Zeitschrift: Technische Mitteilungen / Schweizerische Post-, Telefon- und Telegrafenbetriebe = Bulletin technique / Entreprise des postes, téléphones et télégraphes suisses = Bollettino tecnico / Azienda delle poste, dei telefoni e dei telegrafi svizzeri

Herausgeber: Schweizerische Post-, Telefon- und Telegrafenbetriebe

Band: 64 (1986)

Heft: 4

Artikel: Codierverfahren von Sprach- und Musiksignalen und ihre Beurteilung mit Computer

Autor: Mury, Rolanf

DOI: https://doi.org/10.5169/seals-875024

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 25.07.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Codierverfahren von Sprach- und Musiksignalen und ihre Beurteilung mit Computer

Roland MURY, Bern

Zusammenfassung. *Der vorliegende Artikel soll in gedrängter Form einen allgemeinen Überblick über die hauptsächlichen Verfahren der digitalen Kompression von Sprach- und Musiksignalen geben, mit Hinweisen auf die Sprachsynthese, wie sie bei der Automatisierung der Auskunftsdienste der PTT angewandt wird. Nach der Erläuterung grundsätzlicher Begriffe der in der Fernmeldetechnik am meisten angewandten Codierungs-Methoden werden kurz die Hardware und die Programmstruktur untersucht, die für das Studium der Signalkompressions-Algorithmen verwendet werden (ADDIO-2). Der Artikel schliesst mit einer nicht erschöpfenden Beschreibung der Spezifikationen, der subjektiven Versuche und der Bedingungen für die Aufnahme, die Verarbeitung und das Abhören zur qualitativen Beurteilung von Sprachmustern.*

Procédés de codage des signaux de la parole et de la musique et leur évaluation à l'aide de l'ordinateur

Résumé. *L'objectif de la présente publication est de donner un aperçu général succint des principaux procédés de compression numérique des signaux de la parole et de la musique en passant par la synthèse de la parole appliquée à l'automatisation des services des renseignements de l'Entreprise des PTT. Après avoir rappelé quelques notions fondamentales des méthodes de codage les plus couramment appliquées dans les télécommunications, on examine brièvement l'architecture matérielle de l'équipement et la structure du logiciel utilisées pour l'étude des algorithmes de compression du signal (ADDIO-2). L'article s'achève par une description non exhaustive des spécifications des méthodes d'essais subjectifs et des conditions d'enregistrement, de traitement et d'écoute pour l'évaluation qualitative des échantillons de parole.*

Procedimenti di codifica dei segnali vocali e musicali e loro valutazione mediante elaboratore

Riassunto. *Nel presente articolo sono presentati sommariamente i principali procedimenti di compressione numerica dei segnali vocali e musicali, con alcune indicazioni sulla sintesi della voce, adottati per automatizzare i servizi delle informazioni delle PTT. Dopo aver spiegato i concetti basilari dei metodi di codifica più utilizzati nella tecnica delle telecomunicazioni, l'autore esamina brevemente l'hardware e la struttura del programma impiegati per lo studio degli algoritmi di compressione dei segnali (ADDIO-2). L'articolo termina con una descrizione non esaustiva delle specificazioni delle prove soggettive e delle condizioni di registrazione, di trattamento e di ascolto per la valutazione qualitativa dei campioni di segnali vocali.*

1 Einleitung

Aus Wirtschaftlichkeitsüberlegungen ist es wünschbar, die Bitrate von digital codierten Signalen merklich zu verkleinern. Die Kompression–Expansion erlaubt, die Anzahl übertragener Bit je Abtastwert zu vermindern. Noch vor einigen Jahren konnte man die Wirksamkeit von Kompression-Expansions-(Kompander-)Verfahren nur mit besonderen Schaltungen studieren. Daher war es sehr oft schwierig, ohne grossen Zeitverlust einen Parameter zu ändern oder die Codierungs- oder Kompressionseigenschaften zu optimieren. Heute erledigt der Computer diese Codierungsoperationen oder alle anderen Signalverarbeitungsverfahren in Echtzeit oder im Batch-Modus, je nach Komplexität des Algorithmus. So können die verschiedenen Parameter des zu simulierenden Systems nach Belieben verändert werden. Die Hauptabteilung Forschung und Entwicklung der PTT hat eine programmierbare Einrichtung zur Ein- und Ausgabe von Tonsignalen in einem Grossrechner aufgebaut. Dieses System, genannt ADDIO-2 nach der englischen Bezeichnung «*A*nalog-*D*igital-*D*ata-*I*nput-*O*utput», stellt heute ein unentbehrliches Werkzeug für den Ingenieur dar, der die schwierige Aufgabe hat, den besten Kompromiss zwischen sehr ähnlichen Lösungen zu finden. Die Studien zur 64-kbit/s-Codierung von 7-kHz-Sprache sowie zu den Kompression-Expansions-Verfahren von Hörfunk-Signalen bei 384 kbit/s haben stark von den Möglichkeiten von ADDIO-2 profitiert. Von den zahlreichen im System verfügbaren Programmen, die für Simulationsversuche oder Rechenoperationen verwendet wurden, seien folgende herausgegriffen:

- Simulation der Gesetze der Momentanwert- oder Quasi-Momentanwert-Kompander zur Codierung und Übertragung von Radioprogrammen
- Versuche mit verschiedenen Codier-Algorithmen für Sprache
- Simulation von Quadratur-Spiegel-Filtern für die Subband-Codierung ADPCM
- Kontrolle der Wirksamkeit fehlerkorrigierender Code, angewandt auf digitale Übertragungssysteme von Hörfunkprogrammen
- Messung und Korrektur des Effektivwertes von Sprachsignalen für die subjektiven Hörversuche
- Simulation gewisser Transcodier-Operationen
- Erzeugung und Schneiden von Tonquellen für die subjektiven Hörversuche.

In diesem Zusammenhang werden nachfolgend einige Eigenschaften der Codierung untersucht, besonders jene, die Gegenstand eines kürzlichen Forschungs- und Studienprojektes im Rahmen einer auf die schweizerischen PTT zugeschnittenen Anwendung waren.

2 Methoden der digitalen Signalkompression

Die Codieroperation besteht darin, den Wert des analogen Signals zu diskreten Zeitpunkten zu messen und ihn durch eine aus einer endlichen Anzahl Werten gewählte Zahl darzustellen. Dieser zweite Teil der Operation, die sog. Quantifizierung, kann verschiedenartig ausgeführt werden. Die verbreitetsten Kompressionsmethoden zur Verminderung der Bitrate des digitalen Signals sind in *Figur 1* zusammengestellt. In einem Artikel in den IEEE Transactions on Communication [1] unterscheidet *J. Flanagan* vier grobe Methodenkategorien je nach Qualität, die sie ermöglichen, nämlich:

- Hörfunk-Qualität (30 Hz...8000 Hz)
- Telefon-Qualität (300 Hz...3400 Hz)

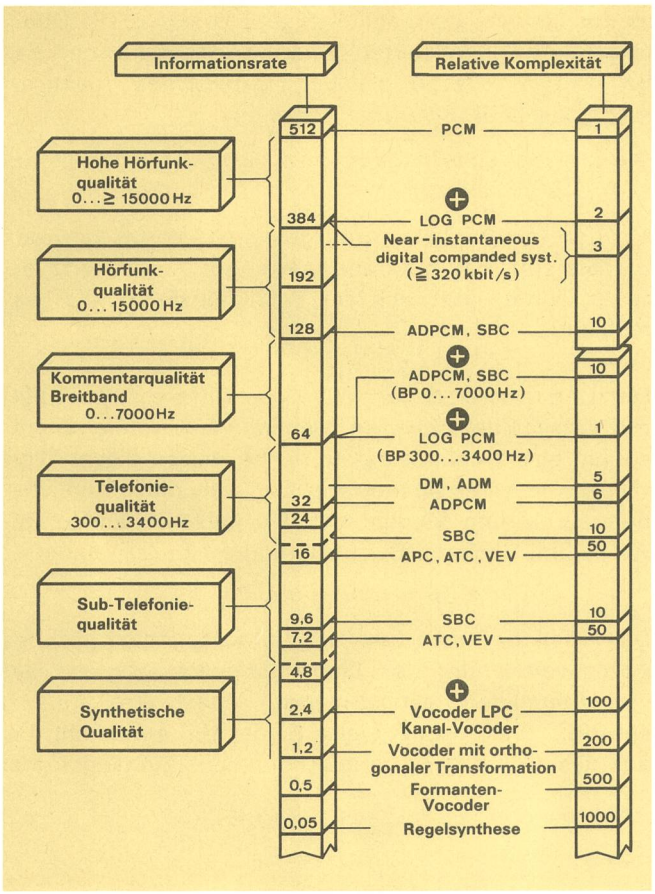

Fig. 1
Digitale Kompander-Verfahren, geordnet nach Qualität, Informationsrate und Komplexität

ADPCM	Adaptive Differenz-PCM-Codierung
SBC	Subband-Codierung
DM	Deltamodulations-Codierung
ADM	Adaptive Deltamodulations-Codierung
APC	Adaptive Prädiktions-Codierung
ATC	Adaptive Transformations-Codierung
VEV	Sprachgesteuerter Vocoder
⊕	Forschungsprojekt der PTT

- Sub-Telefon-Qualität
- Synthetische Qualität

Die Fortschritte der letzten Jahre erlauben, diese Klassifikation zu verfeinern und die Begriffe «hohe Qualität» und «Kommentar» für Radioprogramme und «breitbandige Sprache» für die Telefonie einzuführen.

Die erste Kategorie umfasst die Methoden, die dem decodierten Signal eine hohe Qualität für die Übertragung von Hörfunkprogrammen und eine sogenannte «Kommentarqualität», entsprechend einer Bandbreite von ungefähr 7 kHz, z. B. für Reportagezwecke, ermöglichen. Die Bitraten liegen zwischen 64 kbit/s und 512 kbit/s je Tonkanal.

In die zweite Kategorie fallen die Methoden, die die Telefonbandbreite garantieren, aber auch jene, die seit kurzem erlauben, diese Bandbreite zu verdoppeln, ohne die Bitrate anzuheben. Diese Methoden beruhen auf der adaptiven Codierung. Sie finden neue Anwendungen bei Audio- und Videokonferenz, aber wahrscheinlich auch im künftigen ISDN (*I*ntegrated *S*ervices *D*igital *N*etwork = dienstintegriertes Digitalnetz). Die nötigen Bitraten bewegen sich zwischen 16 kbit/s und 64 kbit/s, man spricht von einer «transparenten» Qualität.

Die dritte Kategorie umfasst die Methoden, die eine erhöhte Kompression mit einer hörbaren Verzerrung des decodierten Signals zulassen. Dieses bleibt jedoch durchaus verständlich. Diese Kategorie wird oft sub-telefon genannt. Sie erfordert Bitraten von ungefähr 4,8 kbit/s...16 kbit/s.

Die Codierungsmethoden dieser drei Kategorien werden als Zeitcodierung bezeichnet, da sie in erster Linie die Eigenschaften des Signals im Zeitbereich zu erhalten trachten.

Die letzte Kategorie greift auf Analyse- und Syntheseverfahren zurück, wobei aus dem Sprachsignal eine kleine Anzahl Parameter extrahiert und in digitaler Form übertragen wird. Das decodierte Signal verliert so einen Teil seiner Verständlichkeit und Eigenart. Das Verhalten im Zeitbereich ist merklich verschieden von jenem des ursprünglichen Signals, was sich subjektiv in einer sogenannten synthetischen Qualität äussert.

Die Bitraten dieser sogenannten Quellencodierungs- oder Vocoder-Verfahren liegen zwischen einigen Hundert bit/s und ungefähr 4,8 kbit/s.

Die Wahl einer Kompressionsmethode hängt nicht nur von der Qualität und der erforderlichen Übertragungs-Bitrate, sondern auch von der Komplexität des Codecs ab. Diese drei grundsätzlichen Parameter werden in Figur 1 zueinander in Beziehung gesetzt. Diese Darstellung ist nicht erschöpfend, aber ausführlich genug, um das Verhältnis von Kosten zu Qualität der dem Fernmeldebereich am meisten angepassten Codierungsmethoden zu würdigen. Die in Figur 1 aufgeführten Kreuze bezeichnen ein neueres Forschungsprojekt der PTT.

3 Sprachcodierung

Die Einführung der digitalen Übertragungstechnik in den Fernmeldenetzen erfolgte auf der Grundlage der PCM-Methode mit 8 Bit und einer Abtastfrequenz von 8 kHz. Die Wahl dieser Methode drängte sich aus folgenden Gründen auf:

a) Diese einfache Codierungsart ergibt eine vom codierten Signal unabhängige Qualität
b) Sie ermöglicht eine genügende Qualität auch bei einer vernünftigen Anzahl Codierungen–Decodierungen zu bewahren, wie sie bei internationalen Verbindungen zu erwarten ist.

Heute erlaubt jedoch die ständig beschleunigte Entwicklung der Technologie, für Multiplexausrüstungen anspruchsvollere Verfahren der Informationsverarbeitung vorzusehen, um optimal die Redundanzeigenschaften der zu codierenden Signale auszunutzen. Zum Beispiel erlaubt die adaptive Differenz-PCM-Methode (ADPCM), die verfügbare Übertragungskapazität einer digitalen 2048-kbit/s-Verbindung, ohne merkliche Verschlechterung der Sprachqualität, besser auszunutzen. Daher zeigen die Fernmeldeunternehmen Interesse an einer digitalen Hierarchie differentieller PCM-Codecs, die auf der Bitrate von 64 kbit/s und Bruchteilen davon (32, 16, 8 kbit/s) beruhen. Eine Alternative, die wachsendem Interesse begegnet, ist die Codierung mit vergrösserter Bandbreite bei 64 kbit/s. Das in diesem Fall angestrebte Ziel ist eine Verdoppelung der Bandbreite eines Telefo-

niekanals und eine merkliche Verbesserung der Übertragungsqualität gegenüber PCM bei 64 kbit/s, für Anwendungen wie Audio- und Videokonferenz. Unter diesem Gesichtspunkt bietet die Sub-Band-Codierung gewisse Vorteile, wie später dargelegt werden wird. Leistungsfähigere Verfahren als ADPCM drängen sich bei Sprachsignalcodierung mit Bitraten zwischen ungefähr 16 kbit/s und 5 kbit/s auf. Die adaptive Prädiktions-Codierung (APC), die adaptive Transformations-Codierung (ATC), die Sub-Band-Codierung und die sprachgesteuerten Vocoder (VEV) sind die vier bekanntesten und wirkungsvollsten Methoden, die im betrachteten Bitratengebiet arbeiten.

31 Prinzip der Differenz-Pulscode-Modulation

Die Studien zur Differenz-Codierung begannen schon vor mehreren Jahren mit der Delta-Codierung, einem Verfahren, worin das Differenzsignal mit seinem Vorzeichen quantisiert wird. Diese sehr einfache Quantisierung wird durch eine erhöhte Abtastfrequenz verbessert oder ausgeglichen. Die Adaption des Quantisierers besteht somit darin, aufgrund des vorhergehenden Vorzeichens auf die Verstärkung der Prädiktor-Schlaufe einzuwirken (*Pissaro*, 1977). Die Leistungsfähigkeit dieses Codertyps ist ungenügend und allzuweit von der im Telefonnetz zulässigen Qualität entfernt.

Die Struktur der adaptiven Differenz-Codierung hingegen kann die Qualität des codierten Signals erheblich verbessern. Das Prinzip der Differenz-Codierung besteht darin, die Differenz e(n) zwischen dem Signal und einem aufgrund von vorhergehenden Werten s(n − 1), s(n − 2)... vorausberechneten Wert p(n) zu quantifizieren.

Die Differenz-PCM-Methode nützt die zeitliche Redundanz im Signal, das codiert werden soll, aus, indem für jeden kommenden Abtastwert eine Voraussage p(n) berechnet wird, die sich auf die vorher codierten und übermittelten Werte stützt. Das Quantisierungsrauschen wird dabei um so schwächer, je kleiner die zu quantisierende Differenz e(n) ist. Da das Differenzsignal e(n) eine nahezu gleichmässige spektrale Verteilung aufweist, nutzt es die Möglichkeiten des Coders viel wirkungsvoller aus als das ursprüngliche Eingangssignal. Die Übertragung des quantisierten Differenzsignals erfordert daher weniger Bits bei gleicher Leistung des Quantisierungsrauschens.

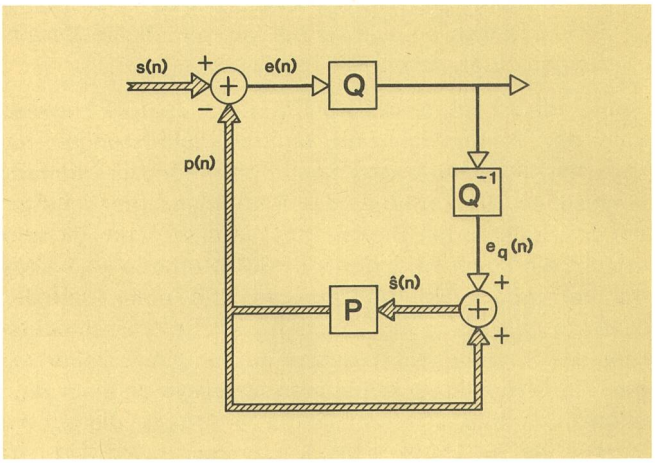

Fig. 2
Blockschaltbild des Differenz-Coders
Q Quantisierer
Q^{-1} Inverser Quantisierer
P Prädiktor

Die grundlegende Arbeitsweise des Differenzcodierers in *Figur 2* ist folgende:

Mit der Kenntnis des vorausgesagten Wertes p(n) des Signals ŝ(n), das aufgrund vorheriger Abtastwerte erzeugt wurde, lässt sich der Prädiktionsfehler e(n) berechnen

$$e(n) = s(n) - p(n)$$

Der Quantisierer Q vergleicht das Differenzsignal e(n) mit Entscheiderschwellen. Jedem Wert von e(n) entspricht eine Nummer des Bereichs, in den dieser Wert fällt. Ein inverser Quantisierer Q^{-1} setzt diese Bereichsnummer in den Medianwert des entsprechenden Bereichs um. Nach der Quantisierung erhält man daher

$$e_q(n) = e(n) + q(n)$$

Der Prädiktor P berechnet den Voraussagewert p(n) des Abtastwertes s(n) des Eingangssignals aufgrund der vorhergehenden rekonstruierten Abtastwerte ŝ(n − 1), ŝ(n − 2),... Das rekonstruierte Signal ŝ(n) entspricht, bis auf das Quantisierungsrauschen, dem Eingangssignal s(n):

$$\hat{s}(n) = e_q(n) + p(n)$$

oder

$$\hat{s}(n) = s(n) + q(n)$$

311 Adaptiver Quantisierer

Das Differenzsignal e(n) am Eingang des Quantisierers weist beträchtliche Pegelschwankungen auf. Daher ist es nötig, den Quantisierer an diese Schwankungen anzupassen und die Quantisierungsschwellen oder -stufen vom Pegel des Differenzsignals abhängen zu lassen.

In der von *Goodman* und *Wilkinson* vorgeschlagenen Adaption des Quantisierers, *Figur 3*, wird das Differenzsignal durch einen Wert Δ(n) geteilt, der proportional zur Signalenergie ist. Dieses Signal $e_s(n)$ wird anschliessend mit den Stufen eines Quantisierers mit festeingestellten Schwellen verglichen. Sei $\hat{e}_s(n)$ der zur Stufe I(n) gehörende Ausgangswert des Quantisierers, dann gilt:

$$\hat{e}(n) = \hat{e}_s(n)\,\Delta(n)$$

Die Adaption des Quantisierers besteht darin, die Sättigungsschwelle in Funktion des vorher gesendeten Pegels $\hat{e}_s(n)$ zu verändern. Daraus folgt der Adaptionsalgorithmus

$$\Delta(n+1) = [\Delta(n)]^{\gamma} \cdot M[(n)]$$

wobei M für jede Stufe I(n) zum voraus fest berechnet wird und γ = 0,98 ist (typischer Wert). Dieser Algorithmus ist einfacher auf einem Signalprozessor zu verwirklichen, wenn die Δ(n) und M in logarithmischer Form dargestellt werden. Dann gilt [2]:

$$d(n+1) = \gamma \cdot d(n) + m\,[I(n)]$$

312 Adaptiver Prädiktor

Der adaptive Prädiktor berechnet im wesentlichen den Schätzwert p(n) aufgrund des quantisierten Differenzsignals ê(n). Die Adaption des Prädiktors besteht darin,

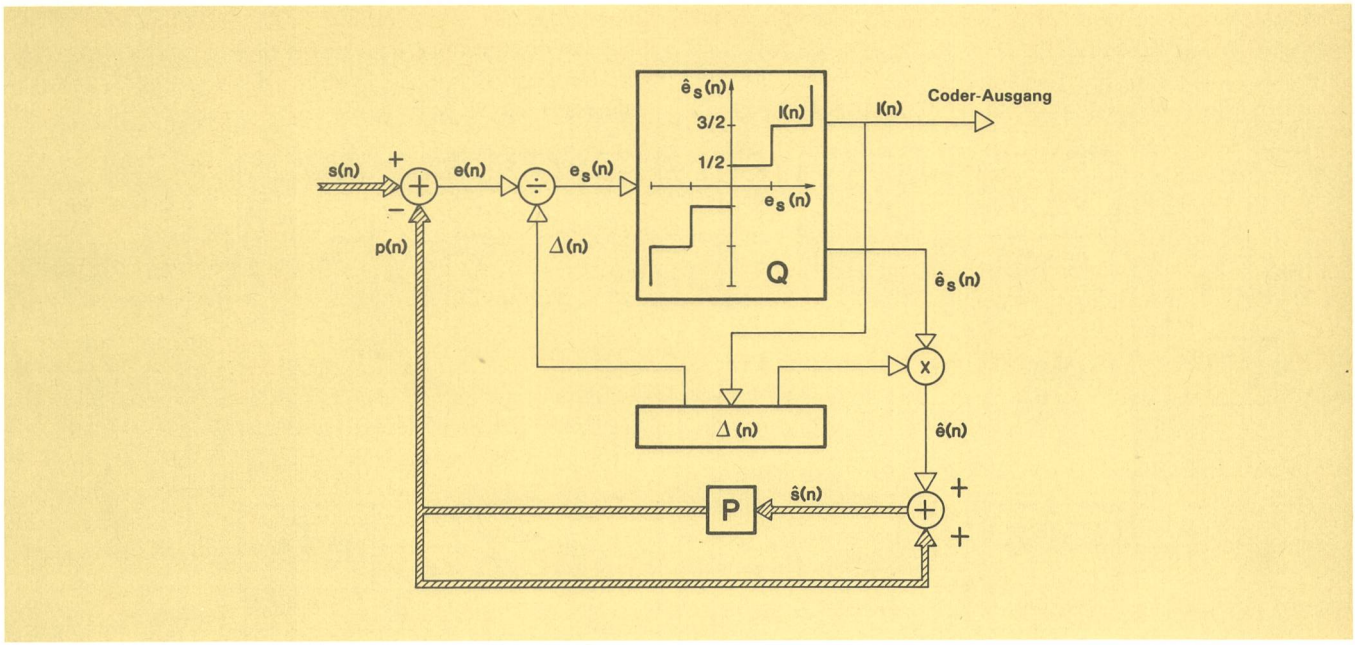

Fig. 3
Adaption des Quantisierers

die Koeffizienten eines linearen Filters je nach dem Verhalten der Korrelationsfunktion oder dem Spektrum des Signals s(n) zu verändern. Der Prädiktorfilter hat eine Übertragungsfunktion P(z), wobei

$$P(Z) = \sum_{i=1}^{N} a_i \cdot Z^{-i}$$

Die Ordnung N und die Koeffizienten a_i des Filters müssen so gewählt werden, dass die Leistung des Differenzsignals e(n) minimal wird. Die Koeffizienten werden nach der Formel [3]

$$R(k) = \sum_{i=1}^{N} a_i \cdot R(k - i),\ 0 \leqq k \leqq N$$

berechnet. Darin erscheinen die Elemente R(k) der Autokorrelationsfunktion des Signals s(n). Im allgemeinen wird N = 4 als vernünftiger Wert für die Ordnung des Prädiktorfilters angenommen. Man findet verschiedene Strukturen für den Prädiktor und zahlreiche Adaptionsalgorithmen. Eine häufig benutzte Struktur für die Verarbeitung von Sprachsignalen ist die lineare Prädiktion mit einem Kreuzschaltungsfilter [4].

32 Subband-Codierung

Bei diesem Codierverfahren wird das Eingangssignal auf eine Filterbank von nebeneinanderliegenden Bandpässen, die das ganze Nutzband überdecken, geführt. Die Ausgangssignale aller Filter werden mit Unterabtastung ins Basisband transponiert. Anschliessend werden die solchermassen gefilterten Signale getrennt codiert und quantisiert [5]. Auf der Empfangsseite wird die umgekehrte Operation ausgeführt: die Signale werden decodiert, mit Überabtastung transponiert, durch gleiche Bandpässe wie auf der Sendeseite gefiltert und schliesslich summiert, womit wieder das Sprachsignal entsteht. Dieser Vorgang ist in *Figur 4* mit dem Blockschaltbild eines Codiersystems mit zwei Subbändern veranschaulicht.

Dieses System wird mit zwei Quadratur-Spiegelfiltern (QMF) verwirklicht. Um über alles eine flache Übertragungsfunktion im ganzen Frequenzband zu gewährleisten, müssen sich die Filter ungefähr bei −3 dB überschneiden. Der Abtastvorgang bewirkt Überlappungen von Spektren eines Subbandes ins andere, was sich in mehr oder minder hörbaren Verzerrungen niederschlägt. Auch die Anwendung sehr steiler Filter löst das Problem nicht, weil dann andere, schwer beherrschbare Echovorgänge auftreten [6]. Die beste Lösung ist die Anwendung einer Bank von Quadratur-Spiegelfiltern, deren interessanteste Eigenschaft ist, dass sie die Überlappung der Spektren eines Subbandes ins andere verhindert. Ohne Quantisierer erlauben die Quadratur-Spiegelfilter, das Eingangssignal beinahe unverändert wiederzugewinnen.

Die Subband-Codierung bietet den Vorteil, dass die Anzahl der zur Quantisierung jedes Subbandes benötigten Bits der betreffenden Signalenergie angepasst werden kann. Berücksichtigt man die spektrale Dichte der mittleren Leistung eines Sprachsignals, wird die Anzahl Bit für das Subband der tiefen Frequenzen grösser als jene für ein hochfrequentes Subband. Die Aufteilung des Eingangssignals kann durch eine Filterbank mit Baumstruktur verfeinert werden. Dabei wird jedes Subband in zwei weitere aufgeteilt usw. Die Anwendung der Subband-Codierung bei 64 kbit/s führt zu einer ausgezeichneten Qualität, indem eine grössere als die Telefoniebandbreite berücksichtigt werden kann. Bei Bitraten von 16 kbit/s...24 kbit/s bleibt noch eine transparente Qualität erreichbar.

33 Andere Codierverfahren bei tiefen Bitraten

Wenn die verfügbare Informationsrate zwischen ungefähr 8 kbit/s und 16 kbit/s liegt, muss unbedingt auf aufwendigere Codierungsmethoden zurückgegriffen werden, um die gute Verständlichkeit der Sprache zu erhalten, obwohl hier einige merkbare Verzerrungen auftre-

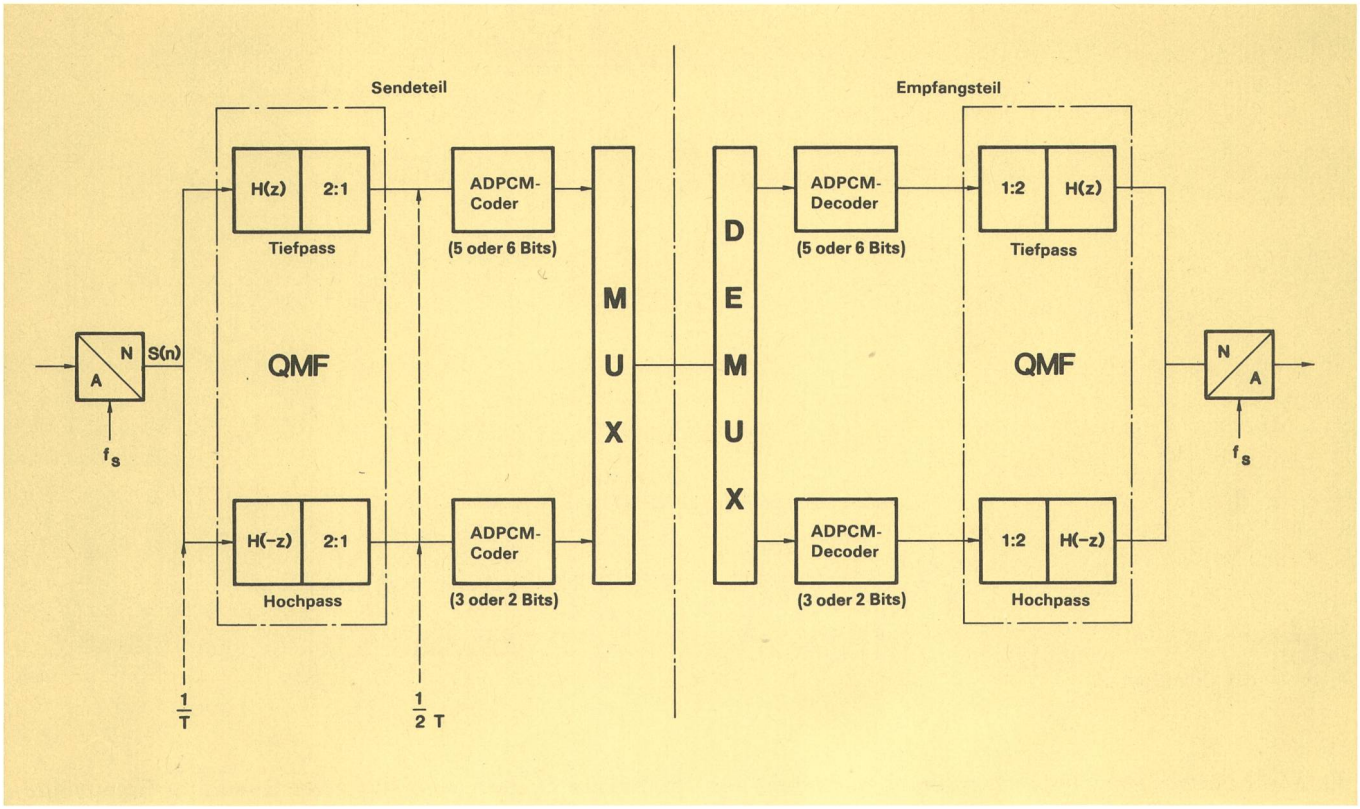

Fig. 4
Blockschaltbild der Codierung mit 2 Subbändern mit einer Struktur aus Quadratur-Spiegelfiltern

ten. Unter diesen Methoden sind nach der Subband-Codierung die bekanntesten:

- die adaptive Prädiktions-Codierung (APC)
- die adaptive Transformations-Codierung (ATC)
- der sprachgesteuerte Vocoder (Voice Excited Vocoder: VEV).

Wir werden hier nicht auf Einzelheiten dieser Coder eingehen. Sie sind Gegenstand vieler Arbeiten, die in [5] zitiert werden. Ihre Anwendung im Gebiet des Fernmeldewesens wird noch studiert, insbesondere die Codierung von Sprachsignalen bei 16 kbit/s für Systeme des Mobilfunks. Es seien lediglich die Voraussetzungen dieser drei Codierverfahren aufgeführt.

- Das angestrebte Ziel der *adaptiven Prädiktionscodierung* ist, die Redundanz des Signals merklich zu verringern, indem ihm das entzogen wird, was aufgrund eines Blocks vorausgegangener Abtastwerte vorausgesagt werden kann. Das solchermassen erhaltene Signal wird quantisiert und übertragen. Die Prädiktion nutzt einerseits die Korrelation zwischen aufeinanderfolgenden Abtastwerten aus, anderseits verwendet sie die fast periodische Eigenart von Stimmsignalen, um einen Teil ihres Spektrums zu unterdrücken. Der gesamte Prädiktionsgewinn beträgt ungefähr 17 dB. Mit anderen Worten: die adaptive Prädiktionscodierung ist ein Verfahren, mit dem man die Prädiktionsenergie zu minimieren sucht, was gleichbedeutend ist mit einer merklichen Verminderung des Quantisierungsgeräusches. Bei 16 kbit/s ermöglicht die APC-Codierung eine sehr gute Qualität in der Sub-Telefonieklasse. Bei 9,6 kbit/s ist die Sprachqualität noch durchaus annehmbar.
- In der *adaptiven Transformations-Codierung* wird das Signal abgetastet und in Blöcken von N Abtastwerten verarbeitet (N = 128...256). Jeder Block wird der diskreten Cosinus-Transformation unterworfen, und die voneinander unabhängig quantisierten Abtastwerte werden gleichzeitig mit den Quantisierungsparametern übertragen [5]. Dieses recht komplizierte Verfahren ist noch in voller Entwicklung und man kann eine gute Sub-Telefoniequalität bei 9,6 kbit/s erwarten.
- Der *sprachgesteuerte Vocoder* ist eine hybride Lösung zwischen Coder und Vocoder, die eine Informationsrate unterhalb 9,6 kbit/s erlaubt bei trotzdem guter Sub-Telefoniequalität. Wie bei der adaptiven Transformations-Codierung sind auch hier noch nicht alle Verbesserungsmöglichkeiten ausgeschöpft [5].

Es ist denkbar, dass in naher Zukunft das eine oder andere Verfahren eine praktische Anwendung in der Codierung mit tiefen Bitraten findet, welche zu den nächsten Studienprojekten im CCITT zählten.

34 Anwendung für die Codierung bei 32 kbit/s

Zahlreiche Fernmeldeunternehmen betreiben Studien über die adaptive Differenz-PCM-Codierung im Blick auf eine verbesserte Ausnutzung der digitalen Übertragungskanäle. Es ist schnell klar geworden, dass die Systeme mit adaptiven Parametern leistungsfähiger für die Sprachübertragung sind als jene mit festen Parametern. Daher hat das CCITT kürzlich eine neue Empfehlung über die adaptive Differenz-PCM-Codierung (ADPCM) bei 32 kbit/s veröffentlicht, welche die Vorschriften für die Umwandlung eines PCM-Kanals nach dem Bildungs-

gesetz A oder μ bei 64 kbit/s in einen 32-kbit/s-Kanal enthält.

Figur 5 zeigt das Blockschaltbild eines PCM-ADPCM-Transcoders. In diesem System befindet sich nach den PCM-Schnittstellen ein Demultiplexer, der die Sprachsignale von den Signalen in den Zeitschlitzen ZO und Z16 trennt (Rahmensynchronwort, Signalisierung, Alarme usw.). Die 2×30 Sprachsignale werden anschliessend auf einen digitalen Prozessor geführt, der die Umrechnung vom PCM-Code gemäss A-Gesetz mit 8 Bits in einen linearen Code zu 16 Bits ausführt. Darauf wird der Bitstrom mit einer Methode, die dem im CCITT festgelegten Algorithmus entspricht, umgewandelt. Die Signale der Zeitschlitze ZO und Z16 gelangen zu einem Multiplexer, der mit der Folge von ADPCM-Worten zu 4 bit einen 2048-kbit/s-Rahmen bildet, der zu den 2-Mbit/s-Leitungsausrüstungen passt. Für die ADPCM-PCM-Umwandlung werden die inversen Schritte ausgeführt [7]. Für die Verwaltungen, die eine Grundbitrate von 64 kbit/s für die Vermittlung gewählt haben, beschränkt sich die Anwendung der 32-kbit/s-Codierung auf die Satellitenübertragung oder auf das zeitweilige Verdoppeln der Kapazität einer überlasteten Übertragungsstrecke.

35 Anwendungen für die 64-kbit/s-Breitbandcodierung

Audio- und Videokonferenz bilden mittelfristig die wichtigsten Anwendungsfälle der Breitband-Sprachcodierung, aber gewisse Fernmeldeunternehmen denken auch an andere Einsatzgebiete, insbesondere an das lautsprechende Telefon in einem ISDN, an Kommentarkanäle oder sogar an Musik in Rundspruchnetzen. Die Arbeiten im CCITT sind ziemlich weit fortgeschritten, und deuten in Richtung einer adaptiven Differenzcodierung in Subbändern (ADPCM-SB) mit 5 Bit oder 6 Bit für den Quantisierer im unteren Subband und 3 Bit oder 2 Bit für jenen im oberen Subband. Die Abtastfrequenz ist 16 kHz. Die subjektiven Hörversuche haben ebenfalls die Überlegenheit einer adaptiven Quantisierung und Prädiktion aufgezeigt. Ein Systementwurf wird den Bedarf nach einem Hilfsdatenkanal von beispielsweise 8 kbit/s...16 kbit/s zu berücksichtigen haben, für die Übertragung von langsam bewegten Bildern oder für das Fernkopieren.

4 Sprachsynthese

Die Stimmsynthese beginnt mehr und mehr im Gebiet des Fernmeldewesens in Erscheinung zu treten. Der sprechende Mikroprozessor ist kein Wunschtraum mehr, sondern Wirklichkeit von morgen. Die starke Ausbreitung des digitalen Netzwerks (ISDN) bringt umgekehrt eine starke Wandlung der Informationsbedürfnisse aller Art mit sich. Die Nachfrage nach Auskünften wächst bereits stärker als die Zahl der Anschlüsse, was früher oder später vielfältige Probleme im Bereich der PTT-Auskunftsdienste zur Folge haben wird, weil die heute eingesetzten Mittel nicht beliebig erweitert werden können. Das Ziel ist deshalb die Begrenzung dieses Wachstums mit Einsatz automatischer Sprachsynthesesystemen, die in Echtzeit arbeiten und nicht nur eine verständliche, sondern soweit wie möglich eine «natürliche» Stimme erzeugen. Folgende Fernmeldedienste sollen in Zukunft teilweise oder sogar vollständig automatisiert werden:

- sprechendes Telefonbuch
- Weckdienst
- Taxierung: der Teilnehmer wählt eine Nummer, worauf die Maschine den Preis der entsprechenden Verbindung bekanntgibt
- Änderung der Telefonnummer
- Datenbank
- Automatisches Dokumentationszentrum
- Auskunftsdienst usw.

Unter den vielfältigen Anwendungsmöglichkeiten, die den PTT-Betrieben angeboten werden, stösst die automatische Stimmsynthese aufgrund von Textvorlagen auf

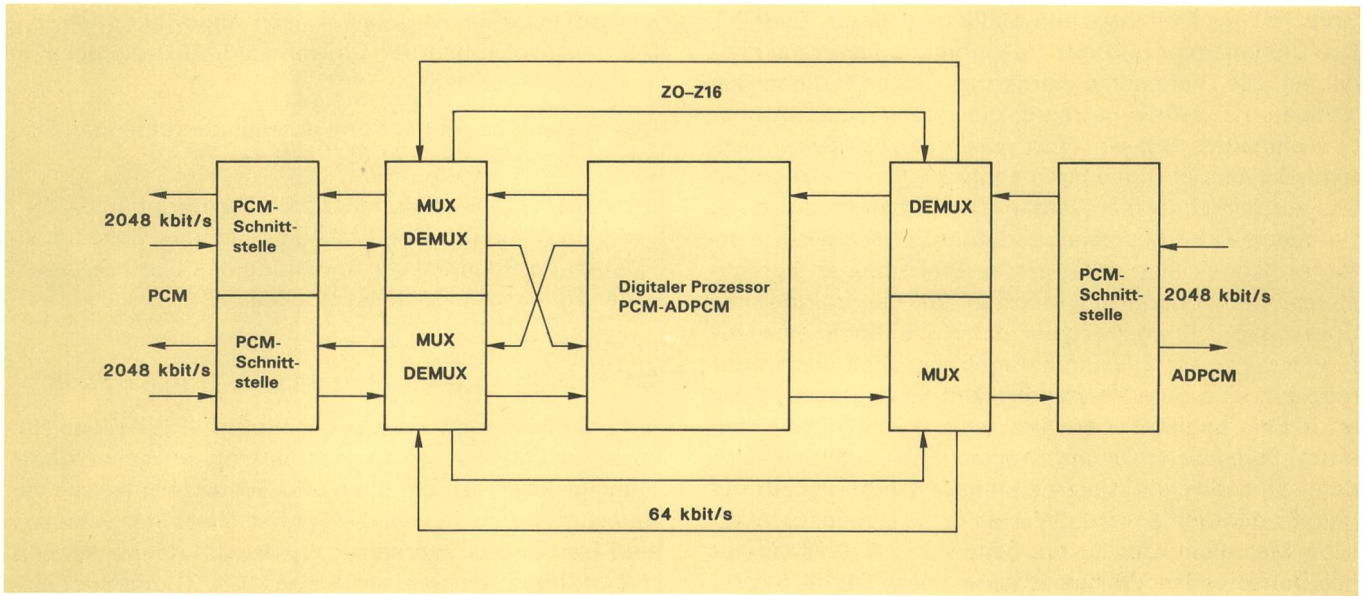

Fig. 5
Schema des Transcoders PCM-ADPCM mit 60 Kanälen

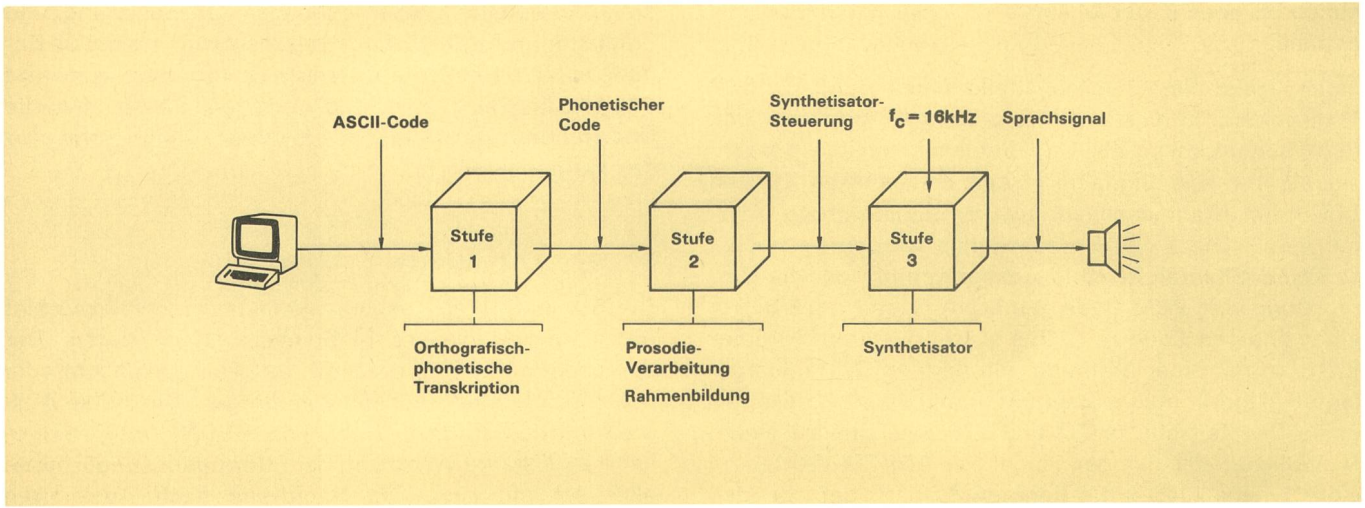

Fig. 6
Umwandlung Text–Sprache

wachsendes Interesse. Die Beweggründe sind stark genug, dass mehr und mehr Forschungsprojekte unternommen werden mit dem Ziel, automatische, sprechende Antwortausrüstungen für die Informations- und Auskunftsdienste mit Zugriff zum Telefonnetz zu entwikkeln.

41 Synthese mit phonetischen Elementen

Die künstliche Erzeugung eines Sprachsignals kann mit verschiedenen sog. parametrischen Verfahren durchgeführt werden. Diese codieren nicht mehr das Signal f(t) selbst, sondern die Parameter eines Erzeugungsmodells, das auf dem Prinzip des Produktes einer Anregungsfunktion und einer Übertragungsfunktion beruht.

Im allgemeinen verlangen die Anwendungen im Fernmeldewesen einen beachtlichen Wortschatz. Es ist daher günstiger, das Syntheseverfahren mit Diphonemen anzuwenden, das theoretisch eine bessere Qualität als andere Methoden erlaubt. In diesem Verfahren sind die Sprachbausteine, die gespeichert sind, Wortbruchstücke. Das Phonem ist die kleinste Lauteinheit einer Sprache. Im Französischen zählt man deren über 37. Das Diphonem bezeichnet die Verbindung aus zwei Phonemen. Die Diphoneme werden voneinander durch fast stationäre Zustände getrennt, die selbst keine Information enthalten. Um ein Wort wiederherzustellen, müssten lediglich die Diphoneme an ihren gemeinsamen Enden zusammengesetzt werden. Die Sprachsynthese mit Diphonemen ist bestechend in ihrer Einfachheit, sie erfordert jedoch eine recht grosse Bibliothek an Sprachelementen. Theoretisch müssten für die französische Sprache $(37)^2$ Kombinationen, d. h. 1369 Diphoneme verfügbar sein. Das Zusammenfügen von Diphonemen ermöglicht eine gute Verständlichkeit der Sprache, diese ist jedoch unglücklicherweise aller «Natürlichkeit» beraubt. Es ist daher unumgänglich, eine Betonung nach dem prosodischen Verarbeitungsverfahren hinzuzufügen. Es geht mit anderen Worten darum, zu jeder Wortsilbe Melodiemuster zu erarbeiten und den Rhythmus des betreffenden Wortes zu berechnen. Da im Silbenablauf Unterbrechungen auftreten können, müssen als Abhilfe gelegentlich prosodische Angaben verwendet werden. Diese feinen Melodieschwankungen werden häufig Verarbeitung der Mikromelodie genannt. In Verbindung mit diesen Verfahren benützt man als Synthetisator vorzugsweise die Kompression mit linearer Prädiktionscodierung, da dies heute die am besten an die Synthese bei niedrigen Bitraten angepasste Methode ist.

Offensichtlich muss die Bibliothek phonetischer Elemente der Art des Synthetisators angepasst sein. Diese Verbindung eines LPC-Verfahrens zur Codierung und Kompression mit der Synthese durch Diphoneme und der prosodischen Behandlung ergaben Lösungen mit ermutigenden Leistungsmerkmalen. Dies hat mehrere Laboratorien bewogen, diese technische Wahl auch bei der Verwirklichung verschiedener, spezifischer Projekte zur automatischen Sprachsynthese ab Text zu treffen [10].

42 Automatische Sprachsynthese ab Text

Bei der automatischen Sprachsynthese ab Text wird von einer Schriftzeichenfolge auf ein Sprachsignal übergegangen. Sie kann vorzugsweise im Gebiet der Auskunftsdienste im Fernmeldewesen angewandt werden. Das Blockschaltbild der Umwandlung Text–Sprache ist in *Figur 6* dargestellt.

Das System arbeitet auf drei aufeinanderfolgenden Stufen:

- die orthographisch-phonetische Transkription
- die Berechnung der Prosodie und das Zusammenfügen der Rahmen zur Steuerung des Synthetisators
- der eigentliche Sprachsynthetisator.

421 Orthographisch-phonetische Transkription

Der Text wird der Einheit auf Stufe 1 im ASCII-Code eingegeben und erscheint wieder als phonetischer Code, miteingeschlossen die Satz- und Syntaxzeichen, die der Bediener in den Text eingefügt hat. Eines der wirksamsten Transkriptionsverfahren wurde vom *Centre National d'Etudes des Télécommunications* (CNET) und der *Ecole Nationale Supérieure des Télécommunications* (ENST) entwickelt [11]. Es benutzt eine Anzahl von baumförmig

sich verzweigenden Tests, die auf den links und rechts benachbarten Text eines Analysefensters angewendet werden. Diese Transkriptions-Software wird ergänzt durch Unterprogramme zur Transkription von Zahlen und Satzzeichen in phonetische Zeichen und durch verschiedene Hilfsprogramme zur Steuerung des Testablaufs.

422 Berechnung der Prosodie

Die prosodische Verarbeitung geschieht auf Stufe 2 in Figur 6. Dabei wird die Kette von phonetischen Zeichen untersucht, d. h. es werden die Anzahl Silben eines Wortes und die Eigenheiten der Wortanfänge und -enden festgestellt, Rhythmus und Melodie berechnet und schliesslich die Rahmen aufgrund vorbestimmter Regeln erarbeitet.

423 Zusammenfügen der Rahmen

Das Zusammenfügen der Rahmen, die die Steuerung des Synthetisators bilden, erfolgt unmittelbar nach den vorhergehenden Verarbeitungsschritten. Schliesslich verwertet der Synthetisator die Steuerparameter, um die Sprache zu erzeugen. Im allgemeinen werden die Parameter, die den Signalwert bestimmen, alle 10 ms...25 ms berechnet.

43 Verfahren zur Sprachoptimierung

Anwendungen, die eine gewisse Sprachqualität verlangen und bei denen der Wortschatz praktisch unbegrenzt ist, erfordern manchmal eine Systemarchitektur, die Versuche zur Sprachoptimierung durchzuführen erlaubt. Zu diesem Zweck werden sehr oft die folgenden Optimierungsverfahren verwendet:

- Korrektur der Aussprache
- Korrektur des Rhythmus
- Veränderung der Stimmeigenschaften.

431 Korrektur der Aussprache

Im allgemeinen sprechen die leistungsfähigsten Systeme die Wörter eines orthographischen Textes richtig aus, aber es kann vorkommen, dass ein Wort oder ein Satzteil in ungenügender Art ausgesprochen wird. Dann kann die melodische Eigenschaft oder die Verständlichkeit des Wortes oder Satzteiles dadurch verbessert werden, dass anstelle der orthographischen Buchstabenfolge eine Kette von phonetischen Zeichen verwendet wird.

432 Korrektur des Rhythmus

Pausen im Satz werden durch Einfügen verschiedener Satzzeichen vorgenommen. Gleichermassen lässt sich der Rhythmus und die Betonung der gesprochenen Silbenfolge mit Satzzeichen und Syntaxregeln verbessern.

433 Änderung der Stimmeigenschaften

Im allgemeinen liest ein automatisches Sprachsynthesesystem einen Text mit einer mittleren Rate von 180 Wörtern in der Minute. Dies ist eine normale und den meisten Umständen und vorgesehenen Informationsarten gut angepasste Geschwindigkeit. Unter gewissen Bedingungen ist es jedoch wünschenswert, diese Rate auf Werte zwischen 120 und 300 Wörter zu erhöhen oder zu vermindern. Zu diesem Zweck erlaubt die Steuerungssoftware, die durch die Satzzeichen bestimmte Pausenlänge zu verändern und so die Sprechgeschwindigkeit und den Stimmklang zu verändern.

44 Das Projekt SAPOR

SAPOR ist die französische Abkürzung für automatische Sprachsynthese mit Computer (*S*ynthèse *a*utomatique de *p*arole par *or*dinateur). Es bezeichnet ein Entwicklungsprojekt zur automatischen Sprachsynthese nach Text. Es soll die Studien verwerten, die eine Forschungsgruppe an der *Eidgenössischen Technischen Hochschule Zürich* mit der Unterstützung und Mitarbeit der Forschungs- und Entwicklungsabteilung der PTT durchführt. Man verfolgt eine erste Versuchsanwendung der Synthese im Rahmen des Betriebs von Sprechmaschinen in einem internen Fernmeldenetz der PTT. Das angestrebte Ziel umfasst folgende beide Hauptaspekte:

- *Forschungsrichtung A*
  - Grobes Auswerten der Sprachqualität für die deutsche Sprache
  - Beurteilen der Betonungsstufen und der Betonungsunterschiede (Melodie)
  - Überwachen der Leistungsfähigkeit der Transkription vom orthographischen auf den phonetischen Code.
- *Forschungsrichtung B*
  - Ermitteln der hauptsächlichen praktischen Grenzen beim Einsatz von Sprachsynthesesystemen
  - Untersuchen der einzelnen Reaktionen und Verhaltensmuster von Benützern gegenüber einer «fast natürlichen» Stimme.

Figur 7 zeigt die vorläufige Anordnung und die Betriebsarten des Versuchssystems für Sprachsynthese nach Text. Das Kernstück des Systems bildet ein Text-Sprach-Umwandler, dessen Text entweder von einem Host-Rechner (VAX 750) im Technischen Zentrum der Abteilung Forschung und Entwicklung der PTT oder von einem Terminal im Redaktionsraum des Sprechmaschinen-Dienstes stammt. Im Endausbau wird die SAPOR-Einrichtung einen Einsatz in den folgenden drei Betriebsarten erlauben:

- In der Betriebsart *on-line* ist SAPOR mit einem Host-Computer verbunden. Diese Konstellation erlaubt, direkt auf die verschiedenen Transkriptions- und Synthesealgorithmen einzuwirken, um sie fortlaufend zu verbessern, was in einem solchen Projekt unerlässlich ist.
- Am Ende der Entwicklung soll SAPOR in der Betriebsart *off-line* arbeiten, d. h. als selbständiges Text-Sprache-Umwandlungssystem. Es wird in der Lage sein, die Befehle, abgegeben von einem Terminal, zu lesen und in gesprochene Sprache umzusetzen.
- In der Betriebsart *set-up* hat der Bediener die Möglichkeit, gewisse interne Parameter von SAPOR zu ändern, um die Text-Sprache-Transkription zu optimieren.

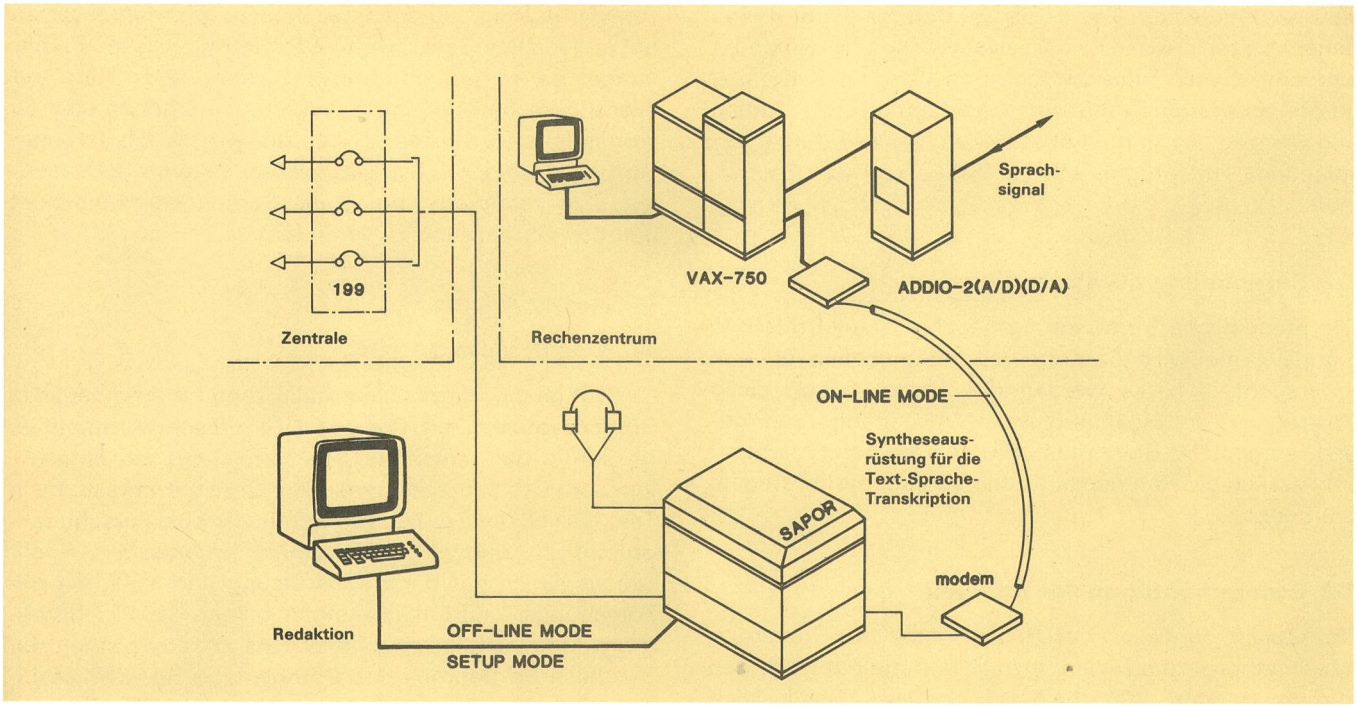

Fig. 7
Anwendungsart der Versuchseinrichtung für Sprachsynthese «SAPOR»

Das Projekt SAPOR wird einen wichtigen Beitrag an die Kenntnisse und die Beurteilung der Möglichkeiten und Grenzen der Sprachsynthese leisten. Zum voraus können folgende Vor- und Nachteile zusammengefasst werden:

- rasche Erneuerung der Informationen und Auskünfte
- Zugriff des Teilnehmers zur Quelle des Informationstextes
- merkliche Senkung der Betriebskosten der Auskunftsdienste der PTT
- grosser Spielraum bei der Einführung neuer Dienste
- Unsicher ist die Akzeptanz einer «fast natürlichen» Stimme durch die Öffentlichkeit. Dieser Punkt hängt weitgehend vom Fortschritt in den Synthesealgorithmen und vom Umfang der Bibliothek mit phonetischen Elementen ab.

5 Codierung von Musiksignalen

Die PCM-Multiplexiersysteme für Tonsignale hoher Güte arbeiten im allgemeinen mit drei verschiedenen Codierverfahren:

- Lineare Codierung mit 13 Bits...16 Bits
- Momentanwertkompander mit 11 Segmenten
- Quasi-Momentanwertkompander mit fünf Stufen.

Zahlreiche Experimente in Form von subjektiven Hörversuchen zeigten, dass eine maximale Auflösung von 14 Bit/Wort nötig ist, wenn für Tonsignale hoher Güte eine lineare Codierung angewendet wird. Eine Auflösung von 16 Bit wird heute bei der Compact-Disc und oft bei neuen Radiostudios verwendet. Die grosse Bandbreite, die dieses Verfahren erfordert, ist ein schwerer Nachteil bei der Übertragung. In diesem Fall greift man vorzugsweise auf eine Form der Kompression-Expansion zurück, die keine hörbare Verschlechterung bringt.

51 Prinzip der Codierung mit Momentanwertkompander mit 11 Segmenten

Beim Kompander-Codierverfahren wird das sog. A-Gesetz verwendet und durch 11 gerade Segmente und 11 Bit/Codewort angenähert. Dies entspricht einer minimalen Auflösung von 9 Bits einer linearen Codierung. Der Verlauf der logarithmischen Kompression ist in *Tabelle I* im Detail aufgeführt. Dieses Gesetz weist eine gebrochene Transferkennlinie auf, mit engen oder weiten Quantisierungsstufen, je nachdem, ob die Signale kleine oder grosse Amplituden aufweisen. Man erzielt so ein vermindertes Quantisierungssrauschen bei kleinen Signalen auf Kosten einer Vergrösserung des Rauschens bei grossen Signalen. Dank deren Maskierungswirkung wird die subjektive Güte des Signals praktisch nicht verändert. Auf diese Weise lässt sich die Anzahl Bits, die je Abtastwert zu übermitteln sind, vermindern. Das 11. (freie) Bit des Segments um den Nullpunkt kann zur Übertragung eines Hilfsdatenkanals dienen, dessen Bitrate von der Amplitudenverteilung des Hörfunksignals abhängt. Einzig während Programmpausen erreicht die Bitrate ihren Höchstwert von 32 kbit/s. Ausserdem schützt ein Paritätsbit die fünf höchstwertigen Bits jedes Abtastwertes gegen Fehler. Die Gesamtbitrate eines Tonkanals wird damit 384 kbit/s und entspricht genau jener von sechs Telefoniekanälen zu 64 kbit/s, wie sie das CCITT definiert hat. Das Gesetz ist im einzelnen in [8] beschrieben.

52 Prinzip der Codierung mit einem Quasi-Momentanwertkompander zu fünf Segmenten

Das Prinzip der Quasi-Momentanwertkompander beruht auf folgender Annahme [9]: Wenn man die binären Zeichen von aufeinanderfolgenden digitalen Abtastwerten

eines Tonsignals betrachtet, bemerkt man, dass die höherwertigen Bits mehr oder weniger langsam ändern, je nach der Eigenheit des Signals. Unter dieser Bedingung schien es vorteilhaft, diese Abtastwerte periodisch in Blöcke zu gruppieren, worin die Bits, die sich kaum oder gar nicht ändern, nur einmal übertragen werden. Das Verfahren besteht somit darin, das Signal in regelmässigen Abständen zu beobachten und jedesmal das Quantisierungsgesetz dem Amplitudenverhalten des Blocks von Abtastwerten anzupassen.

Figur 8 beschreibt das Gesetz der Quasi-Momentanwert-Kompression mit fünf Segmenten genauer. Die

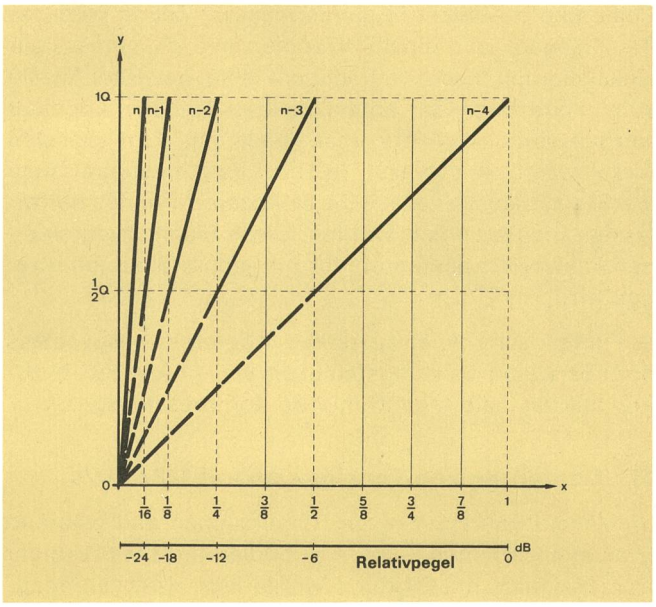

Fig. 8
Charakteristik des Quasi-Momentanwert-Kompanders mit fünf Skalen
Q Gesamtzahl der Quantisierungsstufen im positiven Quadranten
x Normierte Amplitude des Eingangssignals
y Quantisierungspegel
(n-1....4) Auflösung pro Skala

Tabelle I. Codiergesetz A mit 11 Segmenten.

Vorzeichen	Quantisierung		Segment Nr.	Anzahl Bits	Codierung mit 11 Bits										
	Linear	Nicht linear			1	2	3	4	5	6	7	8	9	10	11
Positiv	8192 4097	896 769	1	9	1 1	1 1	1 1	1 1	1 0	1 0	1 0	1 0	1 0	1 0	1 0
	4096 2049	768 641	2	10	1 1	1 1	1 1	0 0	1 0	1 0	1 0	1 0	1 0	1 0	1 0
	2048 1025	640 513	3	11	1 1	1 1	0 0	1 1	1 0	1 0	1 0	1 0	1 0	1 0	1 0
	1024 513	512 385	4	12	1 1	1 1	0 0	0 0	1 0	1 0	1 0	1 0	1 0	1 0	1 0
	512 257	384 257	5	13	1 1	0 0	1 1	1 1	1 0	1 0	1 0	1 0	1 0	1 0	1 0
	256 129	256 129	6	14	1 1	0 0	1 1	0 0	1 0	1 0	1 0	1 0	1 0	1 0	1 0
	128 65	128 65			1 1	0 0	0 0	1 1	1 0	1 0	1 0	1 0	1 0	1 0	× ×
	64 1	64 1			1 1	0 0	0 0	0 0	1 0	1 0	1 0	1 0	1 0	1 0	× ×
Negativ	1 64	1 64			0 0	0 0	0 0	0 0	0 1	0 1	0 1	0 1	0 1	0 1	× ×
	65 128	65 128			0 0	0 0	0 0	1 1	0 1	0 1	0 1	0 1	0 1	0 1	× ×
	129 256	129 256			0 0	0 0	1 1	0 0	0 1	0 1	0 1	0 1	0 1	0 1	0 1
	257 512	257 384	7	13	0 0	0 0	1 1	1 1	0 1	0 1	0 1	0 1	0 1	0 1	0 1
	513 1024	385 512	8	12	0 0	1 1	0 0	0 0	0 1	0 1	0 1	0 1	0 1	0 1	0 1
	1025 2048	513 640	9	11	0 0	1 1	0 0	1 1	0 1	0 1	0 1	0 1	0 1	0 1	0 1
	2049 4096	641 768	10	10	0 0	1 1	1 1	0 0	0 1	0 1	0 1	0 1	0 1	0 1	0 1
	4097 8192	769 896	11	9	0 0	1 1	1 1	1 1	0 1	0 1	0 1	0 1	0 1	0 1	0 1

Die 11. Bits, mit einem x bezeichnet, können zur Übertragung eines Datensignals dienen.

Codierung geschieht folgendermassen: Zuerst wird das Analogsignal mit einem Wandler mit gleichmässiger Quantisierung codiert. Nachdem eine gewisse Anzahl aufeinanderfolgender Abtastwerte im Speicher abgelegt worden sind, betrachtet man jenen mit dem grössten Absolutwert. Aus dieser Beobachtung bestimmt man die Skala, nach der alle anderen Abtastwerte des betreffenden Blocks codiert werden. Diese Skala ist durch einen Faktorwert bestimmt, der mit jedem Block übertragen wird.

Die empirische Analyse dieses Kompressionsgesetzes offenbart eine Leistungsfähigkeit von (n−1) Bit, bezogen auf das halb-logarithmische Kompressionsgesetz.

53 Codierung von Tonsignalen mit 128 kbit/s

Im Verlauf von subjektiven Hörversuchen und während Evaluationsarbeiten über die Codierung breitbandiger Sprachsignale mit 64 kbit/s wurde klar, dass die adaptive Differenz-PCM-Codierung mit Subbändern eine bessere Qualität als die PCM-Codierung mit 128 kbit/s liefert. In diesem Zusammenhang scheint es interessant, die Möglichkeiten einer ADPCM-Codierung mit Subbändern bei 128 kbit/s kennenzulernen, wie sie für Tonfrequenzsignale angewendet werden könnte, aber auch unter dem Gesichtspunkt der möglichen Ablösung des heutigen Hochfrequenz-Telefonrundspruchs durch eine digitale Lösung. Eine Codierung bei 128 kbit/s böte eine mit ISDN vereinbare Anwendung. Die Schweizer PTT sind an dieser Frage interessiert und studieren die folgenden Gesichtspunkte:

- Optimierung des ADPCM-SB-Codierungsalgorithmus von breitbandigen Sprachsignalen bei 64 kbit/s für Hörfunksignale mit 7 kHz Bandbreite
- ADPCM-SB-Codierung mit 128 kbit/s und 7 kHz Bandbreite. Die Quantisierung ist genügend fein, so dass die Qualität jener der Systeme bei 192 kbit/s nahekommt, die mit einem Momentanwertkompander zu 11 Segmenten oder einem Quasi-Momentanwertkompander mit fünf Skalen arbeiten.
- ADPCM-SB-Codierung mit 128 kHz und 15 kHz Bandbreite. Diese Variante bietet zweifellos am meisten Schwierigkeiten in bezug auf Vergrösserung des Quantisierungsrauschens und der Stufigkeit. Diese Verschlechterung kann teilweise durch Zufügen eines zusammengesetzten Schwingungssignals (dither) verdeckt werden, das aus der Überlagerung von weissem Rauschen und einem Rechtecksignal mit der Frequenz, entsprechend der halben Abtastfrequenz, gebildet wird. Andere Methoden werden untersucht, beispielsweise jene, bei der das Restspektrum aus der Verschiebung eines Subbandes verwendet wird, um den subjektiven Höreindruck von höherfrequenten Signalen zu verbessern.
- ADPCM-SB-Codierung mit einer Bandbreite, die den besten subjektiven Kompromiss zwischen Quantisierungsrauschen und Durchlassbereich darstellt. Man kann beispielsweise eine Auflösung von 10 Bit und eine Abtastfrequenz von 25,6 kHz anstreben. Dies entspräche der heutigen Bandbreite des Hochfrequenz-Telefonrundspruchs (12 kHz). Diese Lösung brächte allerdings gewisse Nachteile bezüglich Synchronisation und Kompatibilität mit der Hierarchie der Digitalsysteme.

6 Beurteilung von Codierungsgesetzen mittels Computer-Programmen

Bei der Beurteilung von Codierungsgesetzen wird ständig die Beziehung ergründet zwischen den gemessenen Werten und dem durch Unvollkommenheiten der Codieralgorithmen oder durch Übertragungsstörungen hervorgerufenen subjektiven Eindruck. Diese Bewertungstätigkeit zur Optimierung der Codiereigenschaften bringt unzählige Parameteränderungen mit sich, was nur mit Hilfe des Computers vernünftig durchführbar ist. Zu diesem Zweck benützt man stets dieselbe Einrichtung, deren Verwendung programmgesteuert ist. So verliert man nicht jedesmal Zeit mit Entwerfen und Herstellen von verschiedenen Aufbauten. Heutzutage finden sich eine ganze Reihe von schnellen und besonders für die digitale Signalverarbeitung geeignete Prozessoren, deren Leistungsfähigkeit hinreicht, um sogar die kompliziertesten Algorithmen in Echtzeit zu bewältigen. Allerdings ist in den ersten Phasen von Studien und Forschungen eine Echtzeitsimulation nicht wesentlich, da die subjektiven Hörversuche nur mit Sprachmustern oder Tonfolgen sehr kurzer Dauer (20 s...30 s) durchgeführt werden. Ein genügend grosser Massenspeicher zugunsten eines sehr einfachen und vielseitigen Simulationsverfahrens sind die Voraussetzungen, um eine einigermassen wirksame Entwicklung zu betreiben. Offensichtlich benötigt man anschliessend schnelle Prozessoren, die den entwickelten Algorithmus in Echtzeit verarbeiten können, um die Hardware des Endproduktes zu optimieren. Um die Forschungsbedürfnisse im Bereich der Codierung zu befriedigen, haben die schweizerischen PTT eine Einrichtung entwickelt, die es ermöglicht, Codier- und Übertragungssysteme von Sprach- und Musiksignalen mit einem Grossrechner und einem hochauflösenden A/D-D/A-Wandler zu simulieren und zu studieren.

61 Hardware von ADDIO-2

ADDIO-2 hat drei Hauptteile, wie im Blockschaltbild der *Figur 9* dargestellt. Sie erfüllen die folgenden Aufgaben:

- *Analogteil:* Er umfasst die Aufnahme- und Wiedergabegeräte für die Signale (Mikrofone, Verstärker, digitale Bandgeräte, Synthetisator, Lautsprecher usw.). Diese sind in geeigneten Akustikräumen untergebracht, wie in Abschnitt 75 beschrieben wird.
- *A/D und D/A-Wandler:* Dieser Teil enthält alle Schaltungen zur Analog-Digital- bzw. Digital-Analog-Wandlung, zur Filterung und Signalformung sowie die Ausrüstungen der digitalen und analogen Schnittstellen. Der eigentliche Wandler ist ein DSC-200 der *Digital Sound Corporation.* Seine Auflösung beträgt 16 Bit, was einem theoretischen Aussteuerbereich von 96 dB entspricht. Die maximale Abtastfrequenz ist 50 kHz. Damit können irgendwelche Signale in einem Frequenzband von über 20 kHz, programmierbar mit einer Filterbank von Bandpässen, verarbeitet werden. Bis zu acht Kanäle lassen sich gleichzeitig behandeln.
- *Prozessorteil:* Er umfasst einen VAX11-750-Rechner und einen Massenspeicher RA81 (456 MB). Damit kann in Echtzeit eine Musiksequenz hoher Qualität in Stereo, abgetastet mit 50 kHz, von über 30 Minuten Dauer aufgenommen und wiedergegeben werden. Zu

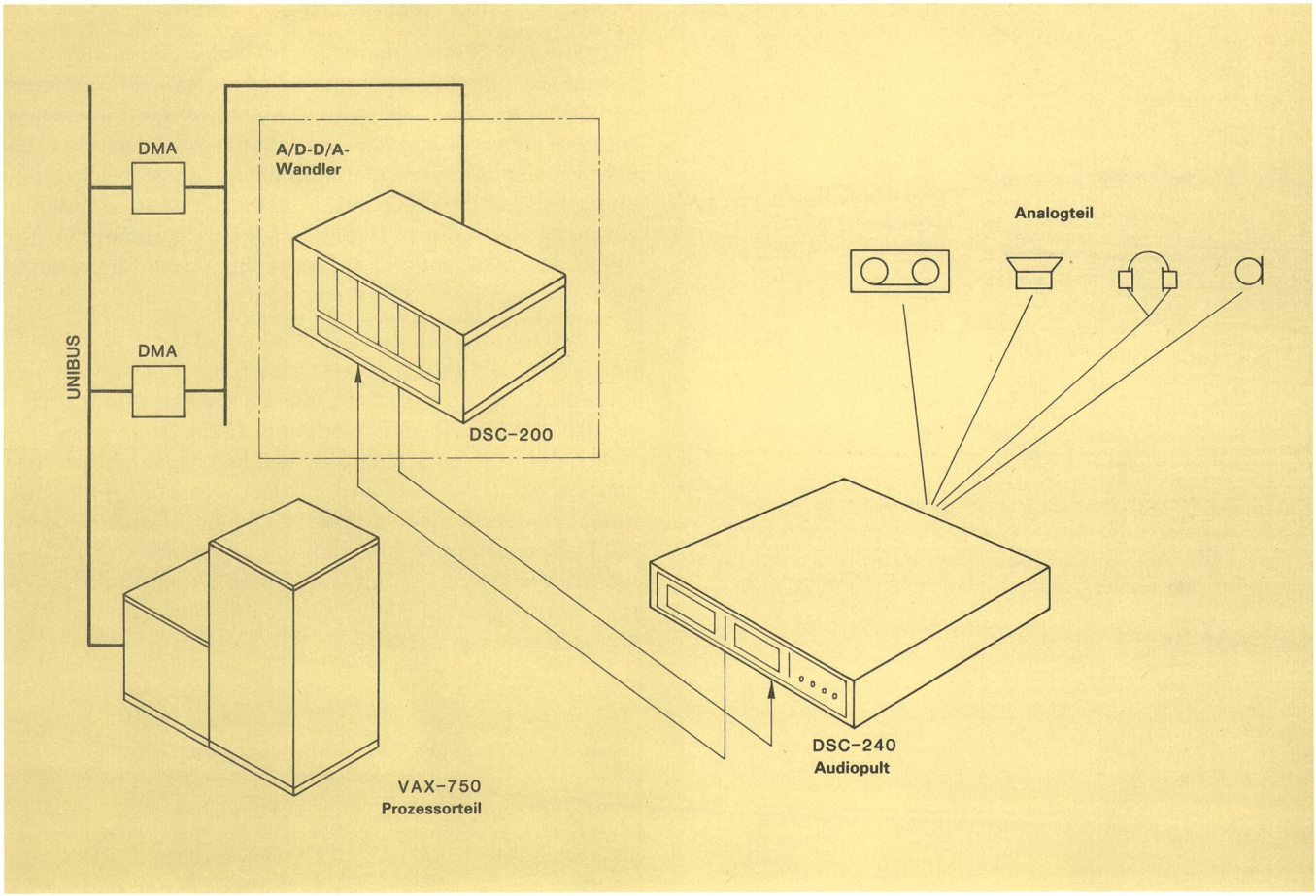

Fig. 9
Blockschaltbild von ADDIO-2

diesem Teil gehören natürlich alle für die Softwareentwicklung nötigen Peripheriegeräte.

62 Simulationsverfahren

Der Systemaufbau erlaubt die Verarbeitung von Sprache oder Musik nach dem vereinfachten Verfahren gemäss *Figur 10*. Das Analogsignal wird, nach der Vorformung (Pegelregelung, Filterung, Preemphasis usw.), mit gleichmässigen Quantisierungsschritten codiert. Damit entsteht eine Folge von 16-Bit-Wörtern. Anschliessend wird das Digitalsignal auf dem Plattenspeicher abgelegt. Die umgekehrte Operation erfolgt ähnlich. Die Datenerfassung und -ausgabe läuft immer in Echtzeit ab, aber nicht gleichzeitig (Halbduplex). Die folgende Beschreibung liefert ein klassisches Anwendungsbeispiel im Bereich der Codierung und digitalen Übertragung von Hörfunkprogrammen.

Problemstellung:

Im Fall einer Multiplexierung von zwei digitalen Audiokanälen in plesiochroner Betriebsart wird der Taktunterschied mit einem sog. Negativ-, Null- oder Positiv-Stopfverfahren aufgefangen. Dazu werden in einem Digitalsignal Stopfinformation sowie ein Synchronisationswort durch das Ersetzen gewisser Bits eingefügt. Bei der Wiedergabe muss die subjektiv wahrnehmbare Verschlechterung durch die zwangsweise Umwandlung von Informations-Bits des ursprünglichen Signals bewertet werden.

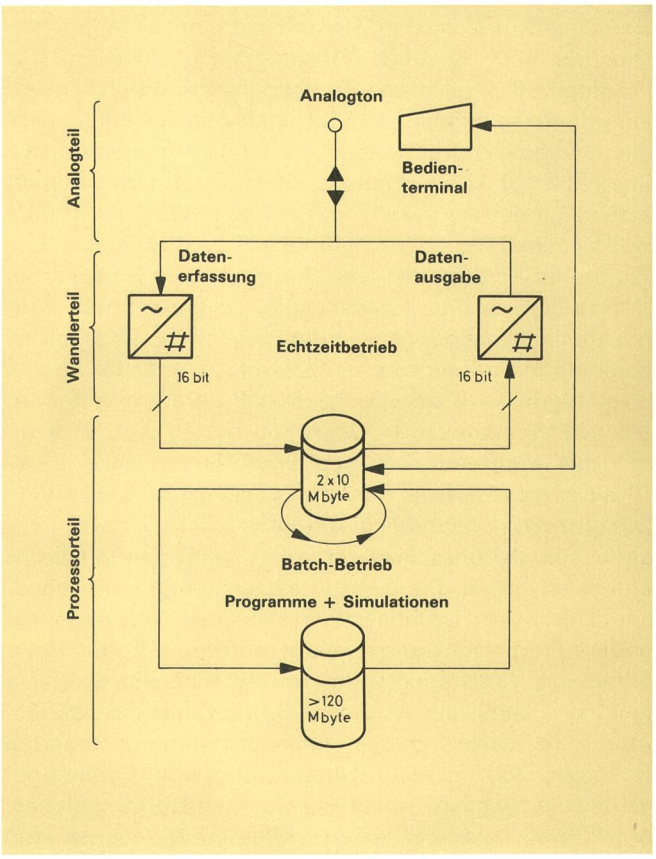

Fig. 10
Prinzipieller Laufweg der digitalen Signale

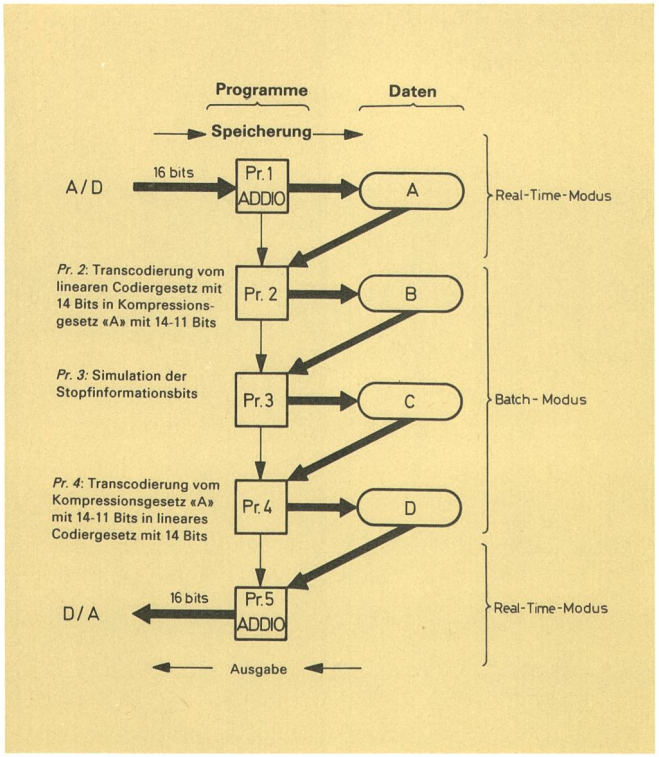

Fig. 11
Software-Simulation

Figur 11 zeigt den Aufbau der Simulationsprogramme durch die Software. Zuerst durchläuft das AF-Signal gewisse Regelungsstufen. Anschliessend wird es in ein digitales Signal mit dem Analog/Digital-Wandler zu 16 Bit mit einer Abtastfrequenz von 32 kHz codiert. Das so gewonnene Signal wird in Echtzeit auf einem Plattenspeicher der VAX abgelegt (Programm Pr.1 ADDIO). Das Programm Pr.2 führt eine Transcodierung durch: Aus einem digitalen Signal zu 16 Bit wird, gemäss einem Momentanwertkompressionsgesetz zu 11 Segmenten, ein Signal zu nur 11 Bit erhalten. Diese Operation simuliert einen digitalen Audiokanal mit 384 kbit/s. Das Programm Pr.3 soll die Simulationen mit dem Digitalsignal aus dem Programm Pr.2 durchführen. Im gewählten Beispiel bewerkstelligt es das Auswechseln, zu bestimmten Zeitpunkten, eines niedrigstwertigen Signalbits (LSB) mit einem andern, das zu einem Codewort gehört, das angibt, ob gestopft wird oder nicht. Das Programm Pr.4 führt die zu Pr.2 umgekehrte Operation durch, d. h. es wandelt das komprimierte Signal in ein linear codiertes zu 16 Bit zurück. Dieses wird Pr.5 und anschliessend dem Digital/Analog-Wandler zugeführt.
Jeder Simulationsschritt verändert somit den Inhalt des Eingangssignals, das dann in seiner neuen Form abgespeichert wird. Grundsätzlich kann das Verfahren auf andere Programme angewendet werden, z. B. die Simulation von Übertragungsfehlern. Mit Hilfe der Software lässt sich somit ein wirklicher Betriebszustand nachahmen. Die Verwendung eines programmgesteuerten Rechners zur genauen Übersetzung von Codieralgorithmen ermöglicht, diese miteinander zu vergleichen, unter Ausschluss der Fehlerquellen infolge verschiedener Hardware-Strukturen. Die Bewertung erfasst nur die Wahl der Software.

7 Subjektive Hörversuche

Bei gewissen Phänomenen in der Entwicklung der Verfahren zur digitalen Signalkompression verfügt man nicht immer über bekannte Massstäbe zur Bewertung der wirklichen Qualität eines Übertragungssystems. Es erweist sich zumeist unumgänglich, sich auf Hörversuche mit gezwungenermassen statistischer Aussagekraft zu stützen, um die mit gewöhnlichen Instrumenten gemachten Messungen zu ergänzen und gesicherte Schlüsse aus dem Ganzen zu ziehen.

Dazu sollen die subjektiven Hörversuche mit Qualitäts- und Verschlechterungsskalen das Mass der subjektiven Beeinträchtigung ermitteln, die durch Fehler oder Störungen im Lauf verschiedener Codierungs-Decodierungs-Schritte entsteht, die das Nutzsignal beeinflussen. Je nach Eigenart und Ziel der Versuche wird man die geeignetste Auswerteskala wählen. *Tabelle II* zeigt die beiden gängigsten Skalen mit fünf Punkten zur Auswertung der eigentlichen Güte eines Systems einerseits und zum Bewerten der subjektiven Wirkung einer Verschlechterung anderseits.

Tabelle II. Qualitäts- und Verschlechterungsskala.

Qualitätsskala	Verschlechterungsskala
5 Ausgezeichnet	5 Nicht wahrnehmbar
4 Gut	4 Wahrnehmbar, aber nicht störend
3 Ziemlich gut	3 Leicht störend
2 Mittelmässig	2 Störend
1 Schlecht	1 Stark störend

Für gewisse Arten von subjektiven Versuchen kann es günstiger sein, eine sogenannte Vergleichsskala anzuwenden. Sie umfasst die folgenden sieben Noten und Einschätzungen:

3 viel besser
2 besser
1 etwas besser
0 gleich
−1 etwas schlechter
−2 schlechter
−3 viel schlechter

Die Versuche müssen so angelegt und durchgeführt werden, dass sie ein Maximum an Aussagen über gewisse Formen von Unvollkommenheiten geben, die den Sprach- und Musik-Codiersystemen eigen sind.

Aus den am häufigsten eingesetzten Verfahren von subjektiven Hörversuchen für die Sprache sei die Methode der paarweisen Vergleiche (paired-comparison procedure) von zwei Sprachmustern A und B sowie die klassische Methode erwähnt, die ein mittleres Urteil oder eine mittlere Note (mean opinion score) verwendet, die durch eine Qualitätsskala mit fünf Noten ausgedrückt wird. Diese zwei Verfahren wurden erfolgreich von den Laboratorien eingesetzt, die im Rahmen der Arbeiten des CCITT an der Festlegung des besten ADPCM-Algorithmus für Breitband-Sprachcodierung mit 64 kbit/s teilgenommen haben. Zu diesem Zweck wurden die folgenden Methoden angewandt:

71 Methode der paarweisen Vergleiche

Bei dieser Methode werden zwei Sprachmuster A und B, die vom selben Sprecher stammen, aber von zwei verschiedenen Codiersystemen digital verarbeitet wurden, miteinander verglichen. Der zur Beurteilung aufgeforderte Hörer gibt an, ob er das Muster A oder das Muster B bevorzugt. Im allgemeinen wird dieses Verfahren für empfindlicher gehalten als die pauschalen Bewertungsmethoden. Es hat jedoch den Nachteil, rasch zu einer recht grossen Zahl von Antworten zu führen.

In der Praxis kann es günstiger sein, zwar beide erwähnten Verfahren einzusetzen, aber bei der Methode der paarweisen Vergleiche die Anzahl der zu prüfenden Kombinationen zu vermindern. Die wichtigsten Schritte zum eigentlichen subjektiven Test sind:

- Vorbereiten der Tonquellen mit verschiedenen Sprechern unter wohldefinierten Aufnahmebedingungen
- Die digital gespeicherten ursprünglichen Signale werden anschliessend durch die zu bewertenden Systeme verarbeitet (Codierung–Decodierung). Die so entstandenen Ausgangssignale werden noch in eine passende Form gebracht und bilden den endgültigen Inhalt der Tonbänder für die subjektiven Versuche
- Durchführen der Sitzungen mit subjektiven Hörversuchen.

Beispielsweise waren 32 Hörer mit über 10 000 Antworten nötig zur statistischen Beurteilung von vier ADPCM-Codecs bei 64 kbit/s unter fünf verschiedenen Bedingungen bezüglich Bitrate, Bitfehlerquote und Transcodierung.

72 Methode der gemittelten Urteile

In diesem Fall wird der Hörer aufgefordert, nacheinander eine Gruppe von Sprachmustern – bestehend aus zwei kurzen Sätzen – zu bewerten. Diese besitzen keinen Zusammenhang und werden von zwei Sprechern vorgetragen. Wenn sich der Versuch zum Beispiel auf vier Codecs bezieht, verwendet man vernünftigerweise 480 verschiedene Sätze, die 240 digital verarbeiteten Sprachmustern entsprechen, um eine genügende Vertrauenszahl zu erhalten. Bei diesem Beispiel werden die Sprachmuster gemäss den Versuchsbedingungen in acht Abschnitte zu je 30 Mustern eingeteilt, die anschliessend den Hörern in einer Folge gemäss dem «lateinischen Quadrat» (8×8) vorgeführt werden.

Mit insgesamt 16 Hörern können 3840 Antworten gewonnen werden. Manchmal ist die Einfügung einiger Takte Musik zum Trennen von zwei Codecs mit ähnlich guten Sprachsignaleigenschaften angezeigt.

73 Herstellung von Digitalaufnahmen für die subjektiven Hörversuche

Die kluge Wahl der Tonquellen, die Qualität der Originalaufnahme, die Hörbedingungen, die Aussteuerungsregelung und die Versuchsteilnehmer sind ausserordentlich wichtig für den Vergleich der Ergebnisse mit jenen anderer Laboratorien. Es ist deshalb erforderlich, alle zu beachtenden Parameter genau zu umschreiben und die Versuchsbedingungen ausführlich darzulegen.

731 Musiksignale

Die Radioprogrammausschnitte müssen im Blick auf ihr sehr kritisches Verhalten besonders ausgewählt werden bezüglich der besonderen Fehler, die das zu testende System bringt. Eine aussagekräftige Wahl von typischen Programmteilen hängt sehr oft von der Art der untersuchten Verschlechterung ab. Die Erfahrung zeigt, dass folgende Tonquellen besonders empfindlich auf die digitale Signalkompression sind:

- Alle Tonsignale mit ausgesprochener Transparenz und sehr wenig Obertönen sind im allgemeinen sehr empfindlich auf subjektive Versuche bezüglich Quantisierungsrauschen und Granularität. Das Klavier und der elektronische Bass gehören zu den kritischsten Instrumenten
- Zur Auswertung der Störfestigkeit eines digitalen Hörfunk-Übertragungssystems gegenüber Bitfehlern sind als Tonquellen Trompetenstösse oder Klänge des Glockenspiels angebracht. Sie enthalten viele Obertöne
- Es wurde beobachtet, dass bei Versuchen mit beispielsweise dem Becken das Einfügen der Präemphase/Deemphase unannehmbare Verschlechterungen bewirken kann.

732 Sprachsignale

Gemäss der genauen Zielvorgabe, die für die Bewertung eines Sprachcodierverfahrens gesetzt wurde, werden verschiedene Formen und Strukturen der Rede in verschiedenen Sprachen verwendet. Wenn die Grenze der Verständlichkeit bestimmt werden soll, empfiehlt sich oft die Anwendung von Prosa: Sie ist eine Ausdrucksweise, die nicht dem Versmass und den Reimregeln unterworfen ist wie die Lyrik.

Verschiedenartige Verzerrungen und ihre Auswirkungen können besser mit dem Test von Reimen erfasst werden. Zu diesem Zweck schlägt *Fairbank* eine Folge von Silben vor, die sich nur in einem Element der festen Zusammensetzung Mitlaut–Selbstlaut–Mitlaut (CVC) unterscheiden. Der Fall C_{var} VC lautet für die deutsche Sprache zum Beispiel:

- Bach, Fach, wach, Dach, Sack, Mach usw.

Für vergleichende Versuche empfiehlt man den Gebrauch von sog. phonetisch ausgewogenen Sätzen. *P. Combesure* schlägt eine Sammlung von zwanzig Listen zu zehn Sätzen vor, worin die relative Häufigkeit jedes Phonems jener im Französischen entspricht [13]. Zum Beispiel:

- Il se garantira du froid avec ce bon capuchon
- Annie s'ennuie loin de mes parents
- Les deux camions se sont heurtés de face

usw.

Von *J. Sotscheck* gibt es Studien über die phonologische Struktur der deutschen Sprache. Der sog. Berliner Test ist der bekannteste. Er umfasst 12 Satzgruppen zu acht Sätzen, welche die Häufigkeitsverteilung der Phoneme in der deutschen Sprache gut widerspiegeln. Zum Beispiel:

- Heute ist schönes Frühlingswetter
- Die Sonne lacht
- Am blauen Himmel ziehen die Wolken

usw.

Für die englische Sprache sei die Liste «1965 Revised List of Phonetically Balanced Sentences» (*Harvard* sentences) erwähnt. Diese Listen werden im allgemeinen für Qualitätstests, vergleichende Beurteilungen oder Beurteilungen nach Verschlechterungsklassen angewendet.

In der Praxis müssen die Sprachaufnahmen mit männlichen und weiblichen Sprechern durchgeführt werden, da sie merklich unterschiedliche Stimmklänge aufweisen.

733 Die digitale Aufzeichnung

Im Gebiet der Breitband-Sprachcodierung wie auch des Hochqualitättons muss die digitale Aufzeichnung für die Durchführung der subjektiven Auswerteversuche eingesetzt werden. Die damit erreichbare Qualität übertrifft bei weitem jene von Analogsystemen. Die Vorteile dieses neuen Verfahrens liegen in der Verbesserung der folgenden Qualitätsparameter:

- Das Signal/Geräusch-Verhältnis ist besser als 85 dB
- Die Geschwindigkeitsschwankungen und der Echoeffekt sind völlig beseitigt
- Die harmonische Verzerrung und die Intermodulation werden nur noch durch die Codier- und Decodier-Schaltungen erzeugt.

Die interessantesten Merkmale zeigt das digitale Bandgerät jedoch beim Kopieren. Beim Schneiden von Tonsequenzen oder beim Simulieren von mehreren hintereinandergeschalteten Codecs sind Kopien manchmal unumgänglich, ohne dafür eine analoge Zwischenstufe durchlaufen und neue Verarbeitungsschritte ausführen zu müssen. Dies bedingt, dass die gleiche Abtastfrequenz für beide Geräte verwendet wird und keinerlei Korrekturen in der Verstärkung, der Klangregelung usw. ausgeführt werden.

Unter diesen Bedingungen erfährt das Signal keine Einbussen, soweit die Bitfehler-Korrekturverfahren, wie sie in allen digitalen Bandgeräten eingesetzt werden, richtig arbeiten.

Tabelle III zeigt die wichtigsten Merkmale eines digitalen Bandgerätes. Was das Mischpult betrifft, scheint es kurzfristig nicht von der Digitaltechnik betroffen zu sein, da die analoge Verstärkungsregelung vor der eigentlichen Codierung immer nötig sein wird.

Tabelle III. Haupteigenschaften eines digitalen Bandgerätes.

PCM	2 Kanäle
Abtastfrequenz	48, 50,4 kHz
Auflösung	16 Bit linear
Frequenzband	20 Hz...20 kHz
Aussteuerbereich	90 dB
Klirrfaktor	0,05 %
Nebensprechen	80 dB
Eingangspegel	+ 6 dBm
Ausgangspegel	+ 6 dBm
Bandgeschwindigkeit	38 cm/s
Breite des Bandes	6,3 mm

74 Rechnerverarbeitung der Aufnahmen

Für Sprachaufnahme ist es sehr schwierig, ja unmöglich, bei der ersten Aufnahme ein Originalband herzustellen, das ohne Korrekturen, Verbesserungen oder Umgruppierungen für die Sitzungen der subjektiven Hörversuche verwendet werden kann. Im allgemeinen erweisen sich bei der Vorbereitung der Versuchsbänder folgende drei Operationen als unumgänglich:

- Regeln des Effektivwerts der Sprache
- Ausmerzen gewisser Fehler bei der Aufnahmesitzung
- Ordnen der Sprachmuster in einer genau definierten Reihenfolge

Alle diese Operationen können dank der praktisch unbegrenzten Möglichkeit der Software zur digitalen Signalverarbeitung mit dem Computer durchgeführt werden. In dieser Hinsicht ist das System ADDIO-2 besonders gut geeignet.

741 Messung und Regelung des Leistungspegels der Sprache

Der Begriff des elektrischen Pegels eines sich ändernden Signals ist immer teilweise mehrdeutig, da ebenso viele verschiedene Ergebnisse wie Integratorzeitkonstanten vorhanden sind. In der Praxis werden zwei Sorten Messinstrumente für Schallpegel verwendet. Die erste Sorte umfasst die Geräte, die das subjektive Empfinden betonen, wie das Vu-Meter mit recht langer Zeitkonstante. Die zweite Sorte betrifft jene, die den Spitzenwert des Signals auswerten. Die Forschung weist jedoch in Richtung eines der Sprache angepassten Gerätes, das den Wert der Sprachleistung während der aktiven Abschnitte eines Gesprächs misst. Dieser Wert lässt sich mit einem Algorithmus berechnen, der aufgrund der Verteilung der Sprachleistungswerte aufgestellt wurde. Abgesehen von Laboranfertigungen sind wenige solche Geräte auf dem Markt erhältlich. Das Problem kann jedoch mit ADDIO-2 elegant gelöst werden, indem ein Programm angewendet wird, das die Effektivwerte der verschiedenen, digital aufgezeichneten Sprachmuster berechnet. Wenn nötig, werden diese so korrigiert, dass der Wert der Sprachleistung jedes Satzes gleich bleibt. Definitionsgemäss liegt der Effektivwert der Sprache 15 dB tiefer als jener eines Sinussignals. Mit einer Sicherheitsmarge von 1 dB muss das Sprachsignal somit auf 19 dB unter die Überlastschwelle des Coders geregelt werden.

Die Bestimmung der Grenze zwischen den aktiven Abschnitten und den Pausen eines Gesprächs muss genau umschrieben werden. Pausen einer gewissen Länge sowie jene zwischen Sätzen müssen aus der Messung des Spracheffektivwertes ausgeschlossen werden. Hingegen müssen in die Berechnung die kurzen Unterbrüche, die Teil der Wortstruktur sind, einbezogen werden. Zum Beispiel erscheint in der üblichen Aussprache des französischen Wortes «acte» ein Unterbruch von ungefähr 200 ms zwischen dem «c» und dem «t», der jedoch berücksichtigt werden muss, da er vom Hörer als stetiger Ton empfunden wird.

742 Schneiden von Versuchsbändern und Korrektur von Aufnahmefehlern

Im allgemeinen bedingen die erwähnten Verfahren zur subjektiven Beurteilung der Sprachqualität eine wohlgeordnete Reihenfolge der von verschiedenen Sprechern vorgetragenen Sätze. Aus offensichtlichen praktischen Gründen entspricht diese Reihenfolge nicht immer jener der Aufnahmesitzung im Studio. Daher sind das Schneiden der Versuchsbänder und das digitale Kopieren mit zwei kompatiblen Aufnahmegeräten unumgänglich. Obwohl dabei keine zusätzliche Verschlechterung des Signals eintritt, ergeben sich zwei wichtige Nachteile: Der erste folgt aus dem noch sehr hohen Anschaffungspreis von professionellen, digitalen Aufnahmegeräten, der zweite bezieht sich auf den Zeitaufwand für die Suche der Sätze auf dem Originalband und für die Ausführung des Schneidens der Versuchsbänder. Dank der Einrichtung ADDIO können alle diese Such- und Zusammenstellungsarbeiten programmgesteuert durchgeführt werden. Dabei werden die Sätze in einer bestimmten Reihenfolge aneinandergefügt, und die Dauer der Abstände aufeinanderfolgender Sätze und irgendwelcher Pausen werden auf eine Abtastperiode genau eingehalten. Offensichtlich muss die Originalaufnahme direkt mit dem ADDIO-System, ohne Umweg über ein digitales Aufnahmegerät, angefertigt werden, damit die Qualität des Sprachsignals erhalten bleibt. Anderseits kann man nicht völlig verhindern, dass im Laufe von langen, ja mühsamen Aufnahmesitzungen, worin eine grosse Anzahl Sätze gelesen werden, Nebengeräusche aus der Umgebung des Raumes oder sogar versteckte Sprechfehler unbemerkt in die Aufnahme gelangen, dann aber bei der Schlussprüfung des Originalbandes auftauchen. Unter den strengen Versuchsbedingungen können sich diese kleinen Fehler trotzdem störend für den an der subjektiven Auswertung beteiligten Hörer auswirken. Dabei ist es nicht immer möglich, eine neue Aufnahmesitzung zu veranstalten.

Auch unter diesen besonderen Umständen erweist sich ADDIO als ein sehr leistungsfähiges Werkzeug. Das folgende Beispiel veranschaulicht die bemerkenswerte Vielseitigkeit und Wirksamkeit des Systems. Es stammt aus den vielen Vorbereitungsarbeiten für Versuchsbänder zur subjektiven Beurteilung von 64-kbit/s-ADPCM-Breitbandcodecs (7 kHz). Während dieser Arbeiten mussten am ursprünglichen Aufnahmeband einige Verbesserungen vorgenommen werden. Das Sprachsignal, das in *Figur 12* eingerahmt ist, soll ausgemerzt werden, da es einem Sprechfehler oder genauer einem unvollständig ausgesprochenen Wort entspricht. Durch aufmerksames Abhören des fehlerhaften Satzes wird dieser Signalabschnitt eingegrenzt und der Fehler anschliessend auf dem Terminalbildschirm markiert. Nachdem die Grenzen des auszumerzenden Signals bekannt sind und die Verbindungsstelle gewählt ist, muss nur noch der fehlerhafte Abschnitt mit einer Software-Routine ausgelöscht werden.

Diese enthält insbesondere eine Marke, die auf den Bildschirm bewegt werden kann, um genau Anfang und Ende des zu behandelnden Signalausschnittes zu bezeichnen. *Figur 13* zeigt das Endergebnis dieser Operationen. Für den Hörer ergibt sich damit eine fehlerlose Aufnahme.

75 Akustische Eigenschaften von Hörräumen [14, 15]

Zum Auffinden der Grenzen der technischen Möglichkeiten eines Verfahrens müssen die Programmabschnitte für die subjektiven Hörversuche im allgemeinen ohne hörbare Fehler sein. Man bemüht sich daher, im Rahmen des Möglichen, die Maskierungswirkung von Umgebungslärm oder anderen Störfrequenzen im Abhörraum zu vermindern. Die akustischen Eigenschaften des Raumes sind sehr wichtig, und seine Form kann vor allem bei kleinen Abmessungen ernsthafte Hallprobleme verursachen, wenn die Verhältnisse und Proportionen nicht sorgfältig gewählt sind. Aufgrund der Erfahrungen in der Akustik von Radiostudios ergeben sich folgende Empfehlungen für die hauptsächlichen physikalischen und akustischen Eigenschaften:

- Raumgrösse: 40...120 m^3, vorzugsweise 120 $m^3 \pm 30\ m^3$
- Raumform: Das beste Seitenverhältnis des Raumes ist 1,9 : 1,4: 1,0. Das Verhältnis 1:2 zwischen irgendwelchen Raumabmessungen ist zu vermeiden

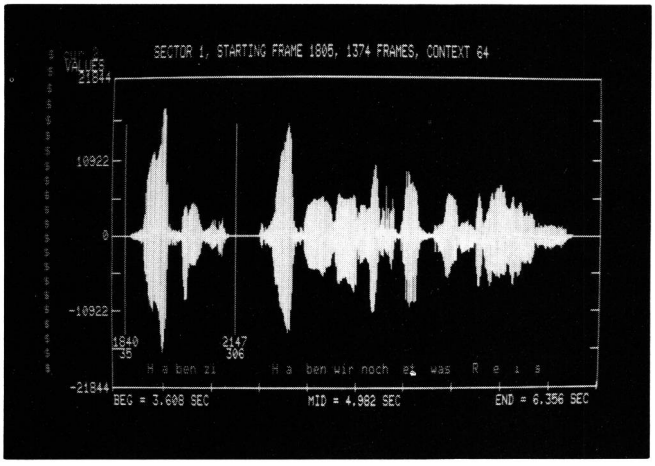

Fig. 12
Zu korrigierendes Sprachsignal
Eingerahmt: auszulöschender Teil

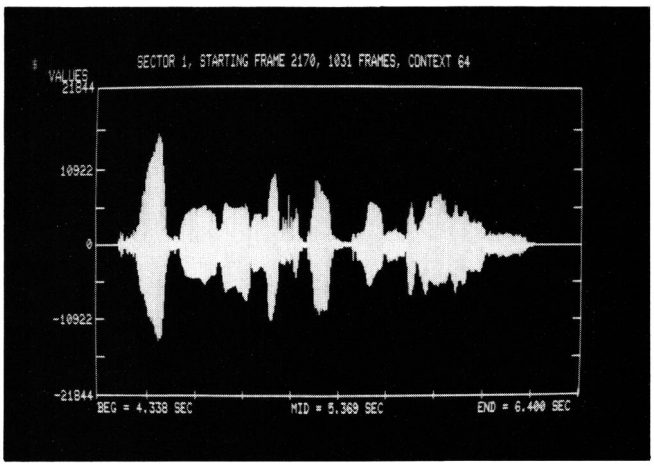

Fig. 13
Sprachsignal nach der Korrektur

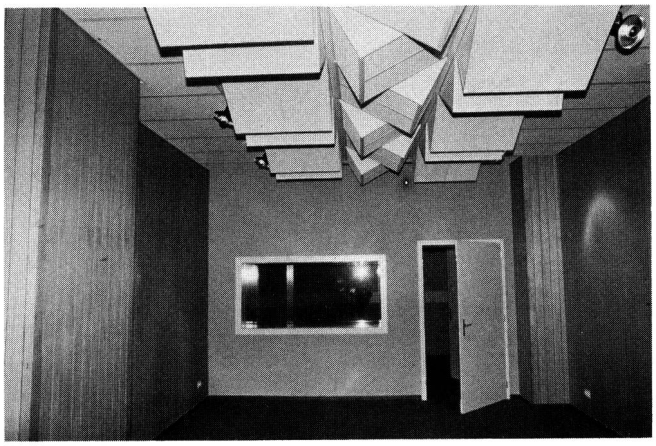

Fig. 14
Ansicht des neuen Akustikraums der PTT während dem Bau

- Nachhallzeit: 250 ms...400 ms (oberhalb 100 Hz)
- Geräuschpegel: Er muss genügend tief sein, um das Maskieren der schwächsten Signale zu verhindern. Am häufigsten werden Geräusche von aussen auf einen Wert unterhalb 25 dB gebracht und nach der «A»-Kurve gewichtet.

Figur 14 zeigt den neuen Akustikraum der Hauptabteilung Forschung und Entwicklung der PTT im Bau. Er ruht vollständig auf Puffern von 3,5 cm Dicke aus Kunststoff, die die Vibrationen und tiefen Frequenzen unterhalb 30 Hz wirksam unterdrücken. Die Anforderungen an diesen Hörraum sind sehr streng, damit das Hintergrundgeräusch auf einen Wert unterhalb 20 dBA gebracht werden kann. Die Isolation wurde so berechnet, dass Störungen aus andern Bereichen unterhalb des Wertes liegen, der im Saal ohne Störungen herrscht, und dies in allen Frequenzbändern.

Alle Aufnahme- und Wiedergabeausrüstungen, Messinstrumente, Recheneinheiten, Prozessoren usw., die eine nicht vernachlässigbare Störquelle darstellen, sind in einem Regieraum zusammengefasst, der vom Hörraum akustisch sorgfältig isoliert ist.

8 Schlussfolgerungen

Die beschriebenen Codier- und Kompressionsverfahren beziehen sich mehr oder weniger direkt auf ein Forschungsprojekt oder eine besondere Studie der PTT im Fernmeldewesen. Die Arbeiten sind angewiesen auf Unterstützung durch die Informatik, d. h. durch einen besonders ausgerüsteten Computer, um Sprachsignale und Musik in Echtzeit aufnehmen und wiedergeben zu können. In dieser Hinsicht hat sich ADDIO-2 schnell als bemerkenswert vielseitiges Werkzeug erwiesen, um ein Kompressionsverfahren zu verwirklichen oder einen Codieralgorithmus abzuändern.

Die Sorgfalt bei der Aufnahme von Tonquellen, bei der Wahl der Verfahren subjektiver Hörversuche und bei den Hörbedingungen sind äusserst wichtig, um sich vor Fehlern in der Würdigung und Beurteilung zu schützen, die von Verzerrungen herrühren, die nicht durch das eigentliche Codiersystem bedingt sind.

Die nächste Nummer bringt unter anderem
Vous pourrez lire dans le prochain numéro

5/86

Wuhrmann K. E.	Das integrierte Fernmeldesystem IFS — Standortbestimmung auf Anfang 1986 Il sistema integrato delle telecomunicazioni IFS — Il punto della situazione all'inizio del 1986
Stadler R.	Télétex et ses protocoles
Debrunner W. E.	Feldstärkeberechnung für die Funkplanung
Herheuser R.	Arbeitsschritte bei der Planung und Realisierung von Fernmeldediensten

---