequenzen die zugehörigen Amplituden und Phasen berechnet. Die Methoden zur Ermittlung der Amplituden sind vielfältig. Die naheliegendste Methode ist die Verwendung der DFT-Koeffizienten [Pur98]. Die Amplitudenbestimmung kann noch verbessert werden durch die im Anhang B.2 beschriebene quadratische Interpolation.

Bei einer hohen Auflösung der DFT kann die Phase der DFT direkt für die Phase der Sinuskomponenten benutzt werden.

Bessere Ergebnisse als die reine Verarbeitung der DFT-Koeffizienten bietet die Methode der kleinsten Quadrate, die in diesem Zusammenhang mit sinusoidaler Regression bezeichnet wird. Diese Methode liefert auch bei tiefen Frequenzen genauere Amplituden und Phasen als die DFT-Methode. Im Anhang B.3 ist eine detaillierte Beschreibung der Methode der kleinsten Quadrate und wie sie für die Ermittlung der genauen Komponenten einer Sinuskomponente angewendet wird, zu finden.

¹ -80 dB [Mas96]

Alle ermittelten Frequenzen, deren Amplituden und Phasen werden in einer Matrix \underline{S}_k der Dimension $M_k \times 3$ geschrieben, wobei M_k die Anzahl der in Paket k erfassten gültigen Maxima (Frequenzen) darstellt.

$$\underline{S}_k = \begin{bmatrix} f_1^k & \varphi_1^k & A_1^k \\ f_2^k & \varphi_2^k & A_2^k \\ \vdots & \vdots & \vdots \\ f_{M_k}^k & \varphi_{M_k}^k & A_{M_k}^k \end{bmatrix} \quad (4.65)$$

Die Anzahl der relevanten Frequenzen in den Analysebereichen vor der Lücke und nachher kann unterschiedlich sein. Somit haben die Sinuskomponenten-Matrizen \underline{S}_{k-1} aus dem vorangegangenen und \underline{S}_{k+1} aus dem nachfolgenden Analysebereich unterschiedliche Dimensionen M_{k-1} und M_{k+1} .

(2) Zuordnung der Sinuskomponenten zu Paaren

Nachdem die Sinuskomponenten vor und nach dem Paketverlust ermittelt wurden, müssen diese nun zu korrespondierenden Sinuskomponenten-Paaren zugeordnet werden. Korrespondierende Sinuskomponenten sind welche, die nur minimal voneinander abweichen.

In Bild 4.41 wird anhand eines Beispiels gezeigt, nach welchem Prinzip die Sinuskomponenten vor und nach der Lücke zu einem Paar zusammengefasst werden.

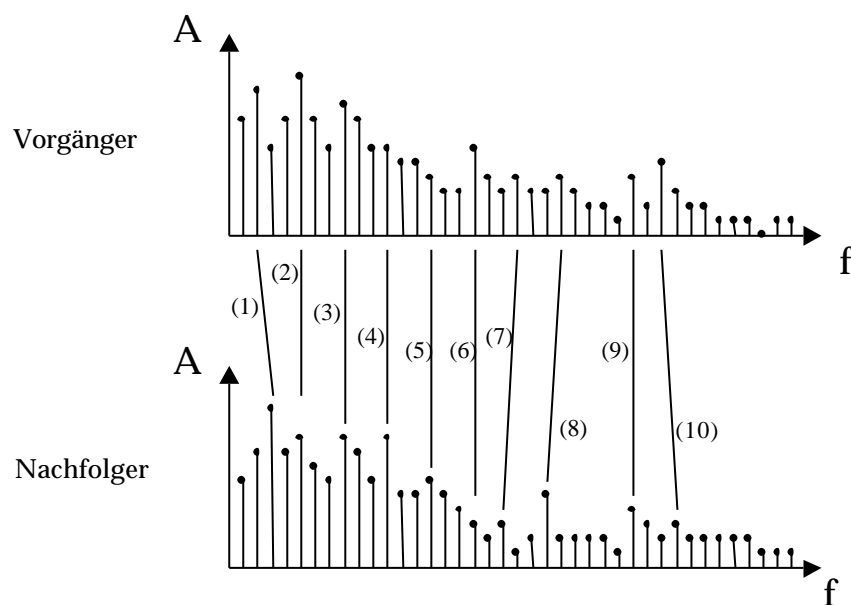


Bild 4.41: Zuordnung der Maxima zu Sinuskomponenten-Paare

- Zuerst werden alle lokalen Maxima gleicher Frequenz zu einem Paar zusammengefasst (2, 3 und 9). Diese entsprechen in den DFT-Vektoren den Maxima mit identischen Frequenz-Indizes.
- Als nächstes werden die Maxima, die zwar unterschiedliche Frequenz-Indizes, aber eine kleine Frequenzabweichung voneinander haben, betrachtet. Diese Abweichung kann z.B. eine Stufe (*bin*) sein (1, 7, 8 und 10).
- Zum Schluss müssen alle lokalen Maxima im Vorgänger- und Nachfolger-Analysebereiche beachtet werden, für die keine gegenüberliegende Sinuskomponente zugeordnet werden konnte. Für solche Maxima wird bei der späteren Analyse die Frequenz fortgesetzt unter Absenkung bzw. Anhebung der Amplitude von dem Niveau vor der Lücke auf das Niveau danach (4, 5 und 6). In diesem Fall sollte eine nachträgliche Korrektur der Frequenzen, Phasen und Amplituden aller Sinuskomponenten, die im Betragsspektrum keine Maxima darstellen, aber mit Maxima aus der gegenüberliegenden Seite der Lücke *getrackt* werden sollten, erfolgen **(3)**. Die Frequenzkorrektur erfolgt weiterhin mit der quadratischen Interpolation, und die der zugehörigen Phasen und Amplituden mit der sinusoidalen Regression.
- Alle Indizes der zu Paaren zusammengeordneten Sinuskomponenten werden in einer Matrix \underline{I} der Dimension $M_k \times 2$ geschrieben, wobei M_k die Anzahl der Sinuskomponenten-Paare bezeichnet.

$$\underline{I}[i] = \{\text{index}_{k-1}, \text{index}_{k+1}\}, \quad i = 0, \dots, M_k - 1. \quad (4.66)$$

Hierbei sind index_{k-1} und index_{k+1} die Indizes der Sinuskomponenten aus den Paketen $k-1$ und $k+1$, die zu einem Paar zusammengefasst wurden.

(4) Tracking und Synthese

Das Tracking-Modul muss die Sinuskomponenten-Paare aus dem Modul **(3)** zusammenführen.

Anhand der Sinuskomponenten-Indizes der Matrix \underline{I} werden zuerst die zu diesen Indizes korrespondierenden Frequenzen, Phasen und Amplituden aus den Matrizen \underline{S}_{k-1} und \underline{S}_{k+1} ermittelt.

Für die Synthese stehen insgesamt M_k Sinuskomponenten-Paare zur Verfügung, aus denen genau so viele Schwingungen erzeugt und zu einem Zeitsignal, das die Lücke ausfüllen soll, überlagert werden. Die Synthese erfolgt nach dem in Kapitel 2.5.3 erläuterten Prinzip. Allerdings müssen zusätzliche Nebenbedingungen berücksichtigt werden, die insbesondere für die Amplituden und Phaseninterpolation wichtig sind.

Im folgenden wird vorausgesetzt, dass sich die Länge der Analysebereiche von der Paketlänge unterscheiden kann. Es gilt:

L_a : Länge der Analysebereiche vor und nach der Lücke

L : Paketlänge

ΔL : Länge des Merging-Bereichs

Der Rekonstruktionsbereich muss also die Länge $L + 2 \cdot \Delta L$ haben.

Als Beispiel wird ein einzelnes Sinuskomponenten-Paar $(\{f_1, \varphi_{01}, A_1\}, \{f_2, \varphi_{02}, A_2\})$ betrachtet, wobei f_1, φ_{01} und A_1 die Frequenz, Phase und Amplitude der Sinuskomponente vor der Lücke und f_2, φ_{02} und A_2 die Frequenz, Phase und Amplitude der Sinuskomponente nach der Lücke darstellen.

Die Phase und Amplitude haben unterschiedliche Bezugspunkte: Die Amplitude bezieht sich auf dem Zentrum und die Phase auf dem Anfang des Analysebereichs (Bild 4.42).

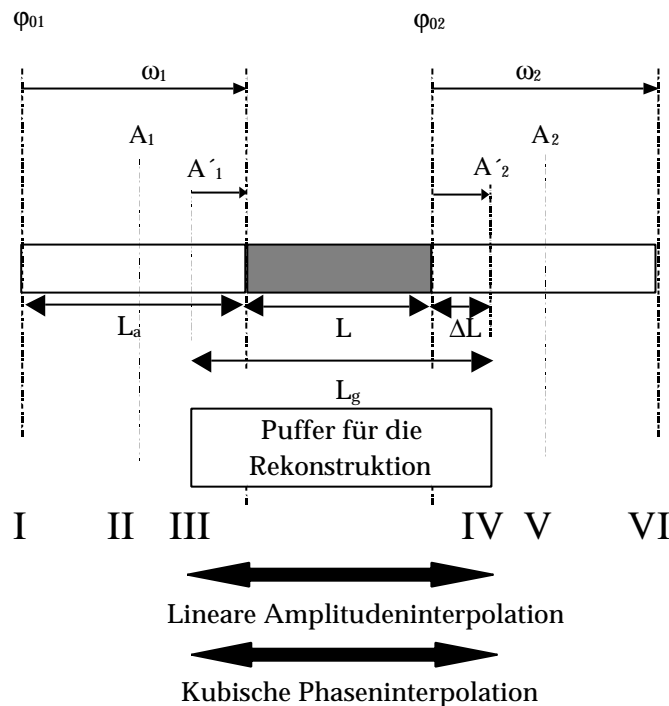


Bild 4.42: Bereiche und Parameter der linearen Amplituden- und kubischen Phaseninterpolation.

φ_{01} ist die Anfangsphase für den vorangegangenen Analysebereich,

φ_{02} ist die Anfangsphase für den nachfolgenden Analysebereich,

$\omega_1 = 2\pi f_1$ und $\omega_2 = 2\pi f_2$

Die Berechnung der Amplituden und Phasen weicht hier von der in Kapitel 2.5.3 aufgeführten Theorie ab. In Kapitel 2.5.3 wird bei der Parameterbestimmung der Amplituden- und Phaseninterpolation vorausgesetzt, dass das rekonstruierte Signal die Länge des analysierten Signals sowie dessen Anfangs- und Endbedingungen aufweist, was hier nicht zutrifft. Aufgrund des verlorenen Paketes und der Länge des Rekonstruktionsbereichs, die sich von der Paketlänge unterscheidet, gelten bei diesem Verfahren andere Randbedingungen.

Amplitudeninterpolation

Für die Amplitude wird eine lineare Interpolation von Punkt III bis Punkt IV angewendet. Die Berechnung erfolgt für alle M_k Sinuskomponenten. Die Amplituden A_1 und A_2 beziehen sich jeweils auf das Zentrum des Analysebereichs (II und V). Die Berechnung der Augenblicksamplitude erfolgt dann gemäß:

$$A_i(n) = A_1' + dA \cdot n \quad n = 0 \dots L_g - 1 \quad (4.67)$$

$$A_1' = A_1 + \left(\frac{L_a}{2} - \Delta L\right) \cdot dA \quad (4.68)$$

$$dA = \frac{A_2 - A_1}{L + L_a} \quad (4.69)$$

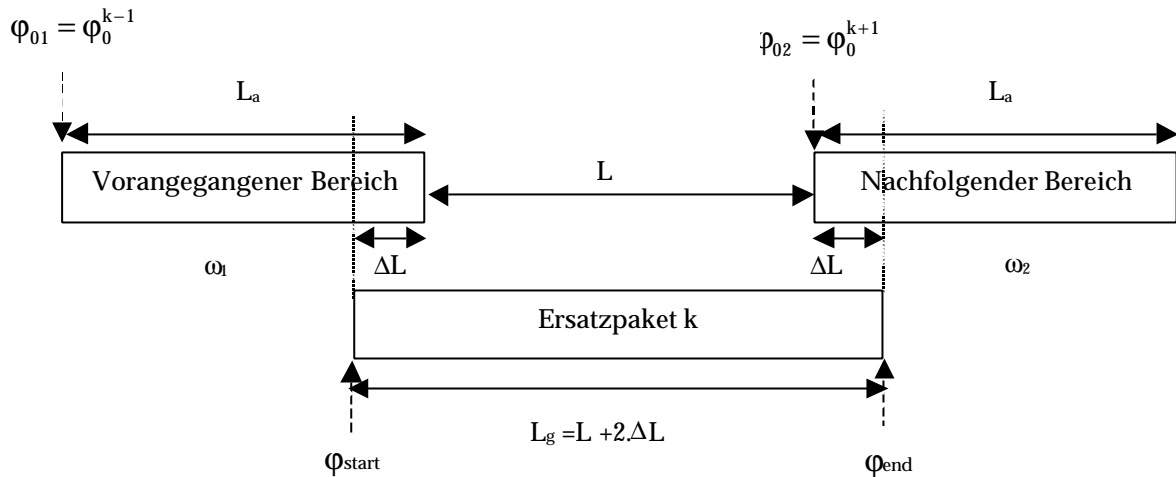
Phaseninterpolation

Die kubische Interpolation der Phase wird durch die Berechnung der Augenblicksphasenfunktion $\varphi_i(n)$ von Punkt III bis Punkt IV erreicht. Die Berechnung erfolgt für alle M_k Sinuskomponenten.

Für den kubischen Verlauf der Phase der i -ten Sinuskomponente im Substitutionspaket k gilt allgemein:

$$\varphi_i^k(n) = \varepsilon + \kappa \cdot n + \alpha \cdot n^2 + \beta \cdot n^3, \quad n=0 \dots L_g \quad (4.70)$$

Im folgenden wird zur Veranschaulichung die Berechnung der Phase für eine Sinuskomponente durchgeführt und deswegen auf den Index i verzichtet.

**Bild 4.43:** Phaseninterpolation

Für die Phase in den Analysebereichen vor und nach der Lücke gilt allgemein:

$$\varphi(n) = \varphi_0 + n \cdot \omega, \quad n = 0 \dots L_a - 1 \quad (4.71)$$

Für die Phase im Substitutionspaket gelten folgende Randbedingungen:

$$\varphi_{\text{start}} = \varphi_0^{k-1} + (L_a - \Delta L) \cdot \omega_1, \quad \omega_1 = 2\pi f_1 \quad (4.72)$$

$$\varphi_{\text{end}} = \varphi_0^{k+1} + (\Delta L - 1) \cdot \omega_2 \quad (4.73)$$

$$\varphi^k(0) = \varepsilon = \varphi_{\text{start}} \quad (4.74)$$

$$\varphi^k(L_g) = \varphi_{\text{end}} + \omega_2 = \varphi_{\text{start}} + \kappa \cdot L_g + \alpha \cdot L_g^2 + \beta \cdot L_g^3 \quad (4.75)$$

Die erste Ableitung der kubischen Phase hat einen quadratischen Verlauf und ist wie folgt definiert:

$$\dot{\varphi}(n) = \kappa + 2\alpha \cdot n + 3\beta \cdot n^2 = \omega \quad (4.76)$$

$$\dot{\varphi}(0) = \kappa = \omega_1 \quad (4.77)$$

$$\dot{\varphi}(L_g) = \omega_1 + 2\alpha \cdot L_g + 3\beta \cdot L_g^2 = \omega_2 \quad (4.78)$$

Für die Berechnung der kubischen Phase sind also nur noch die Parameter α und β unbekannt. Für die Bestimmung dieser Parameter ist das folgende Gleichungssystem zu lösen:

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{bmatrix} \frac{3}{L_g^2} & -\frac{1}{L_g} \\ -\frac{2}{L_g^3} & \frac{1}{L_g^2} \end{bmatrix} \cdot \begin{bmatrix} \varphi_{\text{end}} + \omega_2 - \varphi_{\text{start}} - \omega_1 \cdot L_g \\ \omega_2 - \omega_1 \end{bmatrix} \quad (4.79)$$

Auf die Berechnung des Faktors x (Gleichung 2.25) und daraus des Parameters M , der für eine maximale Flachheit des Frequenzverlaufs sorgen soll, wird hier verzichtet. Das Problem der optimalen Phasenkontinuität wird hier anders gehandhabt. Bei der Berechnung der Augenblicksphase muss drauf geachtet werden, dass die Phasendifferenz zweier aufeinanderfolgenden Phasen minimal gehalten wird. Interessant ist nur die Phasendifferenz, die kleiner π ist. Wird die Phasendifferenz größer π , so muss die neue Phase um den Faktor 2π korrigiert werden (*phase unwrapping*). Diese Bedingung wird erfüllt, wenn:

$$0 \leq \varphi_{\text{end}} + \omega_2 - \varphi_{\text{start}} - \omega_1 \cdot L_g \leq \pi \quad (4.80)$$

Für jedes Sinuskomponenten-Paar werden die interpolierten Amplituden und Phasen abtastwertweise ermittelt und daraus die Schwingung $y_i(n)$ synthetisiert, wobei es gilt:

$$y_i(n) = A_i(n) \cdot \cos(\varphi_i(n)), \quad n = 0 \dots L_g - 1 \quad (4.81)$$

Die einzelnen Schwingungen aller M_k Sinuskomponenten-Paare werden zum Schluss zu einer komplexen Schwingung überlagert.

(5) Addition eines Rauschsignals

Nachdem in den letzten Modulen alle wichtigen Frequenzen aus den DFT-Vektoren entnommen und anschließend getrackt und zu einer komplexen Schwingung synthetisiert wurden, bleiben nun in den DFT-Vektoren noch Koeffizienten, die für die Analyse und Synthese irrelevant waren. Diese „Rest-Koeffizienten“ werden als Rauschkomponenten des Audiosignals angesehen und sollten bei dem Rekonstruktionssignal auch berücksichtigt werden. Die einfache Rücktransformation der nicht benutzten DFT-Koeffizienten entspricht gewissermaßen eine Paketwiederholung nur für Rauschkomponenten. Damit wäre es aber möglich, dass sich Spitzen, die im vorangegangenen oder nachfolgenden Analysebereich auftauchen, im rekonstruierten Signal wiederfinden. Die Färbung der Übertragungsfunktion (Betragsspektrum) der Rauschkomponenten mit einem weißen Rauschen liefert bessere Ergebnisse, da damit die Spitzen auf jeden Fall verschmiert werden.

In diesem Verfahren wird also das Betragsspektrum der Rauschkomponenten¹ mit dem Spektrum eines Rauschens multipliziert und in den Zeitbereich zurücktransformiert. Es er-

¹ Hier werden die Rauschkomponenten vor der Lücke verwendet.

gibt sich ein Rauschsignal $e(t)$, das mit der komplexen Schwingung aus dem letzten Modul zu einem Rekonstruktionssignal überlagert wird.

(6) Merging und Ausgabe

Die Übergänge zwischen dem Rekonstruktionssignal und den angrenzenden Bereichen werden vor der Ausgabe geglättet (*Merging*). Die *Merging*-Länge ΔL kann beliebig sein und darf aber 50 % der Paketlänge nicht überschreiten.

Signalverläufe

Bild 4.44 zeigt die Signalverläufe und die dazugehörigen Spektrogramme des Audiosignals aus einem Einzelnstrument-Musik-Titel (Orgel). Es handelt sich hier um einen Ausschnitt der Länge 5000 Abtastwerte (113 ms) aus dem linken Kanal des Original-Stereosignals (Bild 4.44-a). Bild 4.44-b) stellt das Signal des linken Kanals mit einem Paketverlust dar. Die Paketlänge beträgt hierbei 1024 Abtastwerte (22 ms).

Hier wird auf die Darstellung des rechten Kanals verzichtet, weil das Verfahren unabhängig vom Nachbarkanal funktioniert.

Das Rekonstruktionssignal aus Bild 4.44-c) zeigt, wie das Verfahren die Frequenzen beiderseits der Lücke miteinander verbindet. Diese Frequenzlinien sind ein Hinweis auf ein näherungsweise periodisches Signal. Die fehlenden Frequenzen in dem Substitutionssignal sind Rauschkomponenten, die in diesem Beispiel nicht rekonstruiert wurden.

Subjektiv ist hier kaum Unterschied zwischen Rekonstruktions- und Originalsignal wahrnehmbar.

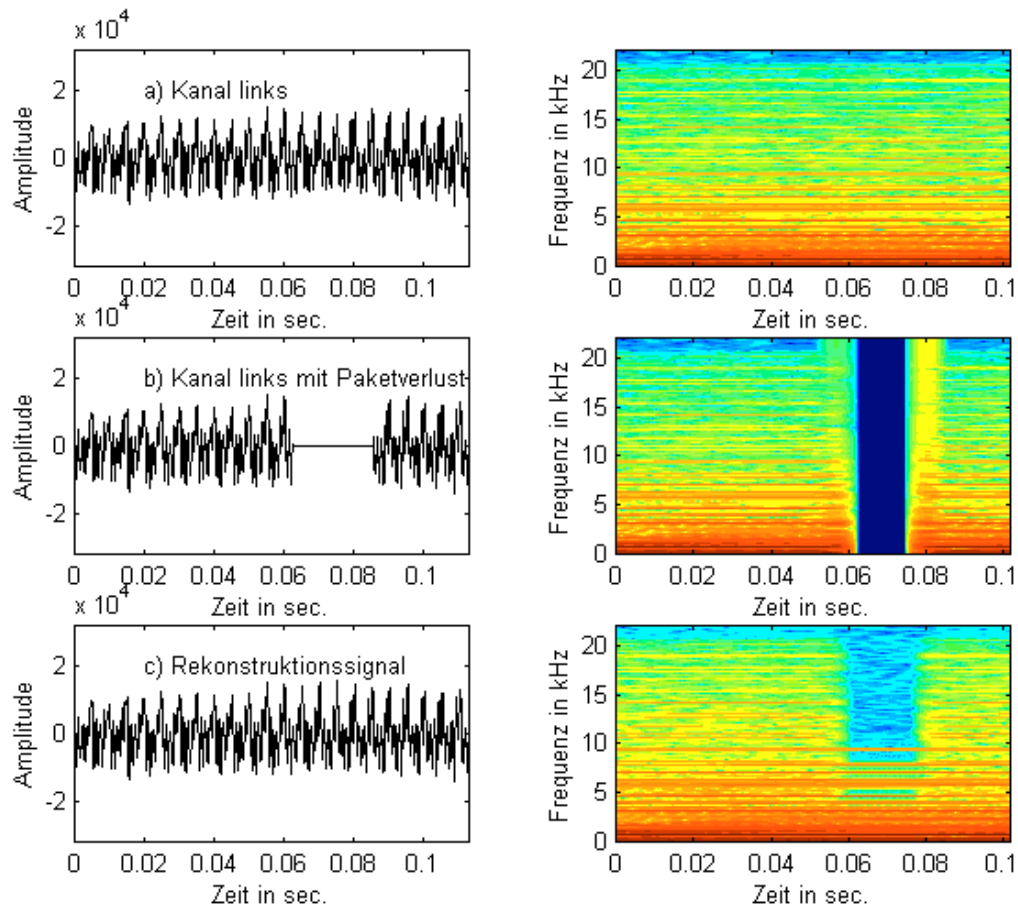


Bild 4.44: Paketsubstitution mit Frequenz-Tracking; Signalverläufe links und die dazugehörigen Spektrogramme rechts. Originalsignal aus einem Einzelinstrument-Musik-Titel (5000 AW: 113 ms); a) Linker Kanal des Stereosignals; b) Linker Kanal mit Paketverlust. Die Paketlänge beträgt hierbei 1024 Abtastwerte (22 ms); c) Mit Frequenz-Tracking rekonstruiertes Signal.

Bild 4.45 zeigt die Signalverläufe und die dazugehörigen Spektrogramme des Audiosignals aus einem Pop-Musik-Titel. Es handelt sich hier um einen Ausschnitt der Länge 5000 Abtastwerte (113 ms) aus dem linken Kanal des Original-Stereosignals (Bild 4.44-a). Bild 4.44-b) stellt das Signal des linken Kanals mit einem Paketverlust dar. Die Paketlänge beträgt hierbei 1024 Abtastwerte.

Der Signalverlauf in diesem Signalausschnitt ist nicht periodisch und weist daher keine eindeutigen Frequenzlinien auf. Es ist dem Verfahren trotzdem gelungen ein Substitutionssignal zu rekonstruieren, dessen Spektrum eine Ähnlichkeit mit dem Spektrum des Originalsignals aufweist. Die fehlenden Frequenzen in dem Substitutionssignal stellen Frequenzen niedriger Amplituden dar, die unterhalb der festgelegten Schwelle liegen und so nicht berücksichtigt werden. Diese haben keinen Einfluss auf die subjektive Audioqualität des Rekonstruktionssignals.

Auch wenn der Signalverlauf des Substitutionssignals an der Fehlerstelle anders ist als der des Originalsignals, ist hier subjektiv nur ein minimaler Unterschied zwischen Substitutions- und Originalsignal wahrnehmbar.

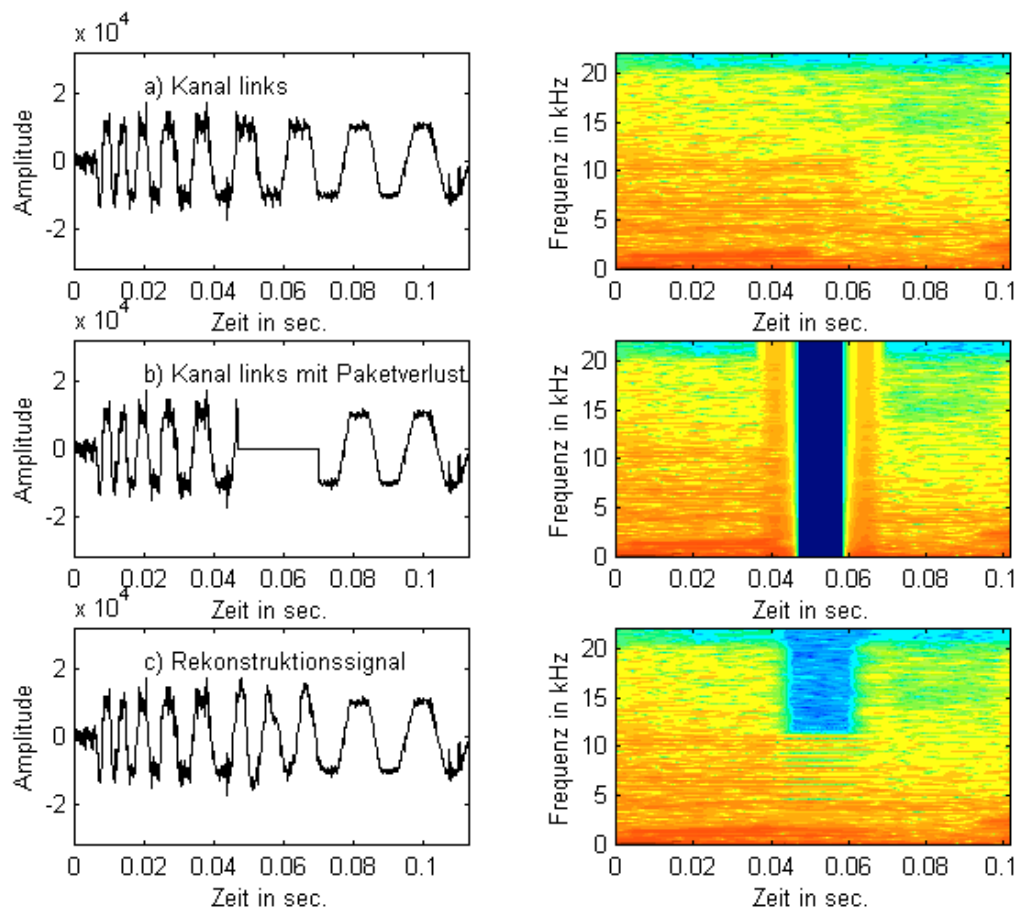


Bild 4.45: Paketsubstitution mit Frequenz-Tracking; Signalverläufe links und die dazugehörigen Spektrogramme rechts. Originalsignal aus einem Pop- -Musik-Titel (5000 AW: 113 ms); a) Linker Kanal des Stereosignals; b) Linker Kanal mit Paketverlust. Die Paketlänge beträgt hierbei 1024 Abtastwerte (22 ms); c) Mit Frequenz-Tracking rekonstruiertes Signal.

Bemerkungen:

- Bei dem vorgeschlagenen Verfahren bestehen die Abtastwerte des vorangegangenen Analysebereichs aus den letzten gepufferten L_a Abtastwerten des Ausgangssignals unmittelbar vor der Lücke. Es ist notwendig, das Ausgangssignal und nicht das Originalsignal zu puffern, um Diskontinuitäten zwischen dem fehlerfreien und dem rekonstruierten Signal zu vermeiden. Diese Diskontinuitäten bewirken eine Verfälschung der analysierten Frequenzen und somit eine Verschlechterung der Audioqualität des rekonstruierten Signals. Mit dicht aufeinanderfolgenden Paketverlusten ist es wahrscheinlicher, dass

der vorangegangene Analysebereich rekonstruierte Signale enthält, was zu einer schlechteren Rekonstruktion führt.

- Bündelfehler können prinzipiell wie Einzelfehler rekonstruiert werden, da der Analysebereich vor der Lücke aus dem Ausgangssignal besteht. Problematisch ist nur, den nachfolgenden Analysebereich für das erste Paket des Bündels zu definieren. Die besten Ergebnisse werden erzielt, wenn alle Paketverluste des Bündels zu einer größeren Lücke zusammengefasst werden und diese wie ein Einzelpaketverlust rekonstruiert wird.
- Ein erster Vorschlag zur **Extrapolation** von fehlenden digitalen Audio-Daten basierend auf der sinusoidalen Analyse/Synthese wurde 1994 von R.C. Maher vorgeschlagen [Mah94]. In seinem Verfahren erfolgte die Analyse von Audio-Daten beiderseits der Lücke mit mehreren kurz aufeinanderfolgenden DFTs, auch *running DFT* genannt. Es wurden dann für alle DFT-Vektoren die Maxima gebildet und nach dem MQ-Analyse-Modell [McA86] ausgewertet. Mit der *running DFT* ist es möglich, Spuren von Frequenzlinien (und wie sie sich über der Zeit entwickeln) zu ermitteln (Spektrogramm). Aufgrund des damit verbundenen hohen Rechenaufwands ist das Verfahren nur für *Offline*-Anwendungen wie z.B. Audiorestauration geeignet [Mah94]. Eine Anwendung dieses Verfahrens im *Online*-Betrieb z.B. für die Rekonstruktion von Paketverlusten in Audio-Daten während der Übertragung über das Internet ist nicht denkbar.

Das Verfahren von Maher unterscheidet sich von dem in dieser Arbeit vorgeschlagenen FTR-Substitutionsverfahren im wesentlichen in den folgenden Punkten:

- Im Gegensatz zu FTR ist das Mahersche Verfahren nicht für die Rekonstruktion von Paketverlusten in Audiosignalen geeignet. Maher geht von einer konstanten Frequenzänderung im Laufe der Zeit aus, mit der die Frequenz am Ende der Lücke zuerst geschätzt wird (Extrapolation). Die Schätzung erfolgt aufgrund der Annahme, dass sich die Frequenzänderung in dem fehlenden Signalausschnitt fortsetzt. In dieser Arbeit basiert das FTR-Verfahren auf einer Interpolation der Frequenzen. Mit FTR werden keine Frequenzlinien in der Vergangenheit ermittelt. Es werden nur die einzelnen Frequenzen vor und nach der Lücke miteinander verbunden. Man erhofft sich dabei, dass keine Diskontinuitäten im Frequenzverlauf entstehen, die das Ohr wahrnehmen könnte.
- Frequenzlinien, die am Ende der Lücke nicht fortgesetzt werden können, wurden mit *Death* bezeichnet und so in der Amplitude und Phase auf Null abgesetzt. In dieser Arbeit werden einzelne Maxima aus der Vergangenheit, die mit keinem Maximum aus der Zukunft getrackt werden können, mit den Sinuskomponenten aus der Zukunft verbunden, welche die gleiche Frequenz aufweisen. Es wird für diese Sinuskomponenten eine Amplituden- und Phasen-Ermittlung durchgeführt. Die Frequenz wird innerhalb der Lücke fortgesetzt, unter Berücksichtigung der Entwicklung deren Amplitude und Phase.
- Die Ermittlung der Frequenzen, Phasen und Amplituden erfolgt nach dem Maherschen Verfahren direkt aus den DFT-Koeffizienten. Mit dem FTR-Verfahren wird abgesehen von der quadratischen Interpolation zur feineren Bestimmung der Frequenz, die Ampli-

tude und Phase dieser Frequenz mit dem Verfahren der minimalen Fehlerquadrate ermittelt.

- Eine Modifikation der Parameter für die Phaseninterpolation ist in dem Verfahren von Maher nicht nötig. Maher hält sich damit an die in [McA86] vorgeschlagene Phaseninterpolation. Hier entspricht die Länge des Rekonstruktionssignals genau der Länge der Lücke (Maher verwendet kein Merging). Aufgrund der von der Lücken-Länge abweichende Länge des Rekonstruktionssignals mussten bei dem FTR-Verfahren die Parameter der Phasen- und Amplitudeninterpolation optimiert werden.
- Die Extrapolation von Audio-Daten nach dem Verfahren von Maher beschränkt sich auf die Betrachtung der sinusoidalen Komponenten des Signals. Hier spielen die Rauschkomponenten keine Rolle.

5 Vergleich der untersuchten Substitutionsverfahren

Bei der Untersuchung von Paketverlusten in Audiosignalen sollen verschiedene Substitutionsverfahren bei unterschiedlichen Paketverlustwahrscheinlichkeiten und Paketlängen verglichen werden. Die Beurteilung der in Kapitel 4 untersuchten Substitutionsverfahren bei Paketverlusten beruht auf der Bewertung der Audioqualität des rekonstruierten Audiosignals. Die Bewertungsergebnisse werden miteinander verglichen, um eine Aussage über der Effizienz eines Substitutionsverfahrens zu geben.

Die Audioqualität des rekonstruierten Audiosignals hängt von den folgenden Faktoren ab:

- dem verwendeten Substitutionsverfahren und dessen Komplexität,
- der Paketverlustwahrscheinlichkeit P_v ,
- der Statistik der Fehlerverteilung (Einzel- oder Bündelfehler)
- und von der Paketlänge L .

Es existieren sowohl objektive als auch subjektive Bewertungskriterien der Tonqualität von Audiosignalen. Einige davon werden in den nächsten Abschnitten näher erläutert.

5.1 Objektive Bewertung

Objektive Bewertungskriterien werden mathematisch gefasst und haben in der Regel eine geringe Aussagekraft hinsichtlich der subjektiven Audioqualität. Für eine relative Bewertung ein und desselben Substitutionsverfahrens mit unterschiedlichen Parametern wie die Paketverlustwahrscheinlichkeit P_v oder die Paketlänge L können mit objektiven Messwerten durchaus zuverlässige Aussagen gewonnen werden.

In [Qua88] wurde untersucht, in wieweit objektive Qualitätsmaße für die Bewertung der empfundenen **Sprachqualität** geeignet sind. Es wurde ein Korrelationsfaktor ρ definiert, der die Korrelation zwischen den objektiven und subjektiven Messergebnissen beschreibt (Tabelle 5.1).

OBJEKTIVES QUALITÄTSMAB	KORRELATIONSFAKTOR ρ
SNR	0,27
SNR _{seg}	0,77
SNR _{f₀-seg}	0,93
LPC-Koeffizienten	0,06

Tabelle 5.1: Korrelation der objektiven mit den subjektiven Ergebnissen [Qua88].

Die größte Korrelation zwischen den objektiven und subjektiven Messergebnissen der Audioqualität liefert die frequenzgewichtete Variante des segmentorientierten SNR. Sie wird mit SNR_{f₀-seg} bezeichnet. Das segmentorientierte SNR, der hier mit SNR_{seg} bezeichnet wird, liefert auch mit seinem Korrelationsfaktor von 0,77 aussagekräftige Ergebnisse.

5.1.1 Signal-to-Noise Ratio

Ein Qualitätsmaß, um z.B. die Leistungsfähigkeit eines Substitutionsverfahrens einzuschätzen, ist der Signal-Rausch-Abstand SNR (*Signal-to-Noise Ratio*). Die Leistung des Originalsignals $x(n)$ wird zu der Leistung des Ersatzsignals $x_e(n)$ in Beziehung gesetzt und als logarithmisches Verhältnis angegeben.

$$\text{SNR} = 10 \log \frac{\sum_{n=0}^{L-1} [x(n) - \mu_x]^2}{\sum_{n=0}^{L-1} [x(n) - x_e(n)]^2} \text{ dB} \quad (5.1)$$

L: Signallänge in Abtastwerte.

n: Abtastwertindex.

μ_x : Mittelwert des Signals $x(n)$, der bei mittelwertfreie Signale entfällt.

Der klassische SNR stellt kein äquivalentes Maß zur Bestimmung der subjektiven Audioqualität dar. Es besteht nur eine geringe Korrelation zur subjektiven Beurteilung. Alle Fehler im Zeitsignal, also Unterschiede der Signalverläufe, werden gleich gewichtet, obwohl sie zu unterschiedlich starken Störungen der Audiowiedergabe führen.

5.1.2 Segment-SNR

In früheren Forschungsarbeiten wurde gezeigt, dass das klassische SNR schlecht geeignet ist für die Schätzung der subjektiven Sprachqualität [McD78], [Nol78].

Um zu berücksichtigen, dass die Leistung eines Audiosignals starken Schwankungen unterliegt, wurde als Verbesserung des klassischen SNR eine segmentorientierte Variante eingeführt (SNR_{seg}) [Nol74]. Es wird für jedes Segment ein SNR berechnet und zum Schluss der arithmetische Mittelwert gebildet (Gleichung 5.2). Die Segmente ergeben sich durch Fensterung des Audiosignals. Die typischen Fensterlängen betragen 15 bis 20 ms [Qua88]. In [Jay84] wird die Länge auf 16 ms¹ festgesetzt.

$$\text{SNR}_{\text{seg}} = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log \left[\sum_{n=L_m}^{L_m+L-1} \frac{x^2(n)}{[x(n) - x_e(n)]^2} \right] \text{dB} \quad (5.2)$$

M: Anzahl der Segmente im Signal.

L: Segmentlänge.

Diese Definition des Segment-SNR bringt folgende Probleme mit sich:

- Stille enthaltene Originalpakete können die Messergebnisse stark verzerren. Segmente, in denen das Audiosignal aus Werten nahe Null besteht, müssen aus der Berechnung ausgenommen werden. Dies kann realisiert werden, in dem man für alle Segmente die Energie berechnet. Bei der SNR-Berechnung werden nur die Segmente berücksichtigt, deren Energien eine festgesetzte untere Grenze² nicht unterschreiten [Jay84].
- Um einer weiteren Verzerrung des Messergebnisses durch substituierte Pakete, welche die Information des Originalpaketes enthalten, zu vermeiden, muss eine obere Grenze³ gesetzt werden. Allerdings wird durch die notwendige Wahl dieser Grenze das Ergebnis stark beeinflusst.

Der SNR_{seg} wurde in vielen Fällen als gut geeignet für die Schätzung der Sprachqualität bewertet [Nol74] [McD78].

Eine sehr erfolgreiche Erweiterung des Segment-SNR bietet die frequenzgewichtete Variante, bei der die Audioprobe in einzelnen Bändern analysiert wird.

¹ 16 ms entsprechen bei 8kHz Sprache 128 und bei 44,1 kHz Audio 706 Abtastwerten. Für die Auswertung von Audiosignalen nach der Substitution wird die Paketlänge 700 AW genommen.

² Die untere Grenze der Kurzzeit-Signal-Energie beträgt -40 dB.

³ Die obere Grenze wird bei 50 dB festgesetzt

5.1.3 Frequenzgewichteter Segment-SNR

Das menschliche Ohr reagiert sehr unterschiedlich auf spezielle Frequenzbereiche. So ist es z.B. im Bereich von 2kHz bis 5kHz sehr empfindlich, während höhere und tiefere Frequenzen schwächer wahrgenommen werden. Darum wurde eine Variante des SNR_{seg} entwickelt, die diese Eigenschaften berücksichtigt [Bar80]. Sie wird im folgenden mit $\text{SNR}_{f_{\omega\text{-seg}}}$ bezeichnet.

Das Audiosignal wird in 20 Teilbänder zerlegt, deren Breite mit steigender Frequenz zunimmt [Zwi61]. Pro Segment wird das SNR für verschiedene Frequenzbereiche ermittelt und, der Hörkurve angepasst, gewichtet ($\omega_{j,k}$).

$$\text{SNR}_{f_{\omega\text{-seg}}} = \frac{1}{M} \sum_{j=0}^{M-1} 10 \log \left[\frac{\sum_{k=1}^K \omega_{j,k} 10 \log [E_{x,k}(m_j) / E_{\epsilon,k}(m_j)]}{\sum_{k=1}^K \omega_{j,k}} \right] \text{dB} \quad (5.3)$$

- M: Anzahl der Segmente
- K: Anzahl der Frequenzbänder
- $\omega_{j,k}$: Gewichtungsfaktor
- E_x : Energie des ungestörten Segments
- E_{ϵ} : Energie des gestörten Segments

Die Energie im Frequenzbereich kann nach dem *Parseval-Theorem* wie folgt berechnet werden:

$$E_U = \frac{1}{N} \sum_{k=0}^{N-1} |U_{\text{DFT}}(k)|^2 \quad (5.4)$$

- N: Anzahl der Stützstellen der DFT

5.1.4 Objective Difference Grade

Eine Beurteilung der Qualität von Audiosignalen ist mit konventionellen Messverfahren (z.B. Störabstand, Verzerrungen etc.) nur unzureichend möglich. Da eine Qualitätsbewertung durch subjektive Hörtests einen Aufwand erfordert, der in vielen Fällen nicht vertretbar ist, werden objektive Messverfahren benötigt, die auf Gehörmodellen für quasistationäre Signale basieren.

Eine Bewertung der untersuchten Substitutionsverfahren wird in dieser Arbeit hauptsächlich mit Hilfe eines im Rahmen eines Forschungsprojektes im Fachgebiet Fernmeldetechnik ent-

wickelten Programms („DIXI“) durchgeführt. DIXI stellt eine Benutzeroberfläche für das Messverfahren PEAQ zur Verfügung. PEAQ ist die von der ITU-R empfohlene Meßmethode zur *objektiven* Messung der empfundenen Tonqualität und verwendet zu wesentlichen Teilen an der TU Berlin erarbeitete Modellierungsansätze.

Als Maß für die objektive Qualitätsbewertung von Audiosignalen liefert PEAQ einen ODG-Wert (**O**bjective **D**ifference **G**rade). ODG-Werte liegen zwischen -4 und 0 und haben die folgende Bedeutung:

ODG	PODG	AUDIO-QUALITÄT
0.0	5.0	Störung nicht wahrnehmbar
-1.0	4.0	Wahrnehmbar aber nicht störend
-2.0	3.0	kaum störend
-3.0	2.0	störend
-4.0	1.0	sehr störend

Tabelle 5.2: ODG-Skala [Thi98]

Im Angleichung an den bei *subjektiver* Qualitätsbewertung üblicherweise verwendeten *Mean Opinion Score* (MOS) (Kapitel 5.2) lässt sich eine aus der objektiven Bewertungsgröße ODG abgeleitete Größe $PODG = 5 + ODG$ (*Positives ODG*) definieren, die Werte zwischen 1 und 5 liefert (vergleichbar mit MOS aus Tabelle 5.3).

5.2 Subjektive Bewertung mit MOS-Messung

Ziel einer Paketsubstitution ist es, das durch Paketverluste gestörte Audiosignal so zu rekonstruieren, dass es für den Hörer angenehmer klingt. Es ist deswegen hier zwingend erforderlich, die Audioqualität durch subjektive Messungen mittels Hörtests zu bewerten. Die subjektive Bewertung der Audioqualität erfolgt mit Hilfe von Hörvergleichen. Der Gesamteindruck des Testsignals wird von Testpersonen bewertet. Aufgrund des immensen Aufwandes und der damit verbundenen Kosten können solche Hörtests aber nicht beliebig oft durchgeführt werden.

Bei der Bestimmung des MOS bewertet jede Testperson die empfundene Audioqualität mit einem Wert aus einer Skala von 1 bis 5 (*Five Grade Impairment Scale*)(s. Tabelle 5.3).

MOS	AUDIO-QUALITÄT
5.0	Störung nicht wahrnehmbar
4.0	Wahrnehmbar aber nicht störend
3.0	kaum störend
2.0	störend
1.0	sehr störend

Tabelle 5.3: MOS-Skala (*Five Grade Impairment Scale*)

Ausgewertet wird die Audioqualität üblicherweise als MOS, also als arithmetisches Mittel der abgegebenen Bewertungen [San95].

MOS-Bewertungen sind akzeptierte Qualitätsmaße für Sprach- und Audiosignale [Yon92], [Pap87], [Suz98]. Sie haben allerdings auch einige Nachteile:

- Manche Testpersonen bewerten grundsätzlich schlechter, bzw. besser als andere, wodurch sich eine hohe Streuung der Resultate ergeben kann.
- Hoher Zeitaufwand, da eine Bewertung durch viele Testpersonen notwendig ist.
- Starke Abhängigkeit von den jeweiligen Testbedingungen: Nebengeräusche, Qualität der Audio-Hardware (A/D-Wandler), Güte der Kopfhörer/Lautsprecher.

5.3 Untersuchungsbedingungen

Verwendete Bewertungskriterien

Die Beurteilung aller in Kapitel 4 vorgestellten Verfahren erfolgt anhand objektiver Messwerte und informeller Hörvergleiche. Bei der objektiven Bewertung werden PODG und SNR_{seg} Werte ermittelt und verglichen, diese haben allerdings in der Regel eine beschränkte Aussagekraft in Bezug auf die empfundenen Audioqualität. Besser geeignet ist die subjektive Bewertung der Audioqualität. Diese erfolgt anhand von Hörtests, die auf einer Bewertung des Gesamteindrucks jedes Testsignals durch die Testpersonen beruht. Ein Testsignal kann entweder das Originalsignal, das Signal mit Lücken oder das Substitutionssignal sein. Die Hörtests wurden von 15 Testpersonen durchgeführt. Die Audio-Wiedergabe erfolgte am PC über einen hochwertigen Kopfhörer (AKG K240).

Paketverlust-Statistik

Zur Simulation von Paketverlusten wurde ein Modul implementiert, das die Aufgabe hat, Paketverluste mit einer vorgegebenen Paketverlustrate und Statistik der Paketverluste zu erzeugen. Es können sowohl statistisch unabhängige (Einzelpakete) als auch bündelförmige Paketverluste mit einer beliebigen Anzahl von Paketen pro Bündel simuliert werden. In [Bol95] wurde gezeigt, dass im Internet hauptsächlich ein Verlust von einzelnen oder von zwei aufeinander folgenden Paketen auftritt. Diese Behauptung wurde in [Boy98] für die Übertragung von MPEG-Daten über ATM-Zellen auch bestätigt. Aus diesem Grund werden in dieser Arbeit Einzelfehler und Bündelfehler mit zwei bis maximal drei Paketen pro Bündel berücksichtigt.

Untersuchte Paketlängen

Die simulierten Paketlängen können beliebig groß sein. In dieser Arbeit beschränkte sich die Untersuchung auf folgende Paketlängen: 256, 512 und 1024 Abtastwerte. Diese entsprechen bei mit 44,1 kHz abgetasteten Audiosignalen einer Dauer von 5,8; 11,61 und 22,22 ms.

Länge des Merging-Intervalls

Nach der Rekonstruktion eines Substitutionssignals, welches die durch Paketverluste entstandenen Lücken im Audiosignal verdecken soll, werden die Übergänge zwischen dem Ersatzpaket und den angrenzenden Paketen durch Überlappung einiger Abtastwerte geglättet. Dies bewirkt eine Verminderung der Knacks- und Knistergeräusche mit hochfrequenten Anteilen.

In [Goo86] und [Clü98a] wurde das Merging-Verfahren (s. Kap. 4.3.7) auf Sprachsignale angewendet. Es wurde ein Überblendintervall von 1 ms verwendet, was bei 8 kHz Sprachsignale 8 Abtastwerten entspricht. Dies würde bei Audiosignalen mit der Abtastfrequenz von 44,1 kHz ungefähr 44 Abtastwerten bedeuten. Mit dem Überblendintervall von 1 ms wurde eine leichte Verbesserung der Sprachqualität nach der Substitution festgestellt. Kürzere Intervalle führten zu schlechteren Ergebnissen, während mit längerer Intervalle keine weitere Verbesserung möglich war.

Die Untersuchung der Auswirkung der Merging-Länge auf die nach der Paketsubstitution empfundene Audioqualität hat gezeigt, dass bei den verwendeten Paketlängen eine Merging-Länge über ungefähr 10 Prozent der Paketlänge optimale Ergebnisse liefert.

5.4 Verwendetes Audiomaterial

Die implementierten Substitutionsverfahren werden mit Audiosignalen in CD-Qualität untersucht. Die Audiodateien haben eine Auflösung von 16 Bit und eine Abtastfrequenz von 44,1 kHz. Es wurden sowohl Mono- als auch Stereosignale untersucht.

Um Substitutionsverfahren auf ihre Robustheit und Effizienz hin zu testen, sollten bei der Untersuchung Audiodateien verwendet werden, die unterschiedliche Merkmale und Eigenschaften aufweisen. Die Simulationen erfolgten in dieser Arbeit überwiegend mit drei Musik-Titeln (s. Tabelle 5.4).

Musik-Titel	Audiosignal	Eigenschaften
Titel 1	Einzelnes Instrument (Orgel)	Der Signalverlauf eines Kanals enthält häufig Segmente mit näherungsweise periodischem Verlauf. Die Signalverläufe in beiden Kanälen haben kaum Ähnlichkeit miteinander Bild 5.2.
Titel 2	Klassische Musik (Beethoven)	Keine Ähnlichkeit des Monosignals eines Kanals weder mit seiner Vergangenheit noch mit dem Nachbarkanal.
Titel 3	Pop Musik (Marusha)	Hohe Ähnlichkeit zwischen den Kanälen des Stereosignals (linker und rechter Kanal).

Tabelle 5.4: Verwendetes Audiomaterial

Titel 1: Einzelnes Instrument

Bild 5.1 zeigt einen Ausschnitt von 1 Sekunden Dauer (44100 Abtastwerte) aus dem Stereosignal von Titel 1 und die dazugehörige Spektraldarstellung. Die Spektren beider Kanäle enthalten über längere Zeitabschnitte konstante (durchgezogene) Frequenzlinien, die auf einen näherungsweise periodischen Signalverlauf in diesen Abschnitten hindeuten.

Bild 5.2 zeigt einen Beispiel eines solchen periodischen Verlaufs im Signal. Es handelt sich hierbei um ein Segment der Länge 5000 Abtastwerte (ungefähr 113 ms) aus der Stereodatei aus Bild 5.1.

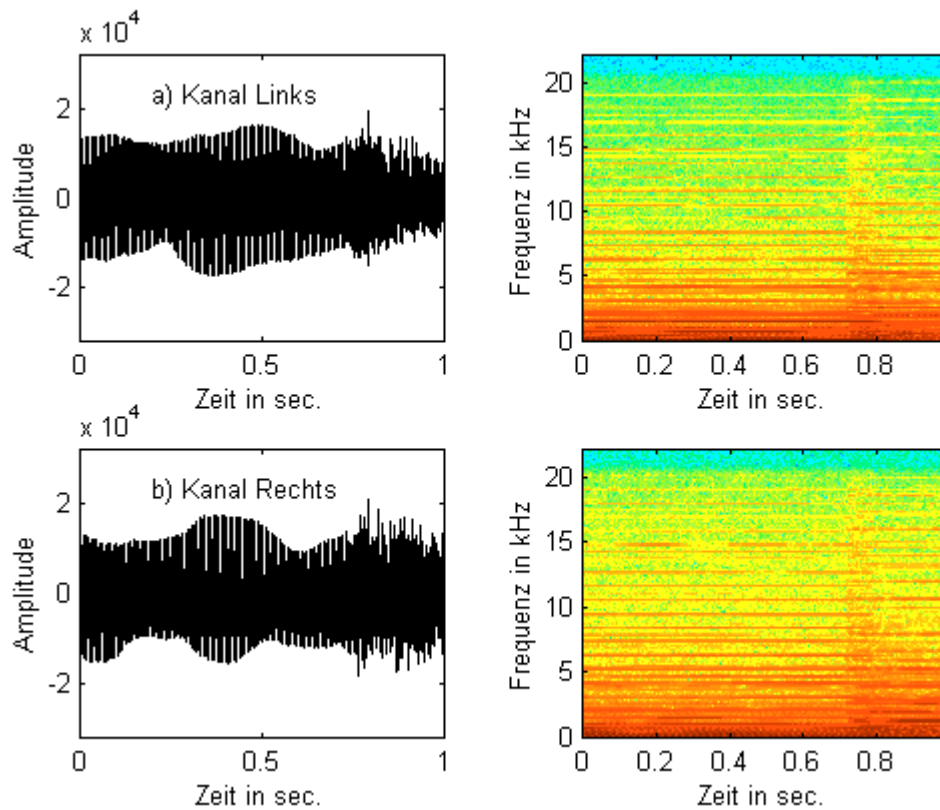


Bild 5.1: Ein Ausschnitt über 1 Sekunde Dauer aus dem Stereo-Audiosignal eines Einzelinstruments (Orgel) und die dazugehörigen Spektrogramme. Signalverlauf und Spektrogramm des linken Kanals (oben); Signalverlauf und Spektrogramm des rechten Kanals (unten).

Beide Kanäle des Stereosignals (Kanal links: oben links und Kanal rechts unten links) zeigen zuerst unterschiedliche Verläufe (Bild 5.2). Sie sind aber beide näherungsweise periodisch und haben in diesem Abschnitt die gleiche Periodendauer (350 Abtastwerte).

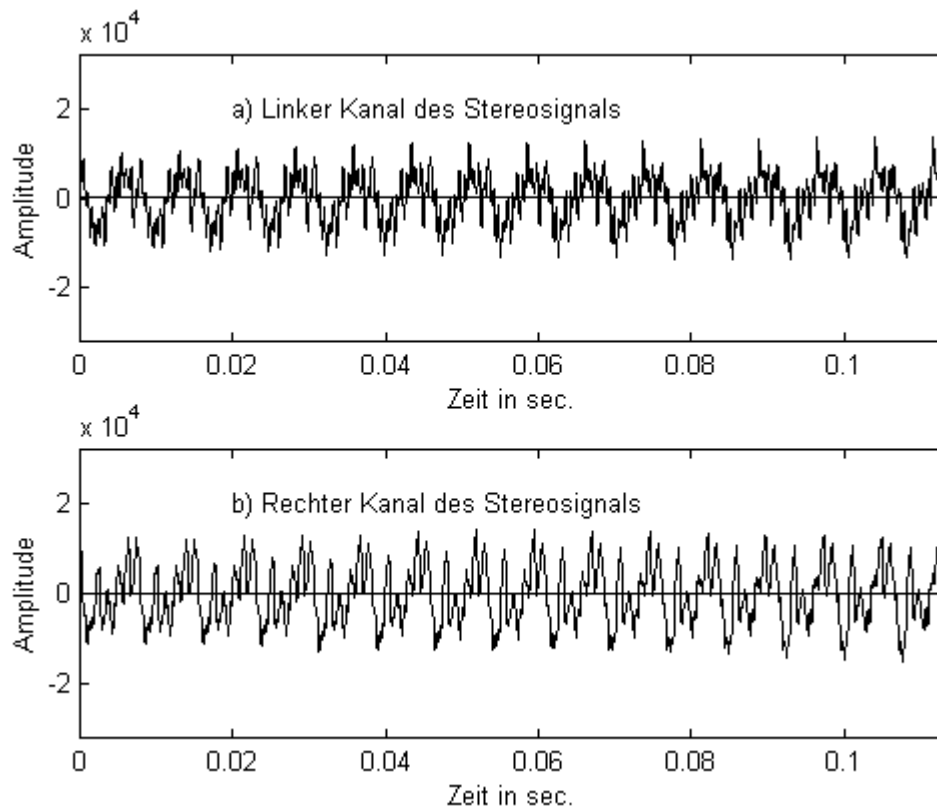


Bild 5.2: Ein Ausschnitt über 5000 Abtastwerte (113 ms) aus dem Stereo-Audiosignal eines Einzelinstrumentes (Ausschnitt aus dem Signal aus Bild 5.1). Beide Kanäle des Stereosignals sind näherungsweise periodisch. a) Linker Kanal des Stereosignals; a) Rechter Kanal des Stereosignals.

Eine Untersuchung der Korrelation des linken Kanals mit sich selbst (Autokorrelationsfunktion in Bild 5.3) stellt einen typischen Korrelationsverlauf periodischer Signale dar. Der Abstand zwischen benachbarten Spitzen im Korrelationsverlauf ist fast konstant und entspricht der Periodendauer eines periodischen Signals. In diesem Abschnitt entspricht die Periodendauer des Signalverlaufs im linken Kanal 350 Abtastwerte. Die Untersuchung der Kreuzkorrelation mit dem Nachbarkanal (Kreuzkorrelationsfunktion in Bild 5.3) zeigt, dass die Kreuzkorrelationsfunktion zwar ihr Maximum nicht genau bei einer Verschiebung $\tau = 0$ hat, weist aber insgesamt einen ähnlichen Verlauf wie die Autokorrelationsfunktion. Dies ist hier nicht auf die Ähnlichkeit des Signalverlaufs zwischen beiden Kanälen zurückzuführen, sondern eher auf die ebenfalls vorhandenen Periodizität im Nachbarkanal, die hier die selbe Periode (und damit auch die Frequenz) wie das Signal aus dem linken Kanal enthält. Man kann es hier mit einer spektralen Korrelation beschreiben.

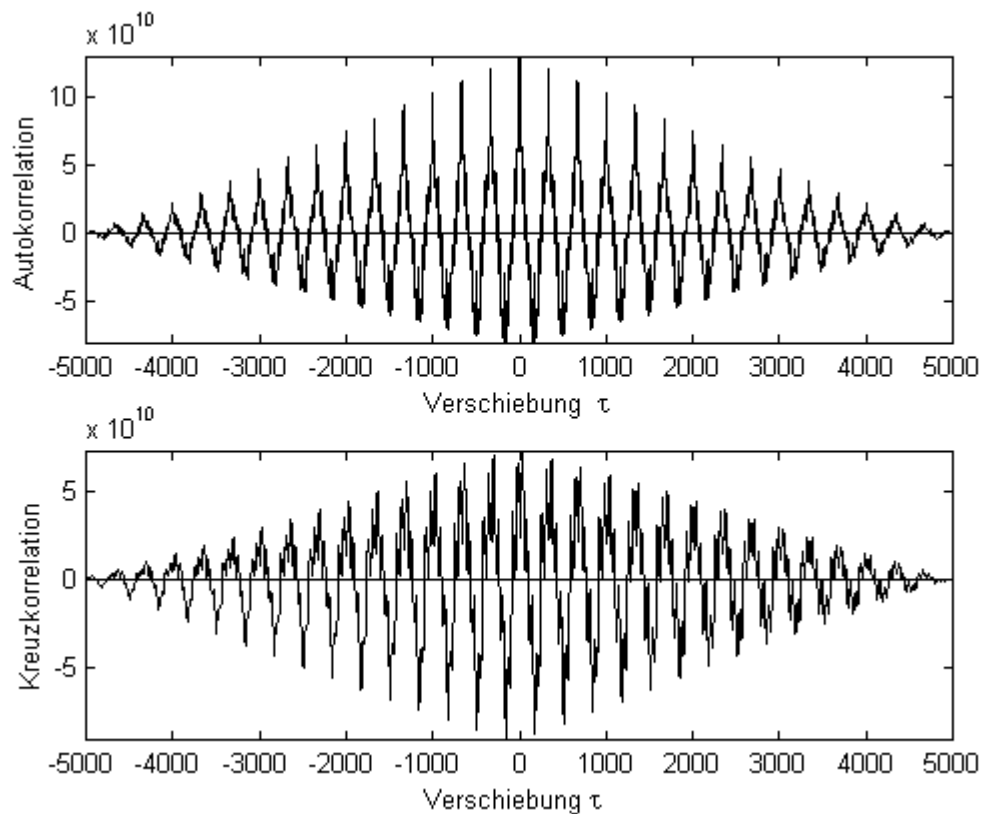


Bild 5.3: Autokorrelations- und Kreuzkorrelationsfunktion des Stereosignals aus Bild 5.2. Die Autokorrelationsfunktion beschreibt die Korrelation des linken Kanals mit sich selbst und die Kreuzkorrelationsfunktion die Korrelation zwischen beiden Kanälen. Die Korrelationsuntersuchung erfolgte für einen Ausschnitt der Länge 5000 Abtastwerte.

Titel 2: Klassische Musik (Beethoven)

Bild 5.4 zeigt einen Ausschnitt von 1 Sekunde Dauer (441 00 Abtastwerte) aus dem Stereosignal von Titel 2 und die dazugehörigen Spektrogramme. Es handelt sich hierbei um einen Orchester-Musik-Titel. Im Gegensatz zum Spektrogramm von Titel 1 sind hier keine eindeutigen Frequenzlinien zu erkennen und somit können keine Periodizitäten im Signalverlauf vorhanden sein.

Bild 2.4 aus Kapitel 2 zeigt einen vergrößerten Ausschnitt aus diesem Signal. Es handelt sich hierbei um ein Segment der Länge 5000 Abtastwerte (ungefähr 113 ms) aus der Stereodatei aus Bild 5.4. Beide Kanäle des Stereosignals haben unterschiedliche Signalverläufe.

Bei klassischer und Orchester-Musik unterscheiden sich die Kanäle stark von einander. Die Instrumente werden sehr unterschiedlich auf die Kanäle verteilt. Eine Abhängigkeit zwischen dem linken und dem rechten Kanal ist in solchen Fällen nicht unbedingt vorhanden bzw. zufällig.

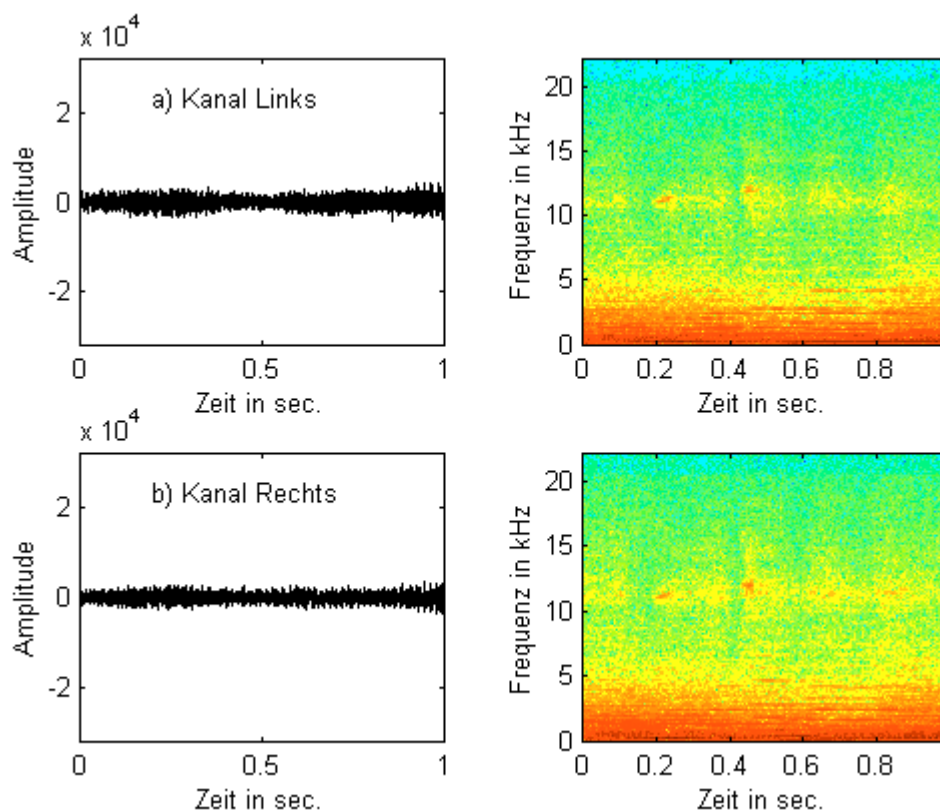


Bild 5.4: Ein Ausschnitt über 2 Sekunden Dauer aus dem Stereo-Audiosignal einer klassischen Musik (Beethoven) und die dazugehörigen Spektrogramme. Signalverlauf und Spektrogramm des linken Kanals (oben); Signalverlauf und Spektrogramm des rechten Kanals (unten).

Eine Korrelationsuntersuchung des linken Kanals mit sich selbst (Autokorrelationsfunktion aus Bild 2.5) und zwischen beiden Kanälen (Kreuzkorrelationsfunktion aus Bild 2.5), zeigt wie unterschiedlich beide Korrelationsverläufe sind. Der Kreuzkorrelationsverlauf zeigt keine eindeutigen Spitzen auf, die auf eine Übereinstimmung der Signalverläufe hindeuten könnten. Hinzu kommt, dass sich der maximale Kreuzkorrelationswert in diesem Signal-Ausschnitt von dem Maximum der Autokorrelation um den Faktor $4 \cdot 10^8$ unterscheidet. Dies alles bestätigt, dass beide Kanäle nicht miteinander korreliert sind.

Titel 3: Pop-Musik

Bild 5.5 zeigt einen Ausschnitt von 1 Sekunde Dauer aus dem Stereosignal von Titel 3 und die dazugehörigen Spektrogramme. Es handelt sich hierbei um einen Pop-Musik-Titel.

Bild 2.2 aus Kapitel 2 zeigt einen vergrößerten Ausschnitt der Länge 500 Abtastwerte aus diesem Signal. Beide Kanäle des Stereosignals (Kanal links oben, Kanal rechts unten) zeigen eine hohe Ähnlichkeit der Signalverläufe

Eine Korrelationsuntersuchung des linken Kanals mit sich selbst (Autokorrelationsfunktion aus Bild 2.3) und zwischen beiden Kanälen (Kreuzkorrelationsfunktion aus Bild 2.3), zeigt die hohe Ähnlichkeit der Kanäle, die durch der eindeutigen Spitze der Kreuzkorrelationsfunktion bei der Verschiebung $\tau = 0$ beschrieben wird. Beide Korrelationsverläufe sind sehr ähnlich.

Bei Pop- und Rock-Musik sind die Kanäle meist sehr ähnlich. Transienten und Attacken sind oft auf beiden Kanälen vorhanden. Stark synthetische Pop-Musik enthält Effekte, die schnelle Frequenz- und Amplitudenänderungen erzeugen.

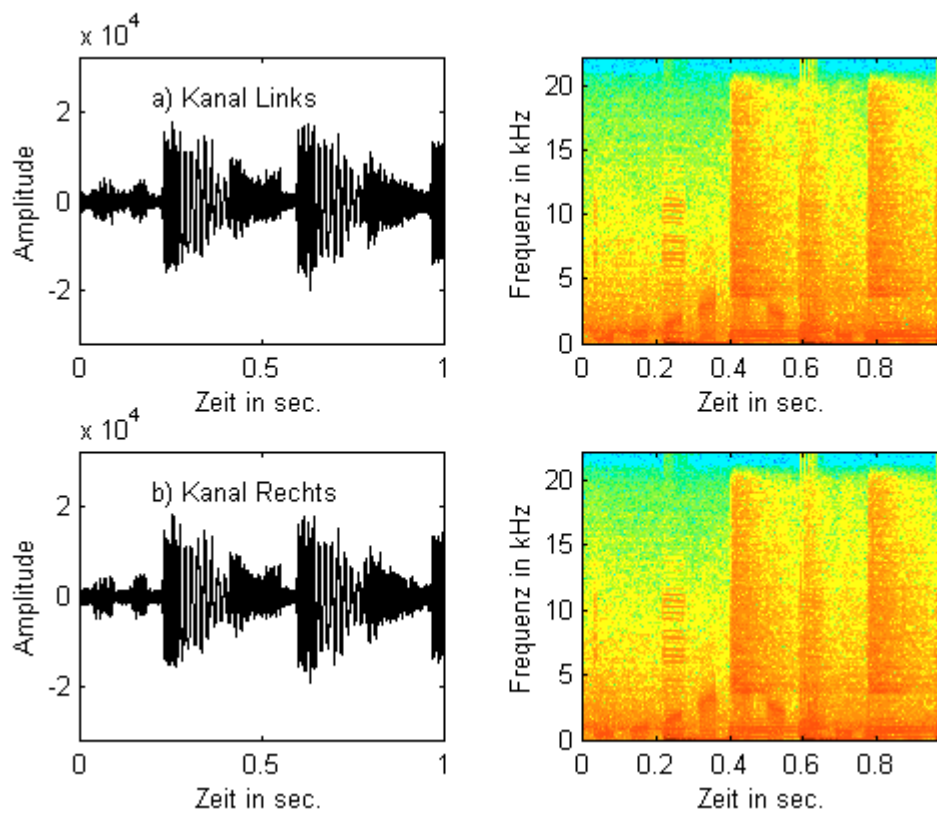


Bild 5.5: Ein Ausschnitt über 1 Sekunde Dauer aus dem Stereo-Audiosignal eines Pop-Musik-Titels und die dazugehörigen Spektrogramme. Signalverlauf und Spektrogramm des linken Kanals (oben); Signalverlauf und Spektrogramm des rechten Kanals (unten).

5.5 Simulation auf Signalebene

In diesem Abschnitt werden einige der in Kapitel 4.3 untersuchten Substitutionsverfahren bei unterschiedlichen Paketverlustraten P_v und Paketlängen verglichen. Diese Verfahren wer-

den direkt auf das PCM-Signal angewendet (Signalebene). Es werden folgende Substitutionsmöglichkeiten untersucht:

- Nullsubstitution (Null) (Werte sind der Tabelle C.1 Im Anhang C zu entnehmen): Stellt das Signal mit Paketverlusten dar und wird daher als Referenz verwendet.
- Paketwiederholung (Wied) (Tabelle C.2).
- Substitution mit Links/Rechts-Ersatz (LR) (Tabelle C.3): Hier wird vorausgesetzt, dass Paketverluste nur im linken Kanal des Stereosignals auftreten. Der rechte Kanal bleibt fehlerfrei, somit ist ein Links-/Rechts-Ersatz der Pakete möglich.
- Verfahren der maximalen Kreuzkorrelation (MK)(Tabelle C.4). Hier wird auch vorausgesetzt, dass Paketverluste nur im linken Kanal auftreten. Das Substitutionssignal wird dem Kanal entnommen, indem das Paket fehlt, wenn eine Monokorrelation festzustellen ist. Andernfalls wird das Substitutionssignal dem Nachbarkanal entnommen (Stereokorrelation).

Um die Robustheit der untersuchten Verfahren gegenüber Paketverlusten zu testen, wird im folgenden eine hohe Paketverlustwahrscheinlichkeit (10 %) untersucht.

Titel 1: Einzelnes Instrument

Die objektive (Bild 5.6) und subjektive (Bild 5.7) Bewertung der Audioqualität der oben erwähnten Substitutionsverfahren zeigen bei Titel 1 eine beträchtliche Verbesserung der Audioqualität gegenüber der Nullsubstitution. Bei Titel 1 handelt es sich um Musik mit einem einzelnen Instrument. Der Signalverlauf dieses Audiosignals weist Periodizitäten auf, die sich über längere Zeitabschnitte erstrecken.

Mit der Paketwiederholung wird in diesem Titel eine nur geringfügige Verbesserung der Audioqualität gegenüber der Nullsubstitution erreicht. Die Nullsubstitution hinterlässt, je nach Paketlänge, durch ihre schwachen Knackse eine holprig klingende Störung. Die Paketwiederholung erzeugt ein zischendes bis hallartiges Hintergrundgeräusch. Die mit der Paketwiederholung erreichbare Audioqualität bleibt insgesamt schlecht. Das wird damit begründet, dass das in der Lücke eingefügte Substitutionssignal nicht in Phase mit dem übrigen periodischen Signal steht, und somit Phasenverzerrungen hervorruft, die für die schlechte Audioqualität verantwortlich sind.

Beide Kanäle des Stereosignals aus Titel 1 sind nicht oder nur schwach miteinander korreliert. Ein Links-/Rechts-Ersatz der Pakete bei Paketverlust in einem Kanal führt nicht zu einer Verbesserung der Audioqualität, in diesem Fall insgesamt sehr störend bleibt.

Das MK-Verfahren ist ein Verfahren, das für die Verdeckung der Lücke die Ähnlichkeit des Signals mit seiner Vergangenheit und mit dem Signal des Nachbarkanals ausnutzt. Aufgrund der in Titel 1 vorhandenen Periodizitäten ist es mit dem MK-Verfahren möglich, diese

Periodizität als Monokorrelation in dem Signal zu interpretieren und ein Substitutionssignal zu erzeugen, das phasenrichtig in das übrige Signal eingefügt werden kann. Die mit diesem Verfahren erreichte Audioqualität ist auch bei hohen Verlustwahrscheinlichkeiten (10 %) kaum störend (Bild 5.8).

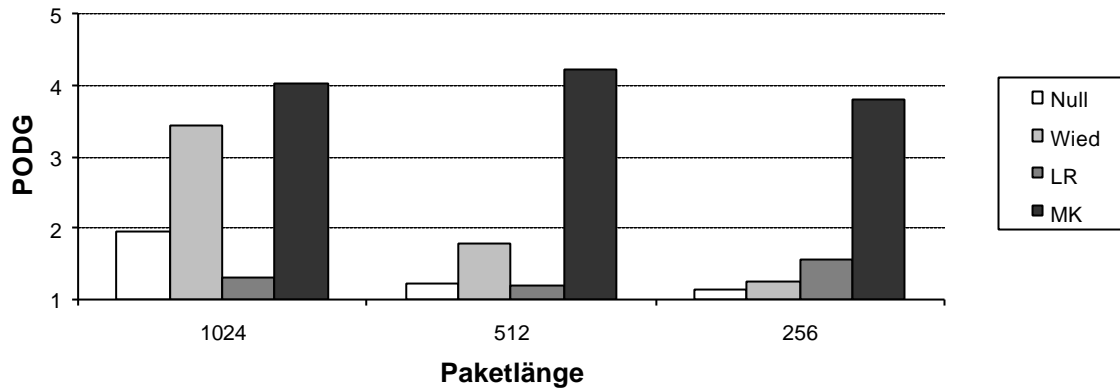


Bild 5.6: Objektive Bewertung der Substitution für Titel 1 in Abhängigkeit von der Paketlänge; Paketverlustwahrscheinlichkeit = 10%; Paketverluste nur im linken Kanal.

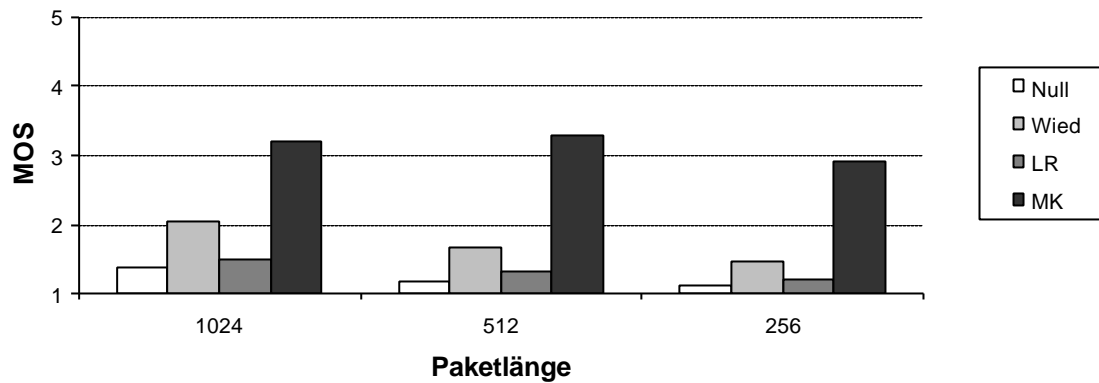


Bild 5.7: Subjektive Bewertung der Substitution für Titel 1 in Abhängigkeit von der Paketlänge; Paketverlustwahrscheinlichkeit = 10%. Paketverluste nur im linken Kanal.

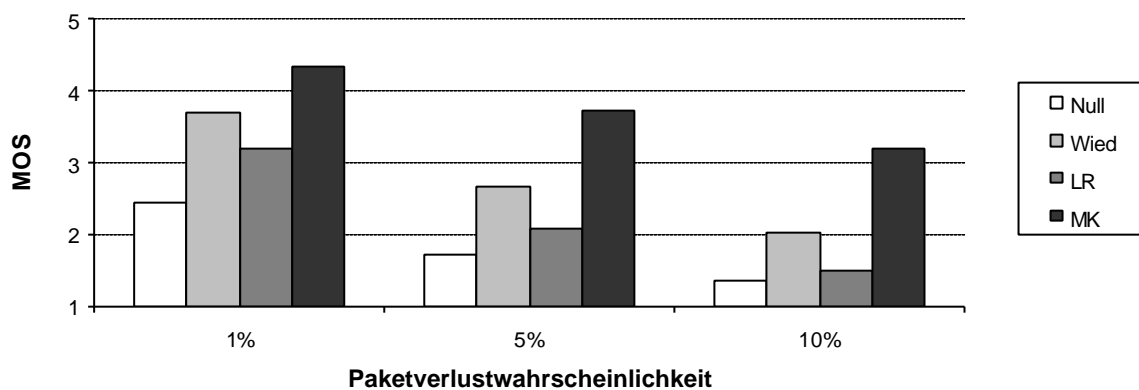


Bild 5.8: Subjektive Bewertung der Substitution für Titel 1 in Abhängigkeit von der Paketverlustwahrscheinlichkeit; Paketlänge = 1024 AW. Paketverluste nur im linken Kanal.

Titel 2: Klassische Musik

Bei Titel 2 handelt es sich um klassische Musik. Hier weist der Signalverlauf keine besonderen Eigenschaften auf. Die Signalverläufe beider Kanäle des Stereosignals sind auch nicht ähnlich, so dass man bei solchen Signalen nicht einschätzen kann, wie der Verlauf des fehlenden Audiosignalausschnittes aussehen könnte.

Die subjektive Bewertung aus Bild 5.10 zeigt, dass für die Paketverlustwahrscheinlichkeit von 10% und die Paketlänge 1024 akzeptable Ergebnisse erzielt werden. Dieses gilt insbesondere für die Paketwiederholung und die maximale Kreuzkorrelation. Die mit beiden Verfahren erreichbare subjektive Qualität ist für die Paketlänge 1024 nahezu gleich. Bei kleineren Paketen wird mit dem MK-Verfahren eine geringfügige Verbesserung der Audioqualität im

Vergleich zu der Paketwiederholung erzielt. Der MOS-Wert liegt hierbei unterhalb 3, was mit einer störenden Audioqualität interpretiert wird (Bild 5.10).

Bis zu einer Paketverlustwahrscheinlichkeit von 5% liegt der MOS-Wert bei einer Paketlänge von 1024 für beide Verfahren oberhalb 4 (Bild 5.11), die Verschleierung der Verluste ist zwar wahrnehmbar, aber nicht störend. Bis zu einem Verlust von 10 % der insgesamt gesendeten Pakete bleibt die Audioqualität kaum störend (MOS größer als 3).

Die geringfügige Verbesserung der Audioqualität, die mit der Paketwiederholung im Vergleich zu der maximalen Kreuzkorrelation (Bild 5.11) bei einer Paketverlustwahrscheinlichkeit von 1% erreicht wurde, liegt bei einem MOS-Wert-Unterschied von 0,2 und ist nicht signifikant. Dies wird mit der objektiven Bewertung aus Bild 5.9 bestätigt, denn da liegt der PODG-Wert-Unterschied bei 0,001. Insgesamt ist die Qualität beider Verfahren ähnlich.

Ein Links-/Rechts-Ersatz der Pakete liefert bei diesem Titel aufgrund der nicht vorhandenen Ähnlichkeiten zwischen den Signalverläufen beider Kanäle schlechtere Ergebnisse als die bei Paketwiederholung und bei Kreuzkorrelation.

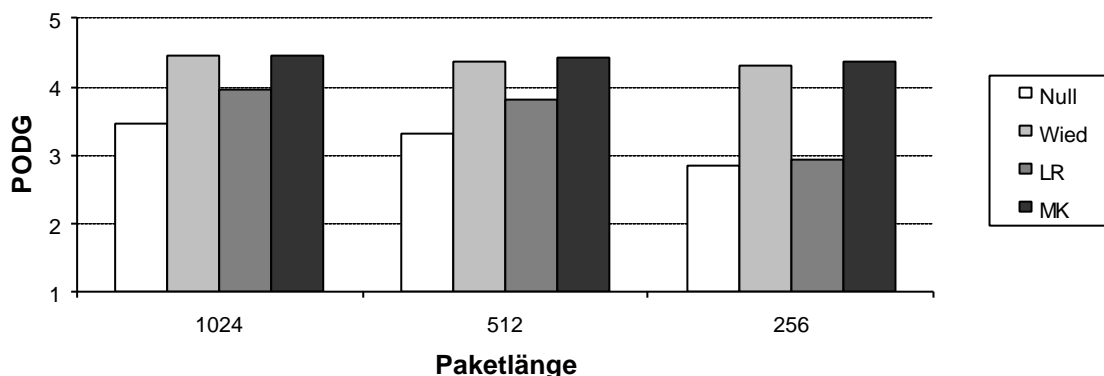


Bild 5.9: Objektive Bewertung der Substitution für Titel 2 in Abhängigkeit von der Paketlänge; Paketverlustwahrscheinlichkeit = 10%. Paketverluste nur im linken Kanal.

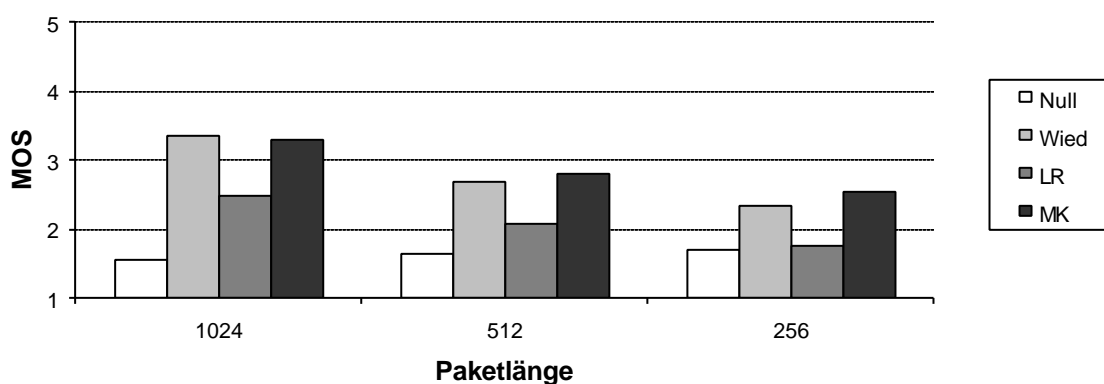


Bild 5.10: Subjektive Bewertung der Substitution für Titel 2 in Abhängigkeit von der Paketlänge; Paketverlustwahrscheinlichkeit = 10%. Paketverluste nur im linken Kanal.

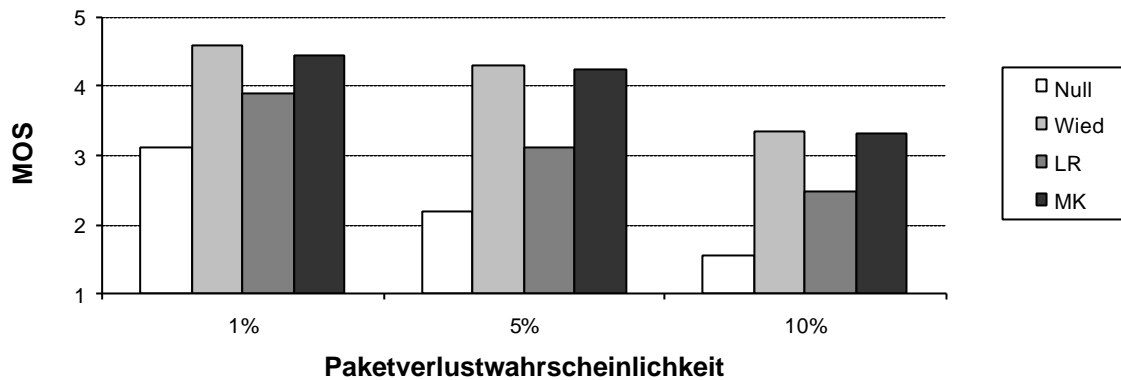


Bild 5.11: Subjektive Bewertung der Substitution für Titel 2 in Abhängigkeit von der Paketverlustwahrscheinlichkeit. Paketlänge = 1024 AW. Paketverluste nur im linken Kanal.

Titel 3: Pop Musik

Das Audiosignal von Titel 3 ist ein Stereosignal mit einer überdurchschnittlichen Korrelation zwischen beiden Kanälen. Die bei Popmusikaufnahmen üblicherweise verwendete Produktionstechnik erzeugt das Stereosignal aus einer Mischung einzelner monophoner Klangquellen. Die räumliche Staffelung („Stereo-Panorama“) der Klangkomponenten wird durch unterschiedliche Zumischungsgrade zum linken bzw. rechten Summenkanal erreicht [Her95]. In diesem Fall können, aufgrund der hohen Kohärenz zwischen den Kanälen, merkbare Verbesserungen durch Links-/Rechts-Ersatz erzielt werden. Auch bei hohen Paketverlustwahrscheinlichkeiten werden mit diesem Verfahren MOS-Werte, die zwischen 4 und 5 liegen, erreicht (s. Bild 5.13). Die Verdeckung der Lücke bleibt in den meisten Fällen unbemerkt.

Die objektive Bewertung aus Bild 5.12 zeigt, dass mit dem Verfahren der maximalen Kreuzkorrelation sogar bessere Ergebnisse erreicht werden als mit dem Links-/Rechts-Ersatz. Das liegt daran, dass beide Kanäle des Stereosignals meistens erst ab einer minimalen Verschiebung maximal ähnlich sind. Diese Verschiebung ist auf einen Phasenunterschied zwischen beiden Kanälen zurückzuführen. Sie kann mit dem MK-Verfahren erfasst werden und so die Lücke mit einem verschobene Signalausschnitt aus dem Nachbarkanal ersetzen.

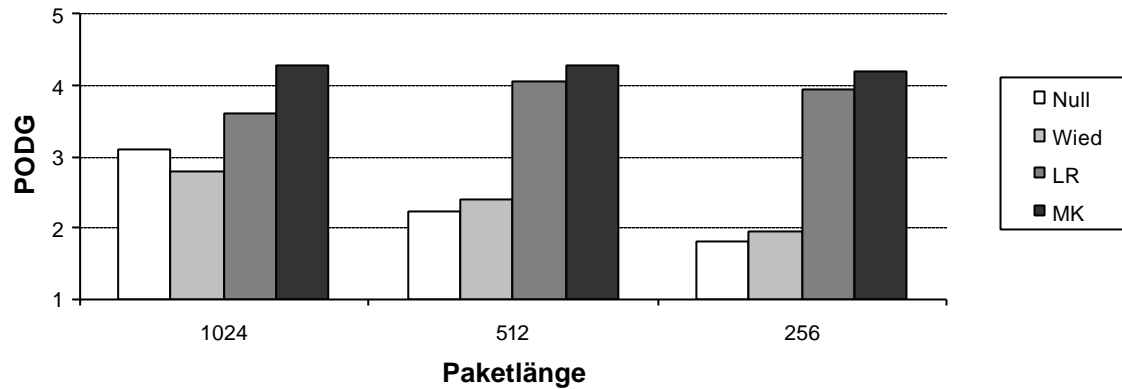


Bild 5.12: Objektive Bewertung der Substitution für Titel 3 in Abhängigkeit von der Paketlänge; Paketverlustwahrscheinlichkeit = 10%. Paketverluste nur im linken Kanal.

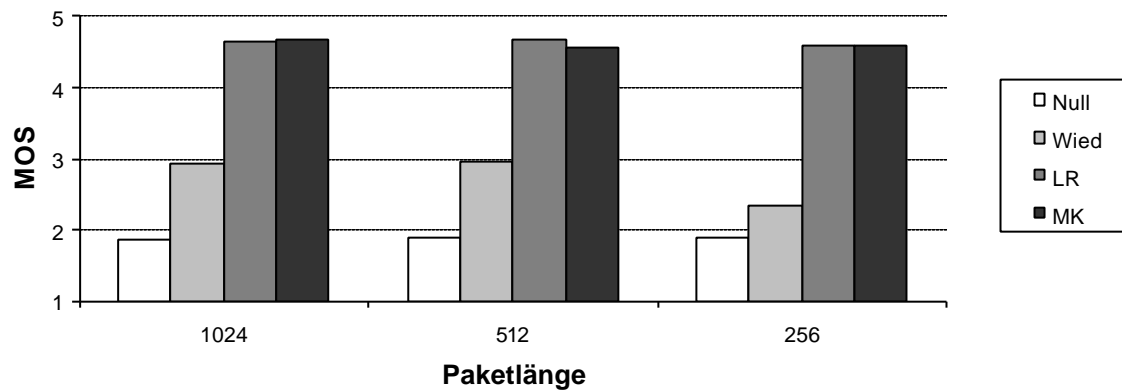


Bild 5.13: Subjektive Bewertung der Substitution für Titel 3 in Abhängigkeit von der Paketlänge; Paketverlustwahrscheinlichkeit = 10%. Paketverluste nur im linken Kanal.

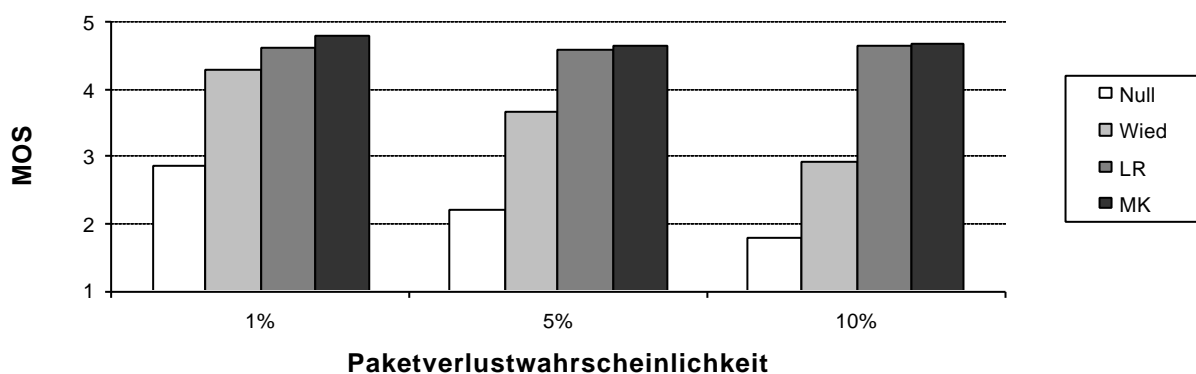


Bild 5.14: Subjektive Bewertung der Substitution für Titel 3 in Abhängigkeit von der Paketverlustwahrscheinlichkeit; Paketlänge= 1024 AW. Paketverluste nur im linken Kanal

Paketverluste in beiden Kanälen

Bei den oben dargestellten Ergebnissen wurde vorausgesetzt, dass einer der Kanäle (in diesem Fall der rechte Kanal) fehlerfrei ist. Bei einem Paketverlust in beiden Kanälen kann kein Links-/Rechts-Ersatz verwendet werden. Das Verfahren der maximalen Kreuzkorrelation erkennt in diesem Fall den Verlust in dem Nachbarkanal und versucht trotzdem die Korrelation des Signals unmittelbar vor der Lücke mit dem Puffer aus dem selben Kanal und dem Puffer aus dem Nachbarkanal zu ermitteln.

In Bild 5.15 ist eine objektive Bewertung der Verfahren LR und MK für die Paketsubstitution in Titel 1, Titel 2 und Titel 3 dargestellt. Hierbei stellen diese Werte die Audioqualität des Substitutionssignals in Fall von Paketverlusten in nur einem Kanal dar (hier ist nur der linke Kanal fehlerbehaftet). Es wurden die Paketlänge 1024 und die Paketverlustwahrscheinlichkeit 10 % verwendet. Als Referenz wurde hier wieder die Nullsubstitution genommen. Mit MK-Pv2K wird das MK-Verfahren mit Paketverlusten in beiden Kanälen bezeichnet.

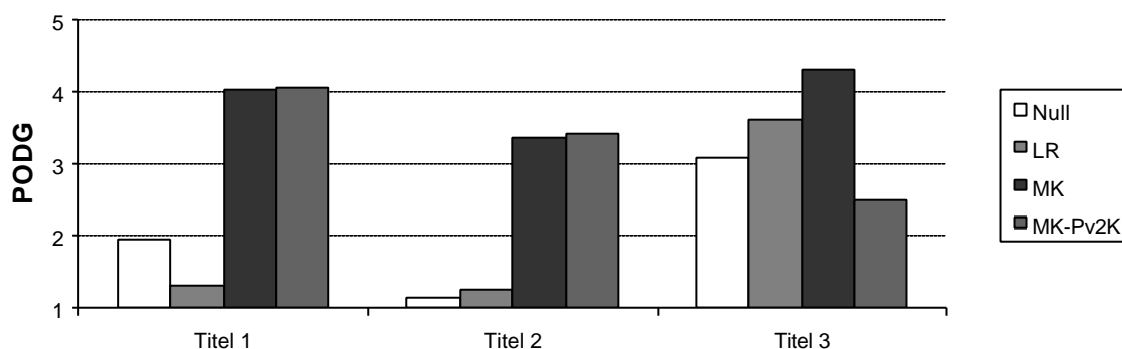


Bild 5.15: Objektive Bewertung der Substitution für eine Paketverlustwahrscheinlichkeit von 10% und eine Paketlänge von 1024 AW. Bei den Verfahren Null, LR und MK wurden Paketverluste nur im linken Kanal simuliert. Mit MK-Pv2K wird das MK-Verfahren mit Paketverlusten in beiden Kanälen bezeichnet.

Der Paketverlust im rechten Kanal hat in Titel 1 und Titel 2 keinen Einfluss auf die Rekonstruktion der Lücke im linken Kanal. Das MK-Verfahren hat eine Monokorrelation des Signals festgestellt und somit für die Lücke ein Substitutionssignal aus dem Puffer des Selben Kanals genommen.

In Titel 3 bewirkt der Paketverlust in dem rechten Kanal eine drastische Verschlechterung der Audioqualität. Beide Kanäle des Audiosignals sind sehr ähnlich. In diesem Fall entspricht der Verlust in beiden Kanälen auch dem Verlust des für die Lückenrekonstruktion besten Signals.

5.6 Simulation auf der Ebene der LPC-Innovation und mit Stereoprädiktion

Die in diesem Kapitel untersuchten Substitutionsverfahren sind:

- Links-/Rechts-Ersatz der Innovationen nach der Dekorrelation der Kanäle des Stereosignals mit der adaptiven linearen Prädiktion der Ordnung 20 (LPC-LR: Werte aus Tabelle C.5).
- Gewichtete Interpolation der Differenzsignale und anschließender Synthese der Innovation (LPC-GI aus Tabelle C.6).
- Rekonstruktion der Innovation nach Dekorrelation des Stereosignals mit der Stereoprädiktion (SLPC: Tabelle C.7). Hierbei wird die Ordnung des Monoprädiktors auf 10 und die des Kreuzprädiktors auf 5 gesetzt. In diesem Verfahren wird für die Rekonstruktion der Lücke die Nullsubstitution der Innovation in Titel 3 und die Wiederholung der Innovation aus dem letzten Paket in den Titeln 1 und 2 verwendet.

Titel 1: Einzelnes Instrument

Aufgrund der schwachen Korrelation der Signalverläufe beider Kanäle des Stereosignals aus Titel 1, ist es mit dem Links-/Rechts-Ersatz der Innovationen nicht möglich, eine gute Audioqualität nach der Rekonstruktion der Lücken in einem Kanal zu erwirken. Dieses gilt auch für das SLPC-Verfahren. Der MOS-Wert, der mit den beiden Verfahren erreicht wurde, liegt schon bei einer Verlustwahrscheinlichkeit von 5 % unterhalb 2,5 mit einer Audioqualität, die kaum besser als die Qualität des Signals mit Lücken ist.

Mit LPC-GI wird in diesem Titel eine Verbesserung erreicht im Vergleich zu der Nullsubstitution und zu LPC-LR. Basierend auf dem Prinzip des MK-Verfahrens untersucht das LPC-GI-Verfahren die Korrelation zwischen den LPC-Differenzsignalen und versucht eine Innovation zu finden, die mit dem fehlenden Differenzsignal der Lücke maximal ähnlich sein sollte. Titel 1 enthält periodische Signalabschnitte. Im dekorrelierten Differenzsignal sind diese Ähnlichkeiten meistens wieder zu finden. Das LPC-GI-Verfahren liefert in diesem Fall nur bis zu einer Verlustwahrscheinlichkeit von 5% gute Ergebnisse (Bild 5.18). Dieses gilt nur für die Paketlänge 1024. Die Tatsache, dass die Ähnlichkeiten im dekorrelierten Differenzsignal manchmal schwer zu finden sind und somit die Möglichkeit besteht, dass das Verfahren falsche Entscheidungen treffen und ungeeignete Innovationen zur Verdeckung

der Lücke nehmen kann, erklärt, dass die Ergebnisse des LPC-GI-Verfahrens schlechter als die des MK-Verfahrens sind.

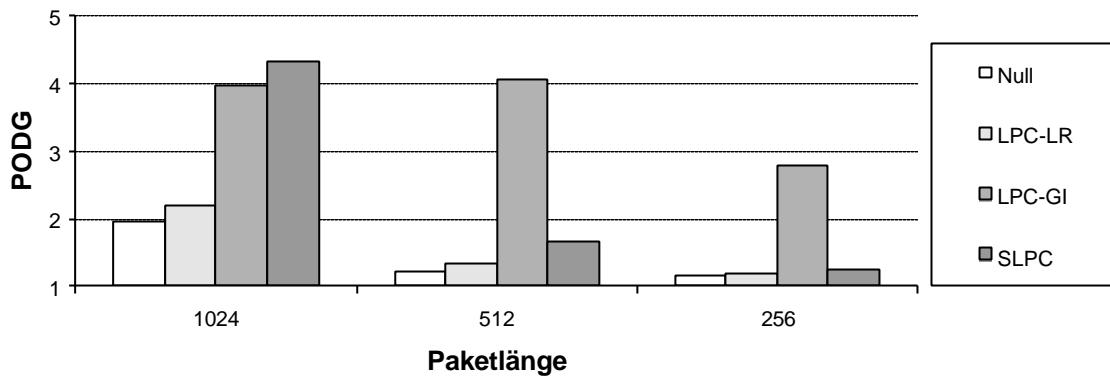


Bild 5.16: Objektive Bewertung der Substitution für Titel 1 in Abhängigkeit von der Paketlänge; Paketverlustwahrscheinlichkeit = 10%. Paketverluste nur im linken Kanal.

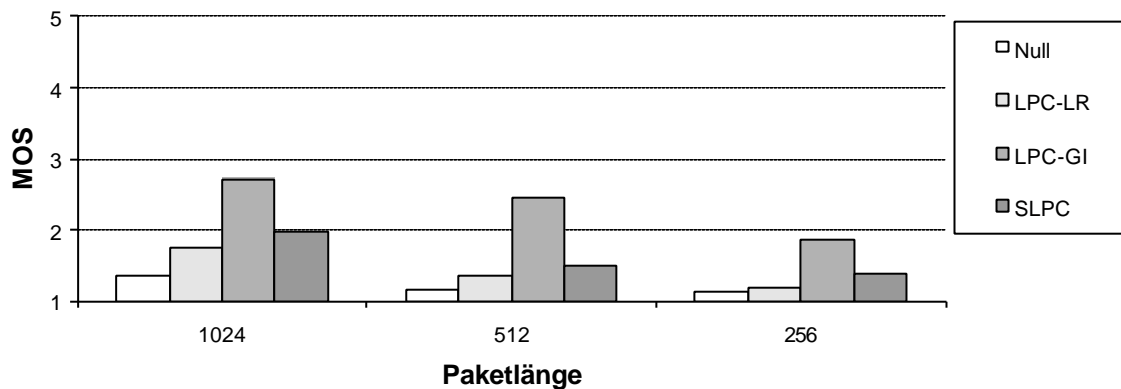


Bild 5.17: Subjektive Bewertung der Substitution für Titel 1 in Abhängigkeit von der Paketlänge; Paketverlustwahrscheinlichkeit = 10%. Paketverluste nur im linken Kanal.

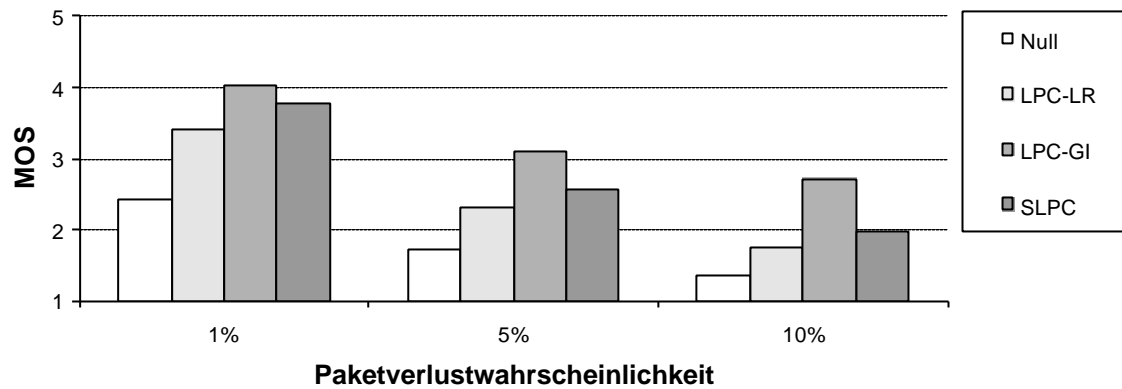


Bild 5.18: Subjektive Bewertung der Substitution für Titel 1 in Abhängigkeit von der Paketverlustwahrscheinlichkeit; Paketlänge = 1024 AW. Paketverluste nur im linken Kanal

Titel 2: Klassische Musik

Mit der Stereoprädiktion wird bis zu einer Paketverlustwahrscheinlichkeit von 10 % eine gute Audioqualität erreicht (Bild 5.21). Dies gilt allerdings nur für die Paketlänge 1024. Insgesamt bleibt die Audioqualität der mit LPC-LR, LPC-GI und SLPC bis zu der Paketverlustwahrscheinlichkeit von 10 % für die Paketlänge 1024 akzeptable. Bei kleineren Paketlängen wird für die Paketverlustwahrscheinlichkeit von 10% die Audioqualität störender.

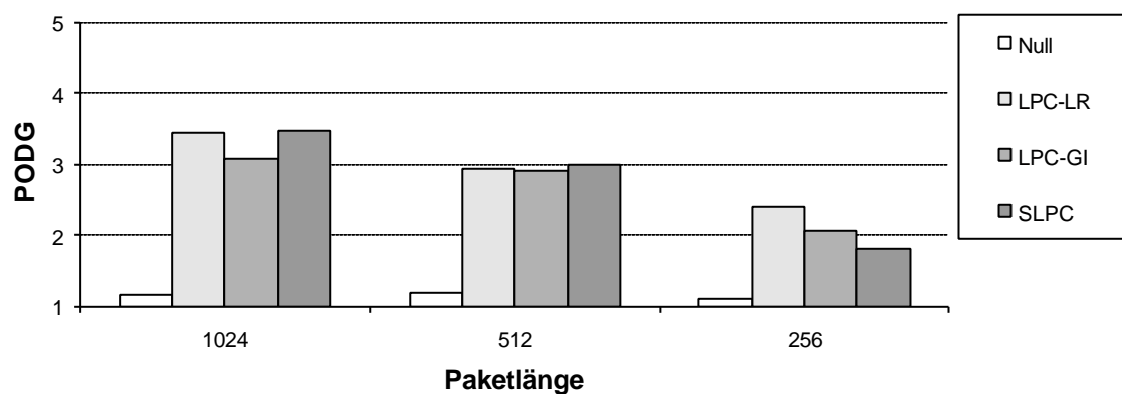


Bild 5.19: Objektive Bewertung der Substitution für Titel 2 in Abhängigkeit von der Paketlänge; Paketverlustwahrscheinlichkeit = 10%. Paketverluste nur im linken Kanal.

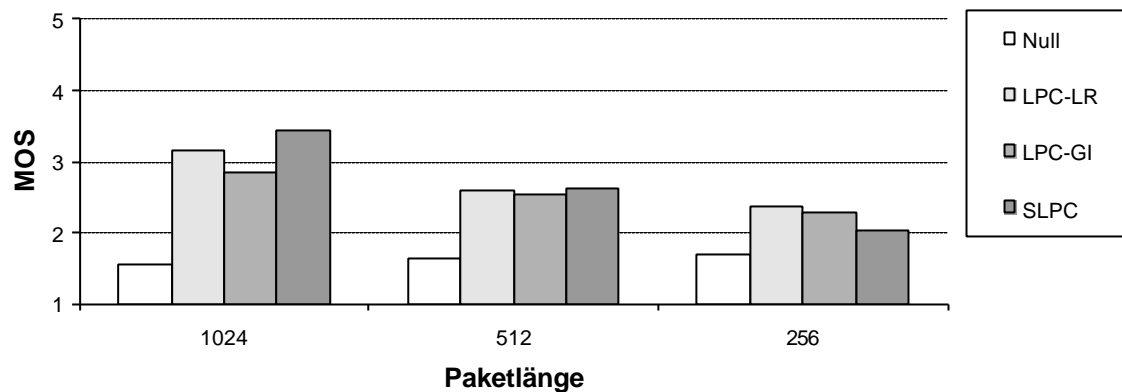


Bild 5.20: Subjektive Bewertung der Substitution für Titel 2 in Abhängigkeit von der Paketlänge; Paketverlustwahrscheinlichkeit = 10%. Paketverluste nur im linken Kanal.

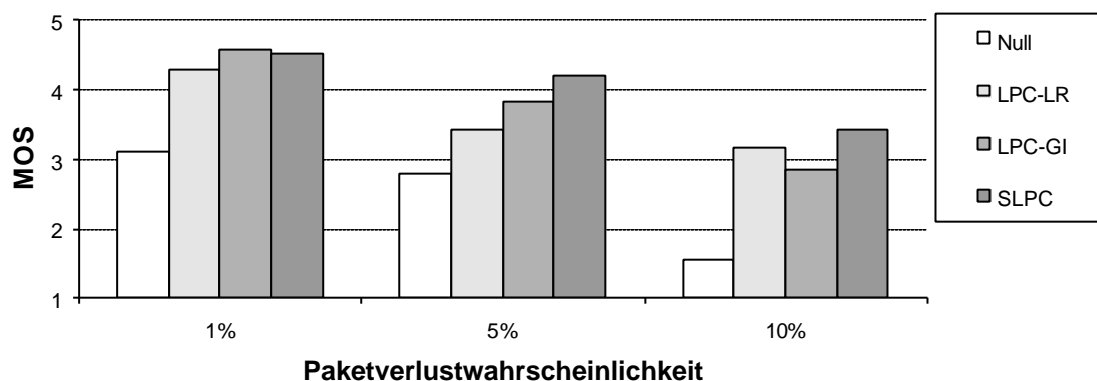


Bild 5.21: Subjektive Bewertung der Substitution für Titel 2 in Abhängigkeit von der Paketverlustwahrscheinlichkeit; Paketlänge = 1024 AW. Paketverluste nur im linken Kanal.

Titel 3: Pop Musik

Die subjektive Bewertung (Bild 5.23) der mit den oben erwähnten Rekonstruktionsverfahren erreichten Audioqualität zeigt die Effizienz der Verfahren LPC-LR und LPC-GI bei der Paketsubstitution in einem Kanal bei vorhandenen Paketen in dem Nachbarkanal des Stereosignals. Dies betrifft Stereosignale mit großer Ähnlichkeit zwischen beiden Kanälen. Sollte das Paket aus dem Nachbarkanal nicht zur Verfügung stehen, so kann mit LPC-LR keine Substitution durchgeführt werden. In diesem Fall kann auch das LPC-GI-Verfahren, analog zu dem MK-Verfahren (s. Bild 5.15) keine gute Audioqualität liefern.

Bei der Substitution von Paketverlusten mit der Stereoprädiktion wurde nur der Fall betrachtet, dass ein Paketverlust in einem der beiden Kanäle des Stereosignals auftritt. Bei Titel

3 sollte es mit dem SLPC-Verfahren möglich sein, die Ähnlichkeit zwischen beiden Kanälen zu erfassen und somit eine gute Verdeckung der Lücke zu gewährleisten. Die subjektive Bewertung dieses Verfahrens hat ergeben, dass das rekonstruierte Audiosignal noch störende Knacksgeräusche enthält. Es werden in diesem Fall mit dem LPC-GI- und dem LPC-LR-Verfahren bessere Ergebnisse erreicht als mit SLPC.

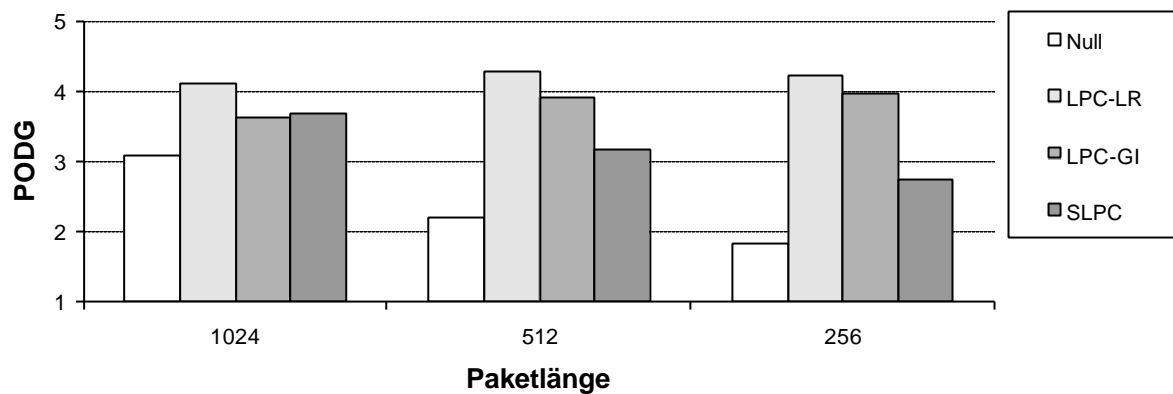


Bild 5.22: Objektive Bewertung der Substitution für Titel 3 in Abhängigkeit von der Paketlänge; Paketverlustwahrscheinlichkeit = 10%. Paketverluste nur im linken Kanal.

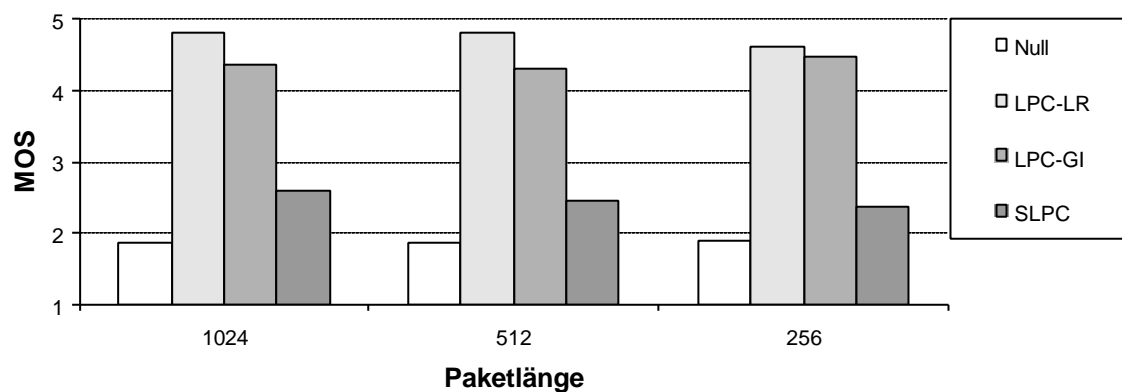


Bild 5.23: Subjektive Bewertung der Substitution für Titel 3 in Abhängigkeit von der Paketlänge; Paketverlustwahrscheinlichkeit = 10%. Paketverluste nur im linken Kanal.

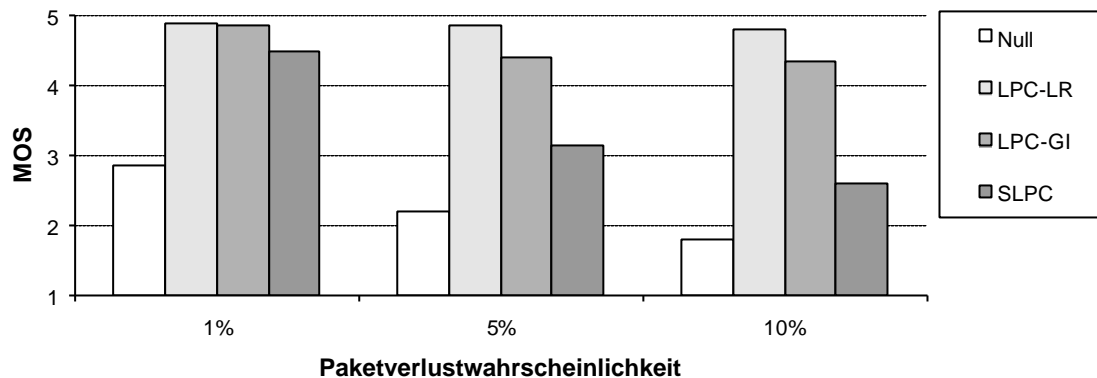


Bild 5.24: Subjektive Bewertung der Substitution für Titel 3 in Abhängigkeit von der Paketverlustwahrscheinlichkeit; Paketlänge = 1024 AW. Paketverluste nur im linken Kanal.

5.7 Simulation auf Frequenzbereichsebene

Bei der Substitution von Paketverlusten im Frequenzbereich wurden zwei Verfahren untersucht:

- **Spektrales Morphing (SM):** Das Verfahren wird auf Stereosignale angewendet. Die Substitution erfolgt anhand der Interpolation der DCT-Spektren beiderseits der Lücke mit Einbeziehung des DCT-Spektrums aus dem Nachbarkanal im Fall einer Stereokorrelation (s. Kapitel 4.6.2).
- **Frequenz-Tracking (FTR):** Mit diesem Verfahren erfolgt eine Rekonstruktion der Lücke in einem Kanal des Stereosignals unabhängig von dem Nachbarkanal. Somit ist das FTR-Verfahren kanalunabhängig. Mit diesem Verfahren werden die Frequenzen beiderseits der Lücke miteinander verbunden (Kapitel 4.6.3). Im Gegensatz zum beispielsweise MK- oder LPC-GI-Verfahren ist das Ziel der Rekonstruktion mit dem FTR-Verfahren nicht mehr den ursprünglichen Signalverlauf der Lücke zu erhalten, sondern mehr ein Substitutionssignal zu erzeugen, das unhörbar in das übrige Audiosignal eingefügt werden kann.

Mit dem Spektrales-Morphing-Verfahren ist in Titel 1 nur bei kleinen Paketverlustwahrscheinlichkeiten (bis 5%) gute Audioqualität erreichbar (Bilder 5.26, 5.28 und 5.30). Das Verfahren zieht zur Substitution die Spektren einiger zuvor gepufferte Segmente des Audiosignals vor. Bei einer höheren Paketverlustwahrscheinlichkeit ist es wahrscheinlicher, dass die analysierten Segmente zum Teil rekonstruierte Substitutionssignale enthalten (s. Bild 4.13). Dieses führt insbesondere bei einer qualitativ schlechten Paketsubstitution zu einer „Fehler-

fortpflanzung“. Die Qualität der Paketsubstitution kann die Audioqualität des Substitutionssignals eines in der nahen Zukunft auftretenden Paketverlusten beeinflussen.

Mit diesem Verfahren werden zwar die durch den Verlust hervorgerufenen Störungen im Audiosignal (Knackgeräusche) nicht mehr wahrnehmbar, allerdings klingt das Substitutionssignal in den meisten Fällen anders als das Originalsignal. Besonders stark fällt das bei Attacken von beispielsweise Schlaginstrumenten (Bild 4.38) auf.

Mit dem Frequenz-Tracking-Verfahren wird bei allen drei Titeln eine gute Audioqualität, die auch bei hohen Paketverlustwahrscheinlichkeiten einen MOS-Wert von ungefähr 4 hat, erreicht.

In Bild 5.27 liegt der MOS-Wert des Audiosignals nach der Rekonstruktion von insgesamt 10 % aller Pakete mit der Paketlänge 256 AW bei 2,56 und entspricht damit einer etwas störenden Audioqualität.

Das Audiosignal von Titel 2 enthält keine periodische Anteile, somit unterscheiden sich die Frequenzen beiderseits der Lücke. Die Frequenzkomponenten des Substitutionssignals entsprechen nicht unbedingt der des Originalsignals. Bei einer höheren Paketverlustwahrscheinlichkeit und kleineren Paketlängen werden für die Rekonstruktion der Lücke Teile von rekonstruierten Substitutionssignalen in der Vergangenheit der Lücke analysiert. Diese können hier in beschränktem Maß die Audioqualität beeinträchtigen. Dies ist beispielsweise bei Titel 1 nicht der Fall, der MOS-Wert liegt hier für die Paketlänge 256 bei 3,61. Aufgrund der in diesem Signal vorhandenen Segmenten mit näherungsweise periodischem Verlauf, werden in diese Segmente mit dem FTR-Verfahren oft die richtigen Frequenzen der Lücke geschätzt.

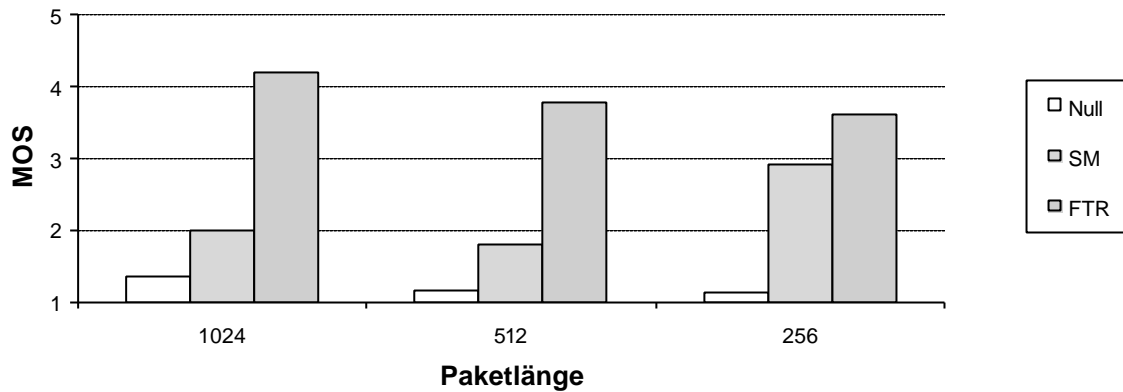
Titel 1: Einzelnes Instrument

Bild 5.25: Subjektive Bewertung der Substitution für Titel 1 in Abhängigkeit von der Paketlänge; Paketverlustwahrscheinlichkeit = 10%.

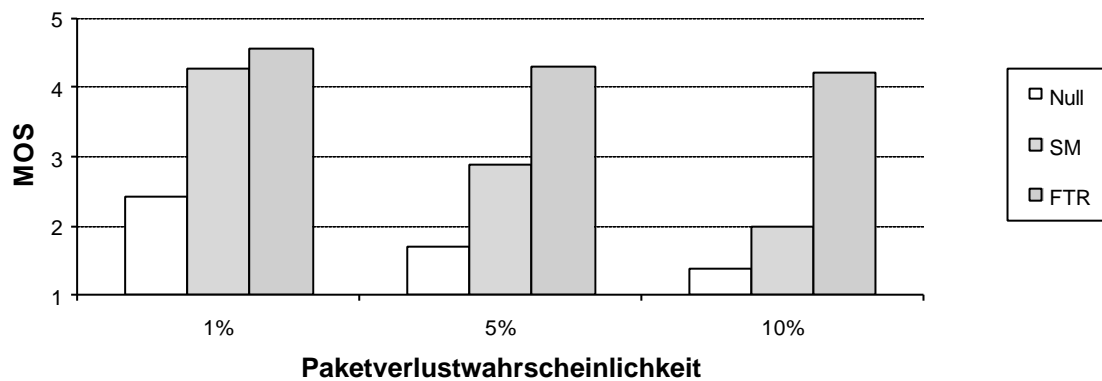


Bild 5.26: Subjektive Bewertung der Substitution für Titel 1 in Abhängigkeit von der Paketverlustwahrscheinlichkeit; Paketlänge = 1024 AW.

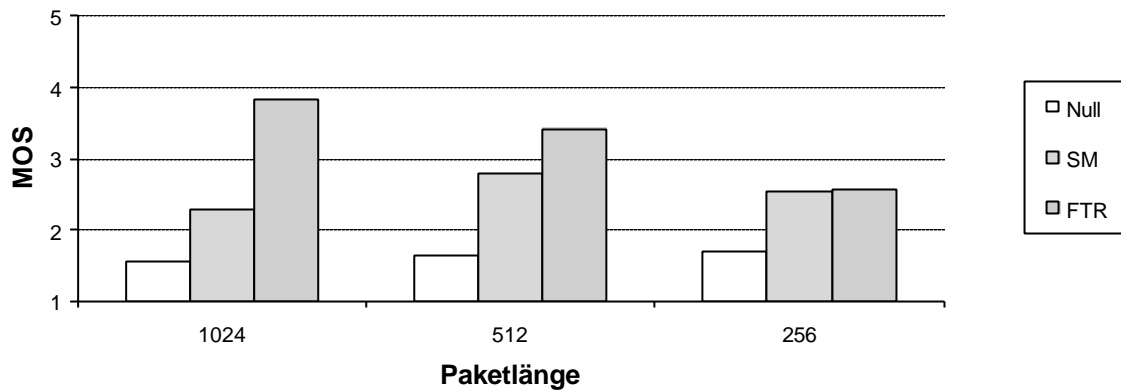
Titel 2: Klassische Musik

Bild 5.27: Subjektive Bewertung der Substitution für Titel 2 in Abhängigkeit von der Paketlänge; Paketverlustwahrscheinlichkeit = 10%.

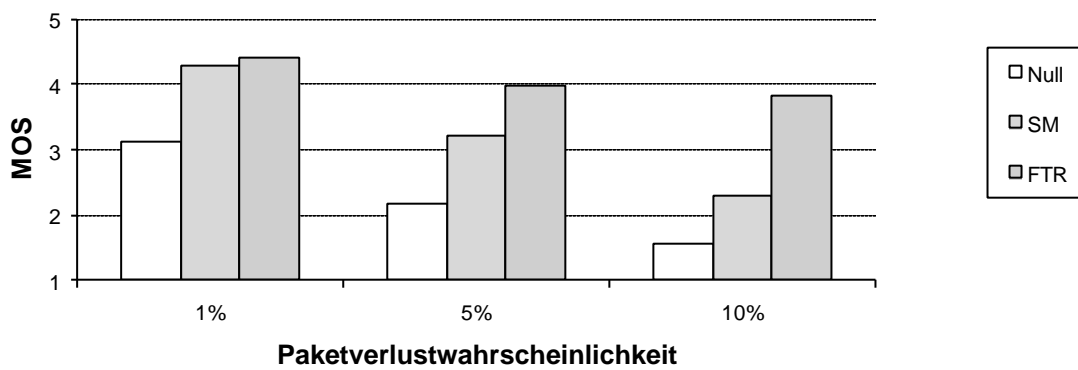


Bild 5.28: Subjektive Bewertung der Substitution für Titel 2 in Abhängigkeit von der Paketverlustwahrscheinlichkeit; Paketlänge = 1024 AW.

Titel 3: Pop Musik

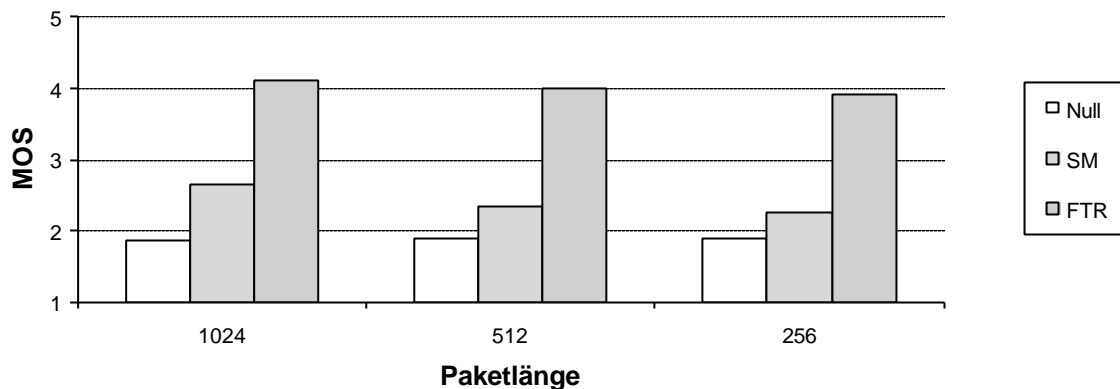


Bild 5.29: Subjektive Bewertung der Substitution für Titel 3 in Abhängigkeit von der Paketlänge; Paketverlustwahrscheinlichkeit = 10%.

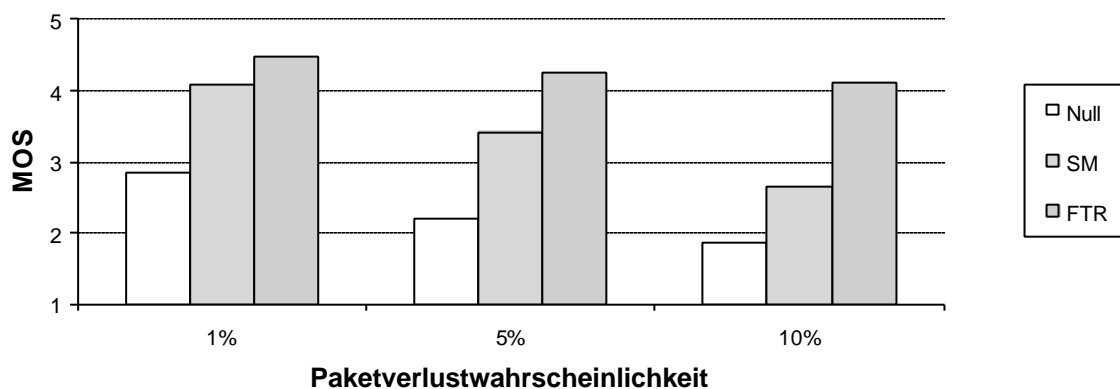


Bild 5.30: Subjektive Bewertung der Substitution für Titel 3 in Abhängigkeit von der Paketverlustwahrscheinlichkeit; Paketlänge = 1024 AW.

Vergleich der besten Verfahren

In Bild 5.31 werden die Verfahren miteinander verglichen, die bei dieser Untersuchung die besten Ergebnisse geliefert haben. Diese Verfahren sind:

- Verfahren der maximalen Kreuzkorrelation (MK): Wird auf das PCM-Signal angewendet,
- Gewichtete Interpolation der LPC-Innovation (LPC-GI) und
- Frequenz-Tracking (FTR): Wird direkt im Frequenzbereich angewendet.
- Die Nullsubstitution wird hier wieder als Referenz genommen.

Abgesehen von Titel 3, wird mit dem FTR-Verfahren immer die beste Audioqualität erreicht.

Die etwas bessere Audioqualität, die mit dem MK bzw. LPC-GI bei Titel 3 erreicht wurde, ist auf die für die Rekonstruktion der Lücke relevante Information aus dem Nachbarkanal zurückzuführen. Hier handelt es sich um eine starke Stereokorrelation des Stereosignals, wobei bei einem Paketverlust in einem Kanal das zeitsynchrone Paket aus dem Nachbarkanal vorhanden ist. Sollte dies nicht der Fall sein, so würden beide Verfahren eine Audioqualität liefern, die weit unter der des FTR-Verfahrens liegen: Mit dem MK-Verfahren beträgt der MOS-Wert bei einem Paketverlust in beide Kanäle 2,59, was im Vergleich zu dem kanalunabhängigen FTR-Verfahren mit seinem MOS-Wert von 4,12 auf einen markanten Qualitätsunterschied hin deutet.

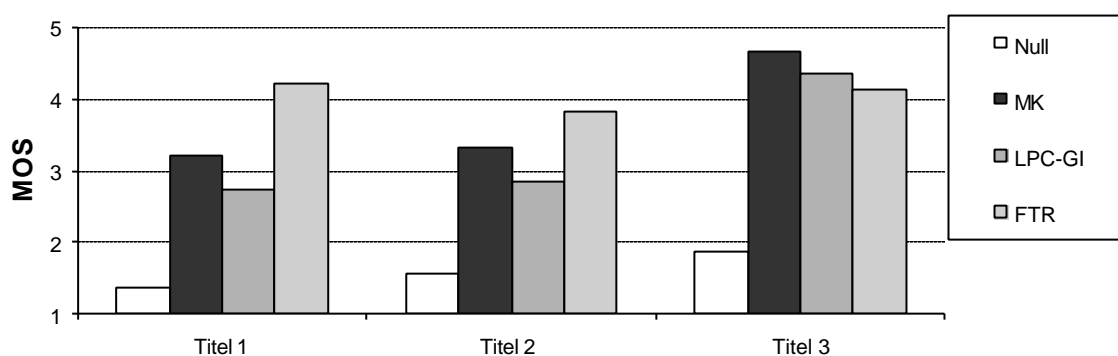


Bild 5.31: Subjektiver Vergleich der besten Verfahren. Paketverlustwahrscheinlichkeit = 10%; 1024 AW. Hier werden für alle drei Titel folgende Verfahren untersucht: Nullsubstitution auf PCM-Signal als Referenz, Verfahren der maximalen Kreuzkorrelation, Gewichtete Interpolation der LPC-Innovation und Frequenz-Tracking.

6 Zusammenfassung und Ausblick

Problemstellung

In paketvermittelnden Netzen - wie insbesondere dem Internet - werden zunehmend auch Audiodaten übertragen. Typische Anwendungen sind z.B. *Real-Time-Streaming* oder *Voice-over-IP*. Diese basieren i.A. auf dem unzuverlässigem Datagram-Service UDP, um Verzögerungen durch erneute Übertragung von verlorenen Paketen (wie bei TCP) zu vermeiden. Die UDP und RTP-Internet-Protokolle bieten keine Möglichkeit die Übertragung von Paketen für Echtzeitdienste wie Audio oder Sprache zu garantieren. Somit entstehen Paketverluste, welche die Dienstqualität beeinträchtigen und einen Qualitätsverlust der Audiodaten verursachen. Es existieren verschiedene Verfahren, um diesen Qualitätsverlust möglichst klein zu halten.

Im Verlauf dieser Arbeit wurde eine Reihe von Verfahren zur Paketsubstitution in Audiosignalen untersucht. Für die Verdeckung einer Lücke, die auf einen Paketverlust während der Übertragung zurückzuführen ist, ist eine weitgehende Parametrisierung, Analyse und darauf beruhende Inter- bzw. Extrapolation des Audiosignals erforderlich. Die Signalbehandlung erfolgte sowohl im Zeit- als auch im Frequenzbereich, ausgehend von den aus der Audiocodierung und -verarbeitung bekannten Algorithmen. Beispiele hierfür sind die lineare Prädiktion, die Stereoprädiktion und die sinusoidale Modellierung von Audiosignalen.

Die Signalbehandlung im Zeitbereich basiert auf abstrakten, mathematisch motivierten Zielvorstellungen, welche das Audiosignal im Zeitbereich betrachten und eine bestmögliche Annäherung an den Signalverlauf des Idealsignals anstreben [Her95]. Für die Rekonstruktion von Paketverlusten in Audiosignalen ist die damit erreichbare subjektive Audioqualität notwendigerweise begrenzt, aufgrund der Diskrepanz zwischen dem zugrundeliegenden mathematischen Konzept und dem völlig andersartigen Funktionsschema des menschlichen Hörsinns [Her95].

Eine entscheidende Verbesserung in der wahrnehmungsbezogenen Anpassung gelang durch den Übergang in der verwendeten Signalrepräsentierung von der Zeitbereichs- in eine mathematisch äquivalente Spektralbereichsdarstellung. Durch den Übergang auf dieses Konzept konnte gegenüber der Signalbehandlung im Zeitbereich eine wesentliche Steigerung der subjektiven Audioqualität erreicht werden.

Zielsetzung

Nach einer fehlerbehafteten paketorientierten Übertragung von Audiodaten kann das Ziel eines empfängerseitigen Substitutionsverfahrens rein intuitiv sein, den fehlenden Ausschnitt des Audiosignals so zu rekonstruieren, dass er dem Originalsignal möglichst nahe kommt. Eine perfekte Rekonstruktion ist nur bei einfachen z.B. näherungsweise periodischen Signalen möglich. Diese Signale besitzen bestimmte Eigenschaften, die sich über einen größeren Zeitraum nicht oder nur unwesentlich ändern.

Ziel der Paketsubstitution soll naturgemäß nicht sein das Originalsignal nachzubilden, sondern ein Substitutionssignal für die Lücke zu finden, das möglichst nicht als Ersatzsignal wahrgenommen werden soll. Der Paketverlust soll also verschleiert und nicht mehr wahrgenommen werden.

Ergebnisse

Die Bewertung der im Rahmen dieser Arbeit untersuchten Substitutionsverfahren erfolgte anhand objektiver und subjektiver Vergleiche. Die objektive Bewertung wurde mit den im ITU-R Standard empfohlenen Messverfahren PEAQ durchgeführt. Mit dieser Bewertung wird nicht der genaue subjektive Eindruck wiedergespiegelt. Die in diesem Fall besser geeignete subjektive Bewertung erfolgte mit Hilfe von informellen Hörtests. Für die Auswertung wurden drei unterschiedliche Audiodateien verwendet, die unterschiedliche Eigenschaften aufweisen (Kapitel 5.4).

Bei der Paketsubstitution auf **Signalebene** ergibt das MK-Verfahren, im Vergleich beispielsweise zur Paketwiederholung oder zum Links/Rechts-Ersatz der Pakete, die beste subjektive Audioqualität. Dieses Ergebnis war bei allen drei untersuchten Audiodateien zu bemerken.

Die Paketsubstitution mittels **linearer Prädiktion** erfordert eine LPC-Analyse der zuvor gepufferten Audiosegmente und gegebenenfalls des Paketes unmittelbar nach der Lücke (LPC-GI). Das Substitutionssignal ergibt sich aus der Synthese einer geschätzten Innovation mit dem LPC-Synthesefilter. Dazu werden die Filterkoeffizienten des letzten Paketes oder andere Koeffizienten, die zu der Innovation am besten passen (LPC-GI), verwendet.

Mit LPC-LR sind nur bei Stereosignalen, die eine hohe Korrelation zwischen beiden Kanälen aufweisen gute Ergebnisse erreichbar. Das gilt insbesondere für Pop-Musik und natürlich nur, wenn im Fall eines Paketverlustes in einem Kanal das zeitgleiche Paket aus dem Nachbarkanal zur Verfügung steht.

Mit LPC-GI wird bei der Paketlänge 1024 bis zu einer Paketverlustwahrscheinlichkeit von 5% eine gute Audioqualität erreicht. Bei kleineren Paketen fällt die Audioqualität schnell ab. Das liegt an der insgesamt höheren Anzahl der zu rekonstruierenden Lücken im Audiosignal. Das Verfahren ist empfindlich gegen dicht aufeinanderfolgende Paketverluste, denn

die Rekonstruktion einer Lücke hängt stark von der Audioqualität zuvor rekonstruierter Lücken ab.

Die **Stereoprädiktion** nutzt zur Rekonstruktion der Lücke die Korrelation zwischen beiden Kanälen des Stereosignals aus. Im Vergleich zu der Monoprädiktion ist die Stereoprädiktion rechenintensiver und empfindlicher gegenüber Fehlern. Zusätzlich zu der notwendigen Optimierung bei der Berechnung der Filterkoeffizienten eines adaptiven Stereoprädiktors ist es erforderlich das Synthesefilter zu stabilisieren. Dazu werden anhand der Filterkoeffizienten alle Polstellen des Synthesefilters berechnet und im Instabilitätsfall modifiziert, um ein stabiles Verhalten dieses Filters zu erzwingen.

Insgesamt bleibt die Audioqualität der mit der Stereoprädiktion rekonstruierten Substitutionssignale niedriger als die von anderen Substitutionsverfahren. Es sollte hier darauf hingewiesen werden, dass die mit SLPC rekonstruierten Paketen zunächst auf triviale Rekonstruktion der Innovation (Wiederholung oder Nullsubstitution) basieren. Dieses sollte allerdings bei der Lücken-Substitution in Titel 3 beispielsweise ausreichen, da beide Kanäle des Audiosignals von Titel 3 eine hohe Stereokorrelation aufweisen (Bild 4.35). Mit den Filterkoeffizienten und den Abtastwerten aus dem Nachbarkanal sollte es möglich sein eine perfekte Rekonstruktion der Lücke zu gewährleisten. Dies ist nicht immer gelungen. Die Audioqualität einiger Substitutionssignale, insbesondere von kleinen Paketen, bleibt noch störend.

Eine wesentliche Steigerung der subjektiven Audioqualität von substituierten Lücken in Audiosignalen gelang durch den Übergang auf die spektrale Interpolation der Signalkomponenten der Lücke und anschließender Rücktransformation in den Zeitbereich. Hierbei wurde das **Frequenz-Tracking**-Verfahren vorgeschlagen, dessen Funktionsweise dem Prinzip der sinusoidalen Modellierung von Audiosignalen entnommen wurde.

Abgesehen von Titel 3, wurde mit dem FTR-Verfahren die beste Audioqualität erreicht. Die etwas bessere Audioqualität, die mit dem MK bzw. LPC-GI bei Titel 3 erreicht wurde, ist auf die für die Rekonstruktion der Lücke relevante Information aus dem Nachbarkanal zurückzuführen. Hier handelt es sich um eine starke Stereokorrelation des Stereosignals, wobei bei einem Paketverlust in einem Kanal das zeitgleiche Paket aus dem Nachbarkanal vorhanden ist. Sollte dies nicht der Fall sein, so würden beide Verfahren eine Audioqualität liefern, die weit unter der des FTR-Verfahrens liegen: Mit dem MK-Verfahren beträgt der MOS-Wert bei einem Paketverlust in beiden Kanäle 2,59, was im Vergleich zu dem kanalunabhängigen FTR-Verfahren mit seinem MOS-Wert von 4,12 auf einen markanten Qualitätsunterschied hindeutet.

Man kann also allgemein schließen, dass das FTR-Verfahren sehr gut geeignet ist für die Substitution von Paketverlusten in komplexen Audiosignalen. Das Verfahren ist kanalunabhängig und kann auch dann eingesetzt werden, wenn ein Paketverlust beide Kanäle des Stereosignals betrifft.

Bei einer hohen Stereokorrelation des Audiosignals und einem Paketverlust in nur einem Kanal, sollten besser Verfahren verwendet werden, die diese Information auch ausnutzen

können und im Vergleich zum FTR-Verfahren weniger Rechenaufwand benötigen. Gute Ergebnisse werden in diesem Fall mit dem MK-Verfahren, der einen MOS-Wert von 4,65 ergibt bei einer hohen Verlustwahrscheinlichkeit (10 %) und großen Paketlängen (1024 AW). Das LPC-GI-Verfahren liefert unter dieser Voraussetzung einen MOS-Wert von 4,35.

Ausblick

Verschiedene Substitutionsverfahren liefern bei unterschiedlichen Signalcharakteristika und Paketlängen unterschiedlich gute Ergebnisse. Um ein Verfahren besser einschätzen zu können, sollte es nur dann Anwendung finden, wenn die betreffende Signalcharakteristik gerade zutrifft. Eine solche an die Signalcharakteristik angepasste Analyse und Rekonstruktion könnte beispielsweise durch das Aussuchen von Segmenten im Audiosignal erfolgen, welche die entsprechende oder gewünschte Signalcharakteristik besitzen.

Die vorgelegten Ergebnisse könnendazu dienen, Verfahren zu entwickeln, bei denen bei verschiedenen Signalcharakteristika und Fehlerumgebungen zwischen verschiedenen Rekonstruktionsverfahren geschaltet wird.

Bei kleineren Paketlängen und hohen Verlustwahrscheinlichkeiten sollte überlegt werden, kurz aufeinanderfolgende Verluste als einen zu betrachten. Das heißt im Grunde das Verwerfen von Paketen, die fehlerfrei empfangen wurden.

7 Literaturverzeichnis

- [Ant93]** A. Antoniou
Digital Filter Analysis, Design and Applications.
Mc Graw-Hill 1993.
- [Aßm91]** U. Aßmus, H. Hessenmüller
Die Übertragung hochwertiger Bild- und Tonsignale in einem auf ATM basierenden Breitband-ISDN.
Aus Poch, H. Seiler, B., "Taschenbuch der Telekompraxis", Schiele & Schön, Berlin 91, S. 77-95.
- [Bar80]** T. P. Barnwell III
A Comparison of Parametrically Different Objective Speech Quality Measures Using Correlation Analysis with Subjective Quality Results.
IEEE Proceedings ICASSP 1980, Denver, pp. 710-3, April 1980.
- [Bol97]** J. C. Bolot, A. V. García
The Case for FEC-Based Error Control for Packet Audio in the Internet.
ACM Multimedia System, 1997.
- [Bol95]** J. C. Bolot, H. Crepin, A. V. García
Analysis of Audio Packet Loss on the Internet.
Proceedings Of 5th Int. Workshop on Network and Operating System Support for Digital Audio and Video, pp.163-174, Durham, April 1995.
- [Boy98]** J. M. Boyce, R. D. Gaglianella
Paket Loss Effect on MPEG Video Sent Over the Public Internet.
ACM Multimedia 98- Electronic Proceeding.
- [Bra71]** P. T. Brady
Effects of Transmission Delay on Conversational Behaviour on Echo-Free Telephone Circuits.
Bell System Technical Journal, 50: 115-134. January 1971.
- [Cam93]** P. Cambridge
Audio Data Compression Techniques.
AES 94th Convention. March 16-19, 1993 Berlin. Preprint 3584 (K1-9)

- [Camp00]** P. Campolucci, F. Piazza
Intrinsic Stability-Control Method for Recursive Filters in Neuronal Networks.
IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing, Vol. 47, No. 8, p. 797-802, August 2000.
- [Che88]** T.M. Chen and D.G. Messerschmitt
Integrated Voice/Data Switching.
IEEE Communication Magazine, 26(6) :16-26, June 1988
- [Clü93]** K. Clüver, L. Schöder, G. Schuller, P. Noll
Digitale paketorientierte Audioübertragung in ATM-Netzen.
Abschlußbericht Berkomp-Projekt. TU Berlin. Januar 1993.
- [Clü98a]** K. Clüver
Rekonstruktion fehlender Signalblöcke bei blockorientierter Sprachübertragung.
Dissertation TU Berlin 1998.
- [Clü98b]** K. Clüver
Skript zur Vorlesung Digitale Nachrichtenverarbeitung.
TU Berlin 1998.
- [Clü96]** K. Clüver
Ein Verfahren zur Rekonstruktion fehlender Sprachsignalrahmen mit linearer Prädiktion und Teilbandanregung.
Frequenz 50 (1996).
- [Del93]** J. R. Deller, J. G. Proakis, J. H. L. Hansen
Discrete-Time Processing of Speech Signals.
Pentrice Hall. Englewood Cliffs 93.
- [Die96]** M. Dietz, H. Popp, K. Brandenburg, R. Friedrich
Audio Compression for Network Transmission.
J Audio Eng. Soc., Vol 44,1 / 2, pp.58-78. 1996.
- [Epp93]** B. Eppinger, E. Herter
Sprachverarbeitung.
Carl Hanser Verlag München Wien 1993. Reihe Informationstechnik, Nachrichtentechnik.
- [For98]** P. Frossard
MPEG-2 over Lossy Packet Networks QOS Analysis and Improvement.

Communication Systems Divison (SSC) Lausanne. 7 July 1998.

- [Fuc93]** H. Fuchs
Improving Joint Stereo Audio Coding by Adaptive Inter-Channel Prediction.
Proceedings of the 1993 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics. October 17-20 1993, Mohonk Mountain House, New Paltz, New York.
- [Gol77]** B. Gold
Digital Speech Networks
Proceeding of the IEEE, vol 65, N°. 12. Dezember 1997, S. 1636-1658.
- [Goo86]** D. J. Goodman, G. B. Lockhart, O. J. Wasem, W.-C. Wong
Waveform Substitution Techniques for Recovering Missing Speech Segments in Packet Voice Communications,
IEEE Trans. On Acoustics, Speech and Signal Processing, Vol. ASSP-34, no. 6. Dezember 1986, S. 1440-1448
- [Hän94]** R. Händel, M. N. Huber und S. Schröder
ATM Networks.
Addison-Wesley Publishing Company, 2. Auflage, 1994.
- [Har78]** F. J. Harris
On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform.
IEEE Proceedings. Vol. 66:1, pp.51-83. 1978
- [Har95]** V. Hardman, S. Angela, M. Handley, A. Watson
Reliable Audio for Use over the Internet.
INET'95 Hypermedia Proceedings 1995.
- [Her93]** J. Herre, E. Eberlein
Evaluation of Concealment Techniques for Compressed Digital Audio.
94th AES Convention. March 16-19, 1993 Berlin.
- [Her95]** J. Herre
Fehlerverschleierung bei spektral codierten Audiosignalen.
Dissertation, Universität Erlangen, 1995.
- [Hes83]** W. Hess
Pitch Determination of Speech Signals.
Springer-Verlag, Berlin 1983.

- [Hof98]** R. Hoffmann
Signalanalyse und -erkennung: Eine Einführung für Informationstechniker.
Springer Verlag Berlin Heidelberg 1998.
- [Hus95]** A. Husain
Reconstruction of Missing Packets for CELP-Based Speech Coders
Proceedings ICASSP V1, p. 245-248. Sep.1995
- [Hut94]** Michael Huter
Die Hörbarkeit von Phasenänderungen in synthetischen Klängen.
Diplomarbeit FH Düsseldorf 1994.
- [Jac92]** V. Jacobson
VAT manual pages.
Lawrence Berkeley Laboratory (LBL), Februar 1992.
- [Jay81]** N. Jayant, S. Christensen
Effects of Packet Losses in Waveform Coded Speech and Improvement Due to an Odd-Even Sample-Interpolation Procedure.
IEEE Trans. on Communications. Vol. Com-29, No. 2, p. 101-109, Febr. 1981.
- [Jay84]** N. S. Jayant, P. Noll
Digital Coding of Waveforms.
Prentice Hall. Englewood Cliffs 1984
- [Kah98]** M. Kahr; K. Brandenburg
Application of Digital Signal Processing to Audio and Acoustics.
Kluwer Academic Publishers 1998.
- [Kon94]** A. M. Kondoz
Digital Speech. (Coding for Low Bit Rate Communications Systems).
John Wiley & Sons. 1994.
- [Kos98]** T. J. Kostas, M. S. Borella I. Sidhu, G. M. Schuster, J. Grabiec, J. Maher
Real-Time Voice over Packet-Switched Networks.
IEEE Network 12(1): 18-27, January/February 1998.
- [Lau01]** P. Lauber, Ralph Sperschneider
Error Concealment for Compressed Digital Audio.
111th AES Convention. Preprint 5460. September 21-24, New York 2001.

- [Lei99]** F. v. Leitner
Audio-Kompression. 2 Dezember 1999.
<http://www.fefe.de/ct/audio.pdf>
- [Mah94]** R. C. Maher
A Method for Extrapolation of Missing Digital Audio Data.
J. Audio Eng. Soc., Vol. 42, N° 5, May 1994.
- [Mak75]** J. Makhoul
Linear Prediction : A Tutorial Review
IEEE Proceedings, Vol. 63, N°4, pp. 561-580, April 1975.
- [Mar76]** J.D. Markel, A. H. Gray
Linear Prediction of Speech.
Springer-Verlag, New York 1976
- [Mas96]** P. Masri
Computer Modelling of Sound for Transformation and Synthesis of Musical Signals.
Dissertation Uni Bristol 1996.
- [Mat87]** N. Matsuo, M. Yuito, Y. Tohunaga
Packet Interleaving for Reducing Speech Quality Degradation in Packet Voice Communications.
Proceedings GLOBECOM, p. 1787-1791, 1987
- [McA86]** R. J. McAulay; T. F. Quatieri
Speech Analysis / Synthesis Based on a Sinusoidal Representation.
IEEE Transactions on Acoustics, Speech and Signal Processing. 1986, Vol. 34:4, pp.744-754.
- [McD78]** B. J. McDermott, C. Scaglia, D. J. Goodman
Perceptual and Objective Evaluation of Speech Processed by Adaptive Differential PCM.
Bell System Technical Journal, vol. 57, no. 5, pp. 1597-619, May 1978.
- [McI92]** C. M. McIntyre, D. A. Dermott
A new Fine-Frequency Estimation Algorithm Based on Parabolic Regression.
IEEE Proceedings ICASSP 1992, p. 541-544.

- [Nol74]** P. Noll
Adaptive Quantization in Speech Coding Systems.
IEEE Int. Zürich Seminar, Oct 1974
- [Nol78]** P. Noll, J.M. Tribolet, B. J. McDermott and R. E. Crochiere
A Study of Complexity and Quality of Speech Waveform Coders.
IEEE Proceedings ICASSP 1978, Tulsa, pp. 586-590, Apr. 1978.
- [Nol86]** P. Noll, V. Lesemann, G. Wessels
Paketorientierte Sprachübertragung.
DFVLR_Mitteilung. Januar 1986
- [Ohm01]** J. R. Ohm
Multimediakommunikation I
Vorlesungsskript, Aachen 2001.
- [Pap87]** P.E. Papamichalis
Practical Approach To Speech Coding
Prentice Hall, Ch. 7, p.186-198. Englewood Cliffs 1987.
- [Per97]** C. Perkins, O. Hodson, V. Hardman
A Survey of Packet-Loss Recovery Techniques for IP-Based Audio-Visual Multi-cast Applications.
IEEE Networks, 11(6):24-36, November/December 1997.
- [Per98]** C. Perkins, O. Hodson, V. Hardman
A Survey of Packet-Loss Recovery Techniques for Streaming Audio.
IEEE Network. September/October 1998.
- [Pie92]** J. R. Pierce
The Science of Musical Sound (2nd ed.)
Publ. W. H. Freeman and co. New York 1992
- [Pre96]** W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery
Numerical Recipes in C. The Art of Scientific Computing. Second Edition.
Cambridge University Press 1996.
- [Pur98]** H. Purnhagen, B. Edler
Objektbasierter Analyse/Synthese Audio Coder für sehr niedrige Datenraten.
ITG-Fachtagung: Codierung für Quelle , Kanal und Übertragung;
Aachen, März 1998.

- [Qua88]** R. Quackenbush, T. P. Barnwell III, M.A. Clements
Objective Measures of Speech Quality.
Prentice Hall Englewood Cliffs, New Jersey 07632, 1988.
- [Rab77]** L. R. Rabiner
On the Use of Autocorrelation Analysis for Pitch Detection.
IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-25,
No. 1, February 1977.
- [Rab78]** L. R. Rabiner, R. W. Schafer, W. Ronald
Digital Processing of Speech Signals.
Prentice-Hall, Englewood Cliffs NJ (USA) 1978.
- [Rah97]** M. Rahkila, J. Huopaniemi
Real-Time Internet Audio – Problems and Solutions.
JAES, Preprint 4477, 102 AES Convention, Mar. 1997.
- [Rak71]** A. Rakowski
Pitch Discrimination at the Threshold of Hearing.
Proceedings of the 7-th International Congress on Acoustics., Vol. 3
Budapest 1971.
- [Ray96]** D. Raychaudhuri, D. Reiniger, R. Siracusa
Video Transport in ATM networks: A Systems View. Multimedia Systems.
Springer-Verlag 1996. 4: 305-315. NEC USA, C&C Research Laboratories
- [Reg95]** P. A. Regalia
Adaptive IIR-Filtering in Signal Processing and Control.
New York Marcel Dekker, 1995.
- [Rob95]** D. P. Robinson, L. D. Fielder
Migration of AC-2 and AC-3 Coding Methods from Professional to Consumer Applications.
AES 98th Convention, Preprint 3940., Paris, France 1995.
- [Rod90]** J. a. Rodriquez-Fonollosa, E. Margrau
A New Algorithm for Adaptive IIR-Filtering Based on the Log-Area Ratio Parameters.
Proceedings EUSIPCO 1990.

- [Rod97]** X. Rodet
Music Sound Signal Analysis/Synthesis: Sinusoidal+Residual and Elementary Waveform Models.
IRCAM, 1997 Paris.
- [Ros01]** J. D. Rosenberg
G.729 Error Recovery for Internet Telephony
Columbia University Science Technical Report. CUCS-016-01; December 19, 2001
- [San95]** H. Sanneck
Fehlerverschleierungsverfahren für Sprachübertragung mit Paketverlust.
Diplomarbeit TU Erlangen, Juni 1995.
- [San96]** H. Sanneck, A. Stenger, K. Ben Younes, B. Girod
A New Technique for Audio Packet Loss Concealment.
Global Internet 96.
- [San00]** H. Sanneck
Packet Loss Recovery and Control for Voice Transmission over the Internet.
Dissertation, TU Berlin 2000.
- [Sch92]** H. Schulzrinne
Voice Communication Across the Internet: A Network Voice Terminal.
University of Massachusetts. Technical Report June 1992.
- [Ser89]** X. Serra
A System for Sound Analysis/ Transformation / Synthesis Based on a Deterministic Plus Stochastic Decomposition.
Ph. D Thesis. Stanford University 1989.
- [Shy89]** J. J. Shynk
Adaptive IIR-Filtering using Parallel-Form Realisation
IEEE Transactions on Acoustics Speech and Signal Processing. Vol. 37, No. 4, April 1989, p. 519-533.
- [Sie93]** G. Siegmund
ATM - Die Technik des Breitband ISDN.
R.v. Decker, 1993.
- [Sla96]** M. Slaney, M. Covell, B. Lassiter

- Automatic Audio Morphing.*
International Conference on Acoustics, Speech and Signal Processing, IEEE Proceedings ICASSP Atlanta, May 1996.
- [Sie95]** Christian Sieckmeyer
Bewertung von adaptiven Ausspielalgorithmen für Paketvermittelte Audiodaten.
Studienarbeit TU Berlin 1995.
- [Spe00]** R. Sperschneider
Error Resilient Source Coding with Variable Length Codes and its Application to MPEG Advanced Audio Coding.
109th AES Convension, preprint 5271, 2000.
- [Sta91]** W. Stallings
Data and Computer Communications.
3 ed. New York, USA. 1991. Macmillian. 817p.
- [Ste96]** A. Stenger, K. Ben Younes, R. Reng, B. Girod
A New Error Concealment Technique for Audio Transmission with Packet Loss".
Proceedings EUSIPCO 96, Trieste, Italy, September 1996.
- [Ste97]** A. Stenger, R. Rabenstein, B. Girod
Fehlerverschleierung für paketierte Sprachübertragung durch Zeitdehnung und Phasen Anpassung.
9. Aachener Kolloquium "Signaltheorie", Aachen, Germany, March 1997
- [Suz98]** J. Suzuki, M. Taka
Missing Packet Recovery Techniques for Low-Bit-Rate Coded Speech.
IEEE Journal on Selected Areas in Communications, Vol. 7(5): 707-717, June 1989.
- [Thi98]** T. Thiede
Perceptual Audio Quality Assessment Using a Non-Linear Filter Bank.
Dissertation TU Berlin 1998.
- [Val89]** R. A. Valenzuela, C.N. Animalu
A New Voice Packet Reconstructtion Technique.
IEEE Proceedings ICASSP 89, p. 1334-1336.
- [Watk95]** C. R. Watkins, J. H. Chen
Improving 16 kbit/s G728 LD-CELP Speech Coder for Frame Erasure Channels.
IEEE Proceedings ICASSP 1995, p. 241-244.

- [Wats95]** S. D. Watson, B. M. G. Cheetham, W. T. K. Wong, A. V. Lewis
Low and Variable Bit-Rate Speech Coding for ATM Networks
Proceedings Eurospeech 95, p. 727-730.
- [Yon92]** M. Yong
Study of Voice Packet Reconstruction Methods Applied to CELP Speech Coding.
IEEE Proceedings ICASSP 92, p. II / 125-128 March 1992.
- [Zwi61]** E. Zwicker
Subdivision of the Audible Frequency Range into Critical Bands.
Journal Acoust. Soc. Amer., Vol. 33, no. 4, p. 248, Feb. 1961.

Web-Seiten

- [WWW1]** Analoge Audiosignale
<http://audio-uni-lueneburg.de/texte/digaudio/analog.htm>

Anhang A

A.1 Abkürzungsverzeichnis

A

AAC	Advanced Audio Coding
AAL	ATM Adaptation Layer
AL	ATM Layer
ACK	Acknowledgement
ADPCM	Adaptive Difference Pulsecodemodulation
ARQ	Automatic Repeat Request
ATM	Asynchronous Transfer Mode

B

B-ISDN	Broadband Integrated Services Digital Network
--------	---

C

CELP	Code Excited Linear Prediction
CRC	Cyclic Redundancy Check

D

DAT	Digital Audio Tape
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DPCM	Difference-PCM
DWT	Discrete Wavelet Transform

F

FEC	Forward Error Correction
FTR	Frequenz-Tracking

G

GI	Gewichtete Interpolation
GSM	Global System Mobile

H

HCR	Huffman Codeword Reordering
HILN	Harmonic and Individual Lines plus Noise

I

IP	Internet Protocol
ITU	International Telecommunication Union

L

LAN	Local Area Network
LDS	Leistungsdichtespektrum
LDSM	LDS Morphing
LMS	Least Mean Square
LPC	Linear Predictive Coding
LSP	Line Spectral Parameter

M

MDCT	Modified Discrete Cosine Transform
MK	Maximale Kreuzkorrelation
MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
MQ	McAulay und Quartieri

O

ODG	Objective Difference Grade
-----	----------------------------

P

PEAQ	Perceptual Evaluation of Audio Quality
PARCOR	Partial Correlation
PCM	Pulsocodemodulation
PL	Physical Layer
PM	Pattern Matching
PODG	Positive Objective Difference Grade
PSN	Packet Switched Network

Q

QOS	Quality of Service
-----	--------------------

R

RSVP	Resource Reservation Protocol
RTP	Realtime Transport Protocol
RTSP	Realtime Streaming Protocol
RTT	Round Trip Time
RVLC	Reversible Variable Length Codes

S

SLPC	Stereo Linear Predictive Coding
SN	Sequence Number
SNP	Sequence Number Protection
SNR	Signal-to-Noise Ratio
SNR_{seg}	Segmentorientierter SNR
SNR_{f_0-seg}	Frequenzgewichteter SNR

T

TCP	Transmission Control Protocol
-----	-------------------------------

U

UDP	User Datagram Protocol
-----	------------------------

V

VCB	Virtual Codebook
VCI	Virtual Channel Identifier
VPI	Virtual Path Identifier

Z

zcr	Zero Crossing Rate
-----	--------------------

A.2 Verzeichnis der verwendeten Formelzeichen

\underline{A}	Matrix
$A_i^k(n)$	Augenblicksamplitude des i -ten Frequenztrack im k -ten Segment zum Zeitpunkt n
a_i, b_i, c_i, d_i	Filterkoeffizienten des Stereoprädiktors
$C(n)$	Kreuzkorrelationsmaß
$D(n), d(n)$	Differenzsignal
ΔL	Länge des Mergingsintervalls
f_a	Abtastfrequenz
f_i^k	Frequenz des i -ten Sinusoid in Segment k
$\varphi_i^k(n)$	Augenblicksphase des i -ten Sinusoids in Segment k zum Zeitpunkt n
$H(z)$	Übertragungsfunktion eines Filters
$I(n)$	Anzahl der Sinusoids in einem Zeitabschnitt
K	Prädiktorordnung
K_a	Filterordnung des Autoprädiktors des linken Kanals
K_b	Filterordnung des Kreuzprädiktors des linken Kanals
K_c	Filterordnung des Autoprädiktors des rechten Kanals
K_d	Filterordnung des Kreuzprädiktors des rechten Kanals
L	Paketlänge
$L(n)$	PCM-Werte des linken Kanals
L_a	Länge des Analysebereichs beim FTR-Verfahren
l_1	Puffer-Länge für die Korrelationsuntersuchung im MK-Verfahren (wird in den Verfahren MK und LPC-GI verwendet)
l_2	Länge des Kreuzkorrelationsfeldes
L_{DFT}	Länge des DFT-Vektors
M_k	Anzahl der getrackten Partial-Paare in Segment k
M_{k-1}, M_{k+1}	Anzahl der Partial im vorgänger- (bzw. nachfolger-) Analysebereich
$\max()$	Funktion zur Bestimmung des Maximums in einem Feld
$\min()$	Funktion zur Bestimmung des Minimums in einem Feld
p_i	PARCOR-Koeffizienten
P_V	Paketverlustrate
$R(n)$	PCM-Werte des rechten Kanals
R_{LL}	Autokorrelation des linken Kanals
R_{LR}, R_{RL}	KKF der Kanäle Links/Rechts bzw. Rechts/Links
R_{RR}	Autokorrelation des rechten Kanals

\underline{S}_k	Matrix mit den gültigen Frequenzen, ihre Phasen und Amplituden in Segment k
$x(n)$	Zeitdiskretes Signal
$\hat{\mathbf{x}}(n)$	Schätzsignal
z_i	Polstelle
z_i^*	Konjugiert komplexe Polstelle zu z_i
zcr	zero crossing rate

Anhang B

B.1 Polstellenverschiebung

In [Ant93] wurde bewiesen, dass eine instabile Polstelle z_i ins Innere des Einheitskreises (auf der Z-Ebene) verschoben werden kann (neue Polstelle: $1/z_i^*$) ohne Veränderung des Amplitudenganges. Der Verstärkungsfaktor der Übertragungsfunktion wird mit einem Vorfaktor V ($V = 1/|z_i|$) multipliziert.

Es folgt ein trigonometrischer Beweis:

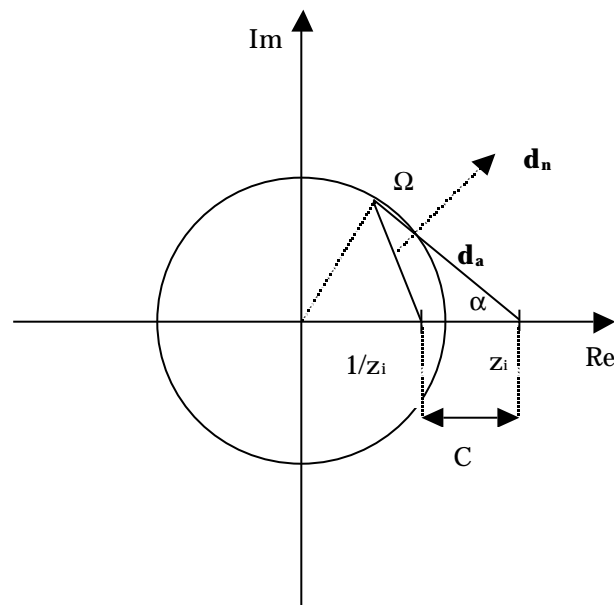


Bild B1.1: Polstellenverschiebung

Es ist also zu beweisen, dass $d_n = V \cdot d_a = d_a / |z_i|$.

Gegeben: d_a, z_i, C .

Gesucht: $d_n = f(d_a)$.

In diesem Beispiel ist die Polstelle reell. Es gilt also:

$$|z_i| = z_i ; z_i^* = z_i.$$

Es gilt:

$$d_n^2 = d_a^2 + C^2 - 2d_a \cdot C \cdot \cos(\alpha) \quad (\text{B1.1})$$

$$1^2 = z_i^2 + d_a^2 - 2d_a \cdot z_i \cdot \cos(\alpha) \quad (\text{B1.2})$$

$\cos(\alpha)$ aus Gleichung (B1.2) ermitteln:

$$\cos(\alpha) = \frac{d_a^2 + z_i^2 - 1}{2 \cdot z_i \cdot d_a} \quad (\text{B1.3})$$

Einsetzen von (B1.3) in (B1.1) ergibt:

$$d_n^2 = d_a^2 + C^2 - 2 \cdot d_a \cdot C \cdot \frac{d_a^2 + z_i^2 - 1}{2 \cdot z_i \cdot d_a} = d_a^2 + C^2 - \frac{C}{z_i} (d_a^2 + z_i^2 - 1), \text{ wobei}$$

$$C = z_i - \frac{1}{z_i} = \frac{z_i^2 - 1}{z_i}$$

$$\begin{aligned} \Rightarrow d_n^2 &= d_a^2 + \frac{(z_i^2 - 1)^2}{z_i^2} - \frac{1}{z_i} * \frac{(z_i^2 - 1)}{z_i} (d_a^2 + z_i^2 - 1) \\ &= d_a^2 \cdot \left[1 - \frac{(z_i^2 - 1)}{z_i^2} \right] + \frac{(z_i^2 - 1)^2}{z_i^2} - \frac{(z_i^2 - 1)}{z_i^2} (z_i^2 - 1) \\ &= d_a^2 \cdot \frac{1}{z_i^2} \end{aligned}$$

$$\Rightarrow d_n = \frac{d_a}{|z_i|} \quad (\text{B1.4})$$

B.2 Quadratische Interpolation zur genauen Frequenzbestimmung.

Wegen der diskreten Natur der DFT mit den Rasterfrequenzen gleicher Frequenzabstände ist die Frequenzbestimmung nur ungenau. Der Fehler gegenüber der genauen Frequenz beträgt $\pm \Delta f/2$. Zur Ermittlung der genaueren Frequenz wird in der Literatur auf die quadratische Interpolation (*parabolic regression*) verwiesen [McI92] [Mas96].

Bei diesem Interpolationsverfahren wird angenommen, dass die Form der Hauptkeule des Spektrums der vor der DFT verwendeten Fensterfunktion durch eine Parabel angenähert werden kann. Die wirkliche Abweichung der realen Frequenz f_r von der Rasterfrequenz f_k kann unter dieser Annahme bestimmt werden. Für die Interpolation werden die Leistungen der Spektrallinie f_k , f_{k-1} und f_{k+1} herangezogen (Bild B2.1).

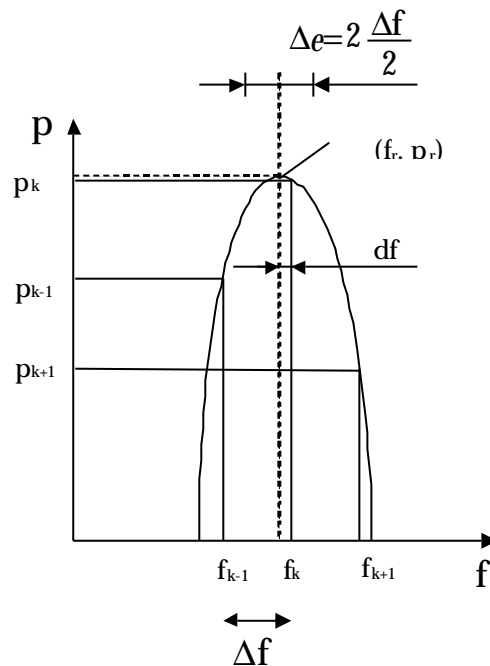


Bild B2.1: *Parabolic regression*

In [McI92] werden folgende Berechnungsformeln für die Frequenzabweichung und der Leistung der realen Frequenz eingeführt:

$$df = \frac{\Delta f}{2} \cdot \frac{p_{k-1} - p_{k+1}}{p_{k-1} - 2p_k + p_{k+1}} \quad (\text{B2.1})$$

$$p_r = p_k - \frac{1}{8} \cdot \frac{(p_{k-1} - p_{k+1})^2}{p_{k-1} - 2p_k + p_{k+1}}, \quad (\text{B2.2})$$

wobei,

p_{k-1} : Leistung der Spektrallinie f_{k-1}

p_k : Leistung der Spektrallinie f_k

p_{k+1} : Leistung der Spektrallinie f_{k+1}

Δf : Frequenzauflösung der DFT

df : Frequenzabweichung von der realen Frequenz und

p_r : Leistung der realen Frequenz $f_r = f_k + df$ bezeichnen.

B.3 Methode der kleinsten Fehler-Quadrate für die Ermittlung der Amplitude und Phase einer Sinuskomponente

In diesem Abschnitt wird erläutert, wie die Methode der kleinsten Quadrate für die genauere Ermittlung der Komponenten eines *Partial* herangezogen werden kann. Diese Methode wird hier benutzt, um die Phase und Amplitude bzw. Real- und Imaginär-Teil einer vorgegebenen Partial-Frequenz ω_k ¹ zu ermitteln.

Für eine genauere Berechnung muss ω_k zuerst korrigiert werden. Für die Frequenzkorrektur wird die quadratische Interpolation (Anhang B.4) verwendet.

Ein Segment der Länge L eines Audiosignals x wird durch die Funktion $f(i)$ angenähert, wobei

$$f(i) = a_k \cos(\omega_k \cdot i) + b_k \sin(\omega_k \cdot i), \quad i = 0..L-1 \quad (\text{B3.1})$$

Die Schätzung der Parameter a_k und b_k der Funktion $f(i)$ erfolgt nach Berechnung und Minimierung des quadratischen Fehlers E zwischen $f(i)$ und x_i , wobei x_i die Abtastwerte des Audiosegments x darstellen.

$$E = \sum_{i=0}^{L-1} (f(i) - x_i)^2 = \sum_{i=0}^{L-1} ([a_k \cos(\omega_k \cdot i) + b_k \sin(\omega_k \cdot i)] - x_i)^2 \stackrel{!}{=} \min \quad (\text{B3.2})$$

Es muss also gelten:

$$\frac{dE}{da_k} = 2 \cdot \sum_{i=0}^{L-1} ([a_k \cos(\omega_k i) + b_k \sin(\omega_k i)] - x_i) \cdot \cos(\omega_k i) \stackrel{!}{=} 0 \quad (\text{B3.3})$$

und

$$\frac{dE}{db_k} = 2 \cdot \sum_{i=0}^{L-1} ([a_k \cos(\omega_k i) + b_k \sin(\omega_k i)] - x_i) \cdot \sin(\omega_k i) \stackrel{!}{=} 0 \quad (\text{B3.4})$$

Für die Ermittlung der Parameter a_k und b_k muss also das folgende Gleichungssystem gelöst werden.

$$\begin{bmatrix} \sum_{i=0}^{L-1} \cos^2(\omega_k i) & \sum_{i=0}^{L-1} \sin(\omega_k i) \cdot \cos(\omega_k i) \\ \sum_{i=0}^{L-1} \sin(\omega_k i) \cdot \cos(\omega_k i) & \sum_{i=0}^{L-1} \sin^2(\omega_k i) \end{bmatrix} \cdot \begin{bmatrix} a_k \\ b_k \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{L-1} x_i \cos(\omega_k i) \\ \sum_{i=0}^{L-1} x_i \sin(\omega_k i) \end{bmatrix} \quad (\text{B3.5})$$

a_k und b_k sind der Real- und Imaginär-Teil der k -ten Frequenzlinie mit der Frequenz ω_k . Die Amplitude A_k und Phase φ_k dieser Frequenzlinie lassen sich wie folgt ermitteln:

¹ ω_k stellt den k -ten Frequenzbin im DFT-Vektor dar.

$$\begin{aligned} A_k &= \sqrt{a_k^2 + b_k^2} \\ \varphi_k &= \arctan\left(\frac{b_k}{a_k}\right) \end{aligned} \tag{B3.6}$$

Die Parameter a_k und b_k werden einzeln für jede Frequenz ω_k des untersuchten Audio-segments ermittelt.

Anhang C: Simulationswerte

Die Bewertung der in Kapitel 4 vorgestellten Verfahren wurde sowohl objektiv als auch subjektiv durchgeführt. Bei der objektiven Bewertung wurden PODG- und SNRseg-Werte ermittelt. Die subjektive Bewertung erfolgte anhand von Hörtests. Es haben insgesamt 15 Testpersonen an den Hörtests teilgenommen. Aufgrund der hohen Anzahl der untersuchten akustischen Beispiele und dem damit verbundenen Aufwand der Hörtests wurde die Anzahl der Hörproben bei den Hörtests reduziert.

Für die Untersuchung wurden folgende Audiobeispiele aufgrund ihrer unterschiedlichen Eigenschaften ausgewählt:

- Titel 1: Einzelnes Instrument (Orgel).
- Titel 2: Klassische Musik (Beethoven)
- Titel 3: Pop Musik (Marusha)

Die untersuchten Musik-Titel haben jeweils eine Dauer von 20 Sekunden. Es wurden einzelne Paketverluste mit der Paketverlustwahrscheinlichkeit $P_v = 1\%$, 5% und 10% untersucht. Die Paketlänge L wurde variiert. Hierbei wurden folgende Paketlängen (in Abtastwerte) verwendet: 1024 (22,22 ms), 512 (11,61 ms) und 256 (5,8 ms).

C.1.1 Substitution des PCM-Signals

Nullsubstitution

Bei der Nullsubstitution des PCM-Signals, die als Referenz dient, wird kein *merging* angewendet.

P_v	L	Titel 1			Titel 2			Titel 3		
		1024	512	256	1024	512	256	1024	512	256
1%	PODG	4,087	4,092	3,733	3,456	3,314	2,836	4,228	4,385	4,009
	MOS	2,43	-	-	3,11	-	-	2,85	-	-
	SNRseg	47,644	47,475	46,863	49,348	49,202	48,62	49,359	49,205	48,644
5%	PODG	3,538	2,432	1,26	1,806	1,284	1,112	3,54	3,51	2,97
	MOS	1,71	-	-	2,18	-	-	2,21	-	-
	SNRseg	44,493	43,122	40,804	46,258	44,676	42,081	46,248	44,698	42,092
10%	PODG	1,949	1,215	1,145	1,152	1,119	1,097	3,087	2,221	1,818
	MOS	1,37	1,17	1,13	1,55	1,64	1,7	1,87	1,88	1,89
	SNRseg	40,866	38,198	33,878	42,738	39,88	35,155	42,73	39,89	35,241

Tabelle C.1: Simulationswerte bei PCM-Null

Paketwiederholung:

Merging: ungefähr 10% der Paketlänge: 1024 → 100; 512 → 50; 256 → 26.

Pv \ L	Titel 1			Titel 2			Titel 3		
	1024	512	256	1024	512	256	1024	512	256
PODG	4,336	4,337	4,144	4,443	4,377	4,315	4,014	4,177	3,705
1% MOS	3,7	-	-	4,66	-	-	4,3	-	-
SNRseg	47,484	47,253	46,7	49,218	49,024	48,496	49,136	48,993	48,424
PODG	4,141	3,652	1,802	4,023	3,784	3,272	3,258	3,13	2,676
5% MOS	2,66	-	-	4,43	-	-	3,67	-	-
SNRseg	44,32	42,926	40,474	45,945	44,166	41,365	45,548	44,061	41,056
PODG	3,438	1,772	1,244	3,654	3,065	2,071	2,797	2,403	1,95
10% MOS	2,04	1,66	1,47	3,48	2,69	2,33	2,92	2,97	2,35
SNRseg	40,066	37,3	32,991	42,08	39,032	33,96	41,366	38,671	33,497

Tabelle C.2: Simulationswerte bei PCM-Wiederholung

Links-/ Rechts-Ersatz

Merging = 10% der Paketlänge.

Pv \ L	Titel 1			Titel 2			Titel 3		
	1024	512	256	1024	512	256	1024	512	256
PODG	4,253	4,271	3,93	3,944	3,803	2,948	4,299	4,473	4,528
1% MOS	3,19	-	-	3,88	-	-	4,61	-	-
SNRseg	47,512	47,314	46,818	49,235	49,068	48,545	49,448	49,267	48,941
PODG	3,128	1,722	1,202	2,031	1,46	1,127	3,692	4,151	4,23
5% MOS	2,08	-	-	3,1	-	-	4,58	-	-
SNRseg	44,61	43,091	40,854	46,061	44,406	41,924	46,965	45,844	43,654
PODG	1,303	1,178	1,155	1,268	1,136	1,114	3,607	4,058	3,959
10% MOS	1,49	1,32	1,19	2,49	2,07	1,77	4,65	4,67	4,6
SNRseg	40,568	37,843	33,794	42,325	39,448	34,79	44,057	42,255	38,508

Tabelle C.3: Simulationswerte bei PCM-LR-Ersatz

Maximale Korrelation (MK):

Es wurden folgende Parameter verwendet:

Für die Paketlänge 1024: Merging = 100 und Pufferlänge = 3500 AW.

Für die Paketlänge 512: Merging = 50 und Pufferlänge = 1800 AW.

Für die Paketlänge 256: Merging = 26 und Pufferlänge = 900 AW.

Pv \ L	Titel 1			Titel 2			Titel 3		
	1024	512	256	1024	512	256	1024	512	256
PODG	4,418	4,532	4,572	4,444	4,433	4,372	4,719	4,684	4,728
1% MOS	4,33	-	-	4,45	-	-	4,8	-	-
SNRseg	47,608	47,532	47,228	49,231	49,057	48,582	49,335	49,19	48,732
PODG	4,298	4,436	4,315	3,984	3,86	3,48	4,409	4,45	4,433
5% MOS	3,74	-	-	4,06	-	-	4,64	-	-
SNRseg	44,976	44,303	42,835	45,8	44,37	41,855	46,511	45,316	43,167
PODG	4,021	4,223	3,809	3,376	3,054	2,301	4,296	4,266	4,198
10% MOS	3,2	3,3	2,91	3,31	2,8	2,5	4,67	4,56	4,58
SNRseg	41,392	40,015	37,554	41,71	39,193	34,663	43,233	41,432	37,352

Tabelle C.4: Simulationsergebnisse bei MK

C.1.2 Substitution der LPC-Innovation

Hier wird die Filterordnung des Prädiktors auf 20 gesetzt.

Es wird im folgenden angenommen, dass ein Paketverlust nur im linken Kanal auftritt. Der rechte Kanal bleibt fehlerfrei.

Links-/Rechts-Ersatz: (LPC-LR)

Pv \ L	Titel 1			Titel 2			Titel 3		
	1024	512	256	1024	512	256	1024	512	256
PODg	4,226	4,289	4,148	4,333	4,41	4,356	4,714	4,701	4,7
1% MOS	3,4	-	-	4,28	-	-	4,89	-	-
SNRseg	47,594	47,448	46,862	49,315	49,159	48,61	49,501	49,338	48,881
PODg	3,545	2,732	1,394	3,883	3,804	3,504	4,402	4,417	4,427
5% MOS	2,31	-	-	3,44	-	-	4,86	-	-
SNRseg	44,3	43,198	40,964	46,107	44,526	42,029	46,891	45,83	43,56
PODg	2,206	1,322	1,195	3,436	2,943	2,388	4,107	4,267	4,218
10% MOS	1,75	1,35	1,2	3,16	2,59	2,38	4,81	4,82	4,62
SNRseg	40,426	38,035	33,82	42,276	39,511	34,954	43,886	42,035	38,132

Tabelle C.5: Simulationsergebnisse bei LPC-LR

Gewichtete Interpolation (GI)

Pv \ L	Titel 1			Titel 2			Titel 3		
	1024	512	256	1024	512	256	1024	512	256
PODg	4,401	4,481	4,477	4,672	4,405	4,299	4,145	4,285	4,557
1% MOS	4,04	-	-	4,58	-	-	4,85	-	-
SNRseg	47,686	47,562	47,183	49,312	49,166	48,91	49,435	49,274	48,858
PODg	4,329	4,341	4,093	3,828	3,849	3,343	3,759	4,249	4,223
5% MOS	3,11	-	-	3,85	-	-	4,41	-	-
SNRseg	44,45	44,209	42,516	45,99	44,514	41,91	46,45	45,462	43,14
PODg	3,974	4,073	2,777	3,083	2,91	2,053	3,619	3,907	3,96
10% MOS	2,72	2,46	1,86	2,84	2,54	2,28	4,35	4,32	4,48
SNRseg	41,45	39,943	36,864	42,1	39,389	34,908	43,431	41,463	37,515

Tabelle C.6: Simulationsergebnisse bei LPC-GI

C.1.3 Substitution mit der Stereoprädiktion

Ordnung des Monoprädiktors = 10.

Ordnung des Kreuzprädiktors = 5.

Paketverlust nur im linken Kanal. Der rechte Kanal bleibt fehlerfrei.

Bei den Titeln 1 und 2 wurde eine Wiederholung des letzten Differenzsignals verwendet, für Titel 3 scheint die Nullsubstitution besser geeignet zu sein.

Pv \ L	Titel 1			Titel 2			Titel 3		
	1024	512	256	1024	512	256	1024	512	256
PODG	4,311	4,384	4,153	4,423	4,322	4,281	4,304	4,393	4,408
1% MOS	3,78	-	-	4,52	-	-	4,48	-	-
SNRseg	47,495	47,262	46,743	49,202	49,014	48,71	49,488	49,108	48,674
PODG	4,107	3,75	1,707	3,91	3,727	3,039	4,01	3,689	3,622
5% MOS	2,56	-	-	4,2	-	-	3,14	-	-
SNRseg	44,05	42,12	40,353	45,941	44,202	41,388	46,546	44,648	42,18
PODG	4,332	1,664	1,245	3,462	2,989	1,823	3,676	3,153	2,741
10% MOS	1,97	1,49	1,39	3,43	2,63	2,04	2,59	2,45	2,37
SNRseg	39,938	37,495	32,726	42,004	38,962	33,961	43,137	39,805	35,463

Tabelle C.7: Simulationsergebnisse bei SLPC

C.1.4 Substitution im Frequenzbereich

Spektrales-Morphing (SM)

Hier wurde der Merging-Bereich automatisch auf 50 Prozent der Paketlänge festgelegt.

Pv \ L	Titel 1			Titel 2			Titel 3		
	1024	512	256	1024	512	256	1024	512	256
PODG	4,43	4,541	4,563	4,192	4,372	4,469	3,622	3,824	4,379
1% MOS	4,27	-	-	4,27	-	-	4,08	-	-
SNRseg	47,179	46,974	46,79	48,913	48,702	48,285	48,878	48,682	48,26
PODG	3,7	4,162	4,306	2,907	3,737	3,58	3,031	3,703	3,6
5% MOS	2,89	-	-	3,23	-	-	3,42	-	-
SNRseg	42,826	42,222	41,504	44,623	43,407	41,318	44,57	43,573	41,121
PODG	2,068	3,262	3,292	1,49	2,352	1,676	1,969	2,9	2,374
10% MOS	2	1,81	2,91	2,29	2,8	2,55	2,65	2,35	2,25
SNRseg	37,381	36,2	35,05	39,123	37,48	33,658	39,1	37,711	33,402

Tabelle C.8: Simulationsergebnisse bei SM

Frequenz-Tracking (FTR)

Die Länge der Analysebereiche vor der Lücke und danach ist unabhängig von der Paketlänge und beträgt 1024 AW.

Es wird eine DFT der Auflösung 2048 angewendet.

L	Titel 1			Titel 2			Titel 3		
	1024	512	256	1024	512	256	1024	512	256
PODG	4,636	4,713	4,598	4,356	4,491	4,189	3,806	3,792	4,34
1% MOS	4,56	-	-	4,41	-	-	4,47	-	-
SNRseg	47,736	48,231	48,555	49,221	49,672	49,543	49,306	49,726	49,9
PODG	4,519	4,526	4,343	4,076	3,974	3,44	3,759	4,005	3,807
5% MOS	4,29	-	-	3,97	-	-	4,25	-	-
SNRseg	44,973	46,85	47,203	46,157	46,848	46,403	46,062	47,69	47,537
PODG	4,311	4,291	3,946	3,382	3,05	2,826	3,042	3,532	3,311
10% MOS	4,21	3,77	3,61	3,82	3,42	2,56	4,12	3,99	3,92
SNRseg	42,106	44,64	44,3	42,207	43,625	41,877	42,137	44,816	42,812

Tabelle C.9: Simulationsergebnisse bei FTR