



19 BUNDESREPUBLIK
DEUTSCHLAND



DEUTSCHES
PATENT- UND
MARKENAMT

12 Patentschrift
10 DE 198 06 015 C 2

51 Int. Cl.⁶:
H 04 M 1/58
H 04 M 1/60
H 04 M 9/08
H 03 H 21/00

21 Aktenzeichen: 198 06 015.7-31
22 Anmeldetag: 13. 2. 98
43 Offenlegungstag: 26. 8. 99
45 Veröffentlichungstag
der Patenterteilung: 23. 12. 99

DE 198 06 015 C 2

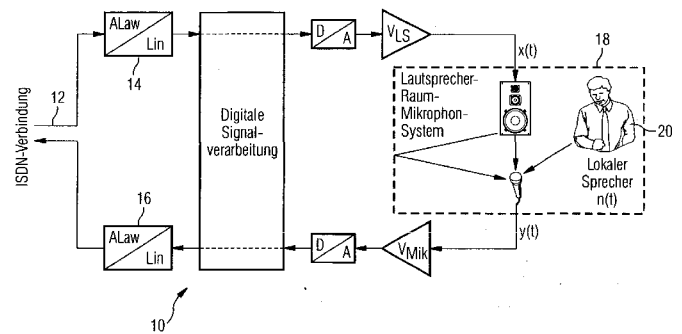
Innerhalb von 3 Monaten nach Veröffentlichung der Erteilung kann Einspruch erhoben werden

73 Patentinhaber:
Siemens AG, 80333 München, DE
72 Erfinder:
Schmidt, Gerhard, 63303 Dreieich, DE

56 Für die Beurteilung der Patentfähigkeit in Betracht
gezogene Druckschriften:
DE 42 27 327 A1
Werveprospekt der Fa. NEC: RefleXion Acoustic
Echo
Canceller on the uPD7701xFamily, 1996;
KNOX, M., ABBOT, P.: CYOX: A Highly Integrated
H320 Audiosubsystem using the Motorola DSP600x
Digitalprozessor;
TURBIN, V., GILLOIRE, A., SCALART, P.:
Comparison
Of Three Post-Filtering Algorithmus For Residual
Acoustic Echo Reduction, ICASSP97, International
Workshop on Acoustic Speech and Signal
Processing,
München 1997;
MARTIN, R.: An Improved Eco-shape Algorithm for
Acoustic Echo Control, EUSIPCO 96, 8th European
Signal Processing Conference, Triest, Italien,
1996;

54 Verfahren zur Verbesserung der akustischen Rückhördämpfung in Freisprecheinrichtungen

57 Verfahren zur Verbesserung der akustischen Rückhördämpfung in Freisprecheinrichtungen mit einer Pegelwaage (22) und einer frequenzselektiven steuerbaren Echokompensation (28) mit Teilbandverarbeitung, wobei das abgehende Signal nach der frequenzselektiven Echokompensation (28) einer Nachfilterung in einem weiteren frequenzselektiven Filter (30) mit Einstellalgorithmus gemäß einem Wiener-Ansatz unterworfen wird (Wiener-Filterung), dadurch gekennzeichnet, daß eine einzige Steuergröße (Schrittweitenvektor $\vec{\alpha}(k)$) sowohl für die Steuerung der frequenzselektiven Echokompensation, als auch für die Steuerung des weiteren Filters (30) verwendet wird.



DE 198 06 015 C 2

Die vorliegende Erfindung betrifft ein Verfahren zur Verbesserung der akustischen Rückhördämpfung in Freisprecheinrichtungen mit einer Pegelwaage, einer frequenzselektiven, steuerbaren Echokompensation mit Teilbandverarbeitung und einer Restfehlernachfilterung.

Bei Freisprecheinrichtungen ist es unbedingt erforderlich, die vom Lautsprecher ausgesandten und damit vom Mikrofon wieder aufgenommenen Signale des entfernten Teilnehmers zu unterdrücken, da sonst unangenehme Echos die Verbindung stören. Bisher wurde zur Unterdrückung dieser Echos, also zur akustischen Rückhördämpfung, üblicherweise eine Pegelwaage vorgesehen, die abhängig von der Gesprächssituation den Sende- oder den Empfangspfad stark dämpft. Dadurch wird jedoch ein Gegensprechen (Voll-Duplex-Betrieb) praktisch unmöglich.

Mit der bisherigen Technik wurde bereits versucht, eine ausreichende Rückhördämpfung trotz akzeptabler Gegensprechbetriebseigenschaften zur Verfügung zu stellen. Hierzu wurde zusätzlich zu der Pegelwaage eine frequenzselektive, steuerbare Echokompensation vorgesehen. Diesbezüglich wird auf die noch unveröffentlichte Patentanmeldung DE 197 14 966 der Anmelderin verwiesen. Andere Verfahren sind beispielsweise dem Werbeprospekt der Firma NEC "RefleXion™ Acoustic Echo Canceller on the μ PD7701x Family", 1996, oder aus der Beschreibung des Motorola DSP5600x Digitalprozessors (M. Knox, P. Abbot, Cyox: A Highly Integrated H320 Audiosubsystem using the Motorola DSP5600x Digitalprozessor" beschrieben. Auch diese Verfahren können jedoch bei den langen Signallaufzeiten von Videokonferenzverbindungen bzw. bei GSM-Verbindungen keine ausreichende Echounterdrückung bieten, wenn gleichzeitig ein Gegensprechen möglich sein soll.

Es wurde daher bereits vorgeschlagen, eine zusätzliche Nachfilterung nach der frequenzselektiven Echokompensation mit Teilbandverarbeitung vorzusehen. Eine solche Nachfilterung ist beispielsweise in dem Artikel "V. Turbin, A. Gilloire, P. Scalart: Comparison Of Three Post-Filtering Algorithmus For Residual Acoustic Echo Reduction" ICASSP97, International Workshop on Acoustic Speech and Signal Processing, München 1997, oder aus dem Artikel von R. Martin "An improved Echo-shape Algorithm for Acoustic Echo Control", EUSIPCO96, 8th European Signal Processing Conference, Triest, Italien, 1996, bekannt. Diese Konzepte ließen sich bisher nur schwer verwirklichen, da ja sowohl für die Echokompensation mit Teilbandverarbeitung als auch für die Nachfilterung eine digitale Signalverarbeitung vorzusehen ist, und die dafür erforderlichen Rechenleistungen vor den derzeit verfügbaren Prozessoren nicht mit angemessenem Aufwand erbracht werden können.

Es ist daher Aufgabe der Erfindung, ein Verfahren zur Verbesserung der akustischen Rückhördämpfung in Freisprecheinrichtungen anzugeben, bei dem der Rechenaufwand so minimiert ist, daß sowohl eine frequenzselektive Echokompensation mit Teilbandverarbeitung als auch die erforderliche Nachfilterung auf gebräuchlichen "Consumer-Prozessoren" verwirklicht werden können. Diese Aufgabe wird gelöst mit einem Verfahren mit den Merkmalen von Patentanspruch 1. Vorteilhafte Ausgestaltungen dieses Verfahrens sind in den Unteransprüchen angegeben.

Nach der Erfindung wird daher lediglich eine einzige Steuergröße, nämlich der Schrittweitenvektor, sowohl für die Steuerung der frequenzselektiven Echokompensation, als auch für die Steuerung des weiteren Filters verwendet. Vorzugsweise können dabei mehrere unterschiedliche Abstraten verwendet werden. Dadurch kann der Rechenaufwand weiter verringert werden.

Ebenso ist es bevorzugt, sowohl für die Echokompensation als auch für das weitere Filter adaptive Filter zu verwenden.

Die Echokompensation wird vorzugsweise mittels einer Filterbank in Frequenzteilbändern implementiert.

Vorzugsweise werden für die Adaptions- bzw. die Schrittweitensteuerung sowohl leistungsbasierende Schätzungen als auch korrelationsbasierende Analysen verwendet.

Ebenso ist es bevorzugt, zur Schrittweitenbestimmung Leistungsübertragungsfaktoren in Teilbändern zu schätzen.

Ebenso ist es bevorzugt, daß sowohl die Echokompensatoren als auch die Restfehlernachfilterung die Schätzwerte für die durch sie eingebrachte Echodämpfung liefern, da diese Schätzwerte bevorzugt zur Steuerung der Dämpfung der Pegelwaage verwendet werden können. Dadurch kann die von der Pegelwaage einzubringende Dämpfung weiter reduziert und damit die Gesprächsqualität beim Gegensprechen weiter verbessert werden.

Zusätzlich ist es bevorzugt, die gleichzeitige Aktivität beider Gesprächsteilnehmer (Gegensprechen) zu detektieren. Es ist dann beispielsweise möglich, die Gesamtdämpfung der Pegelwaage im Gegensprechfall zu reduzieren, um die Gegensprechfähigkeit (Full-Duplex-Betrieb) der Freisprecheinrichtung weiter zu verbessern.

Die vorliegende Erfindung wird im folgenden anhand des in den beigefügten Zeichnungen dargestellten Ausführungsbeispiels näher beschrieben. Es zeigt:

Fig. 1 ein vereinfachtes Modell einer Freisprecheinrichtung mit Anschluß an eine digitale Verbindung;

Fig. 2 ein Blockschaltbild der erfindungsgemäßen Freisprecheinrichtung;

Fig. 3 Kurven für die Dämpfungsanforderungen an die Freisprecheinrichtung in Abhängigkeit von der Echolaufzeit;

Fig. 4 eine Übersichtsdarstellung des erfindungsgemäßen Verfahrens;

Fig. 5 die Struktur der Adaption der Teilbandechokompensatoren;

Fig. 6 eine Modellvorstellung für die Leistungsübertragungsfaktoren;

Fig. 7 eine Darstellung der Signale des fernen und des lokalen Teilnehmers anhand derer im folgenden das erfindungsgemäße Verfahren erläutert wird;

Fig. 8 die daraus resultierende Anregung und der gestörte Fehler im Band 1;

Fig. 9 den geschätzten Leistungsübertragungsfaktor unter den Bedingungen gem. **Fig. 7** und 8 im Band 1;

Fig. 10 die von der Schrittweitensteuerung gewählte Schrittweite im Band 1 unter den Bedingungen gem. **Fig. 7** und 8;

Fig. 11 die erfindungsgemäße Glättung der Dämpfungsabsenkung;

Fig. 12 eine Detaildarstellung der Nachfilterung des Fehlersignals;

Fig. 13 die erfindungsgemäße Glättung der Schrittweiten (Teil A für gleiche Zeitkonstanten, Teil B für unterschiedliche Zeitkonstanten);

Fig. 14 ein weiteres Beispiel für die Signale des fernen und des lokalen Teilnehmers, die in den weiteren Figuren der

Verarbeitung zugrunde liegen;

Fig. 15 den Abgleichverlauf und die Dämpfung durch das weitere Filter im Band 1;

Fig. 16 die Dämpfung durch das weitere Filter in Band 1;

Fig. 17 die Übergabe der Dämpfungswerte an die Pegelwaage; und

Fig. 18 die Anregungs- und Fehlerleistung im Gesamtband (jeweils für den Eingangssignalverlauf gem. **Fig. 14**).

In **Fig. 1** ist ein vereinfachtes Modell einer Freisprecheinrichtung **10** mit Anschluß an eine digitale Verbindung **12** dargestellt. Die im europäischen ISDN-Netz verwendete A-Law-Codierung bzw. Decodierung ist in den beiden linken Blöcken **14, 16** dargestellt. Auf der rechten Seite ist das Lautsprecher-Raum-Mikrophonsystem **18** (LRM-System) mit dem lokalen Gesprächsteilnehmer **20**, dem Benutzer der Freisprecheinrichtung, skizziert.

Durch die akustische Kopplung zwischen Lautsprecher und Mikrophon kommt es zum Übersprechen über das LRM-System. Dieses Übersprechen wird vom fernen Teilnehmer als störendes Echo wahrgenommen. Akustische Wellen treten dabei aus dem Lautsprecher aus und breiten sich im Raum aus. Durch Reflexion an den Wänden und anderen sich im Raum befindlichen Gegenständen entstehen mehrere Ausbreitungspfade, durch die unterschiedliche Laufzeiten des Lautsprecher Signals entstehen. Das Echosignal am Mikrophon besteht somit aus der Überlagerung einer Vielzahl von Echoanteilen und ggf. dem Nutzsignal $n(t)$: dem lokalen Sprecher.

Auch die Verbindung zwischen den Teilnehmern kann an Übergängen zwischen verschiedenen Übertragungssystemen Echos erzeugen. Die Netzbetreiber versuchen jedoch, direkt an den kritischen Stellen besondere Maßnahmen gegen derartige Echoquellen zu treffen, so daß diese Echos hier außer Acht gelassen werden können. Auch Gabelechos, die in Telefonen mit analogem Interface durch Fehlanpassung der Leitungsnachbildung an die Leitungsimpedanz entstehen, können durch die Verwendung von digitalen Verbindungen außer Betracht gelassen werden.

In **Fig. 2** ist eine Übersicht der erfindungsgemäßen Freisprecheinrichtung dargestellt. Zentrales Element ist eine Pegelwaage **22**, welche im linken Teil der **Fig. 2** dargestellt ist. Optional können zwei Verstärkungssteuerungen **24, 26** (Automatic Gain Control = AGC) in den Sende- und den Empfangspfad eingeschaltet werden. Die Pegelwaage **22** garantiert die durch die ITU- bzw. ETSI-Empfehlungen vorgeschriebenen Mindestdämpfungen, in dem sie abhängig von der Gesprächssituation Dämpfungen in den Sende- und/oder den Empfangspfad einfügt. Bei Aktivität des fernen Teilnehmers wird der Empfangspfad freigeschaltet und das Signal des fernen Teilnehmers wird ungedämpft auf dem Lautsprecher ausgegeben. Die bei abgeschalteten oder schlecht abgeglichenen Kompensatoren entstehenden Echos werden durch die in den Sendepfad eingefügte Dämpfung stark verringert. Bei Aktivität des lokalen Sprechers kehrt sich die Situation um. Während der Empfangspfad stark bedämpft wird, fügt die Pegelwaage **22** in den Sendepfad keine Dämpfung ein und das Signal des lokalen Sprechers wird ungedämpft übertragen. Schwieriger wird die Steuerung der Pegelwaage im Gegensprechfall. Hier erhalten beide Pfade (und damit auch die Teilnehmersignale) jeweils die Hälfte der einzufügenden Dämpfung oder bei nicht optimaler Steuerung wird zumindest einer der beiden Signalpfade gedämpft. Gegensprechen ist damit nicht oder nur eingeschränkt möglich.

Abhilfe schafft hier der Einsatz von adaptiven Echokompensatoren **28** – dargestellt im rechten Teil der **Fig. 2**. Diese versuchen das LRM-System digital nachzubilden, um dann den Echoanteil des fernen Teilnehmers aus dem Mikrophonsignal herauszurechnen. Je nachdem, wie gut die Kompensatoren dies bewerkstelligen, kann die durch die Pegelwaage einzufügende Gesamtdämpfung reduziert werden.

Die Echokompensation wurde in Frequenzteilbändern implementiert, wobei die Breite der einzelnen Bänder vorzugsweise zwischen 250 Hz und 500 Hz bei 8 kHz Abtastrate bzw. zwischen 500 Hz und 1000 Hz bei 16 kHz Abtastrate liegt. Der Einsatz einer frequenzselektiven Echokompensation hat mehrere Vorteile. Zum einen kann durch Verwendung von Unter- und Überabtastrung das System als Multiratsensystem betrieben werden, wodurch sich der Berechnungsaufwand verringert. Zum anderen kann durch die Teilbandzerlegung die "Kompensationsleistung" unterschiedlich auf die einzelnen Frequenzbereiche verteilt werden und somit eine effektive Anpassung der "Kompensationsleistung" an Sprachsignale erreicht werden. Weiter hat die Teilbandverarbeitung eine decorrelierende Wirkung, wenn die Gesamtbandverarbeitung mit den einzelnen Teilbandsystemen verglichen wird. Für Sprachsignale bedeutet dies eine Erhöhung der Konvergenzgeschwindigkeit der adaptiven Filter. Neben diesen Vorteilen darf der Nachteil einer Teilbandverarbeitung nicht außer Acht gelassen werden. Die Zerlegung eines Signals in einzelne Frequenzbereiche bewirkt stets eine Laufzeit – im vorliegenden bevorzugten Verfahren **32** ms bei 8 kHz Abtastrate bzw. 16 ms bei 16 kHz Abtastrate. Da das Verfahren jedoch für Videokonferenzen bzw. in GSM-Mobiltelefonen eingesetzt wird, sind solche Laufzeiten zulässig.

In Videokonferenzsystemen wird die Laufzeit hauptsächlich von der bildverarbeitenden Komponente bestimmt. Da im allgemeinen versucht wird, dem lokalen Teilnehmer Bild und Ton des fernen Teilnehmers lippensynchron auszugeben, kann sich die Laufzeit der akustischen Echos auf mehrere hundert Millisekunden erhöhen. In **Fig. 3** sind die Ergebnisse einer Studie dargestellt, in der versucht wurde, herauszufinden, welche Echodämpfung abhängig von der Laufzeit dieses Echos notwendig ist, damit 90, 70 bzw. 50 Prozent der Befragten mit der Gesprächsqualität zufrieden waren.

Basierend auf dieser Studie sind bei der reinen Audiolaufzeit von 30–40 ms (bei 8 kHz Abtastrate) lediglich 35 dB Echodämpfung notwendig. Bei lippensynchroner Ausstrahlung von Bild und Ton und einer damit verbundenen Laufzeit von beispielsweise 300 ms erhöht sich die Anforderung auf 53 dB. Auch in GSM-Verbindungen kann die Laufzeit mehr als 100 ms betragen. Die Anforderungen, die an Echokompensationsverfahren in Videokonferenz- und GSM-Systemen gestellt werden, sind somit höher als die Anforderungen an herkömmliche Freisprechtelefone.

Da die Echokompensatoren in ihrer Leistungsfähigkeit begrenzt sind und derart hohe Echodämpfungen mit der zur Verfügung stehenden Hardware nicht erreichen können, wurde ein sog. Postfilter **30** eingeführt. Dieses wertet die Schrittweiten der einzelnen Teilbänder zusammen mit den anderen Detektoregebnissen aus und filtert das Synthesefilterausgangssignal nochmals frequenzselektiv. Da der Einstellalgorithmus des Filters **30** gemäß einem Wiener-Ansatz entworfen wurde, wird diese Postfilterung im folgenden auch mit Wiener-Filterung bezeichnet.

Die Steuerung der Echokompensatoren erfolgt in mehreren Stufen. Alle leistungsbasierenden Steuereinheiten **32** arbeiten für jeden Kompensator autonom, also unabhängig von den restlichen Frequenzbereichen. In **Fig. 2** ist daher für jeden Kompensator eine eigene Adaptions- und Steuereinheit **32** skizziert. Die auf Korrelationsanalysen des Lautsprecher- und des Mikrophonsignals basierende Stufe der Steuerung wird zur Gegensprechdetektion verwendet und daher in allen

Frequenzbereichen gleichermaßen ausgewertet. Eine weitere Stufe trägt der durch die Festkommaarithmetik begrenzten Genauigkeit Rechnung und steuert die Adaption in Abhängigkeit der Aussteuerung.

Die endgültige Gegenseprechererkennung erfolgt ebenfalls gesondert mit einer eigenen Einheit, die sich sowohl auf die Detektoren der Pegelwaage als auch auf die der Echokompensatoren stützt. Diese Einheit veranlaßt die Pegelwaage in Gegenseprechsituationen die einzufügende Gesamtdämpfung nochmals (gemäß der ITU-Empfehlung G. 167) zu reduzieren.

In Fig. 4 ist eine Übersichtsdarstellung des oben beschriebenen Zusammenhangs wiedergegeben. Zentrales Element ist hierbei die Berechnung des Schrittweitenvektors $\bar{\alpha}(k)$. Dieser wird sowohl zur Steuerung der Teilbandechokompensatoren als auch zur Berechnung der Koeffizienten des Postfilters verwendet. Die beiden Teilverfahren berechnen jeweils die durch sie hervorgerufene Echodämpfung und teilen diese Informationen der Pegelwaage 22 mit. Die Waage 22 reduziert dann die vom Benutzer eingestellte Gesamtdämpfung und fügt nur noch die restliche Dämpfung in den Sende- bzw. den Empfangspfad ein.

Da sich die vorliegende Erfindung auf die Kombination der oben erwähnten Wiener-Filterung und der Adaptionsteuerung der Teilbandechokompensatoren bezieht, werden beide Verfahren in eigenen Kapiteln detailliert beschrieben. Neu an dem vorgestellten Ansatz ist die Verwendung einer einzigen Steuergröße – dem Schrittweitenvektor $\bar{\alpha}(k)$ – für beide Verfahren. Durch den hierdurch verringerten Rechenaufwand (weniger 100 Zyklen/Abtasttakt für die Postfilterung) wird es ermöglicht, beide Verfahren auf preiswerten "Consumer"-Signalprozessoren zu implementieren und damit die Qualität der Freisprecheinrichtung zu erhöhen.

Bisherige Ansätze zur Fehlernachfilterung verwenden zunächst eine (aufwendige) FFT-Analyse bzw. andere rechenleistungsintensive Berechnungsverfahren und betrachten die Steuerung der Postfilterung stets getrennt von der Steuerung der Echokompensation.

Die für die Teilbandverarbeitung notwendige Frequenzbandanalyse und -synthese ist als Polyphasenfilterbank implementiert.

Zuerst wird – zunächst unabhängig von der späteren Verwendung innerhalb der Wiener-Filterung – eine Schrittweitensteuerung beschrieben, welche eine schnelle und stabile Adaption der Teilbandechokompensatoren gewährleistet. Zusätzlich werden Verfahren vorgestellt, welche die erreichte Echodämpfung schätzen. Die Pegelwaage 22 kann somit – basierend auf diesen Schätzwerten – die Gesamtdämpfung reduzieren. Für die Dämpfungsschätzung ist es dabei unerheblich, ob die Dämpfung von gut abgeglichenen Echokompensatoren, durch die akustische Anordnung von Lautsprecher und Mikrophon oder durch eine entsprechende Wahl der analogen Verstärkungen erreicht wird.

Die Adaption der Teilbandechokompensatoren wird mittels eines auf den verwendeten Signalprozessor angepaßten NLMS-Verfahrens durchgeführt. Um die Notation der folgenden Beschreibung zu erläutern, ist in Fig. 5 eine Strukturdarstellung des Adaptionsprozesses wiedergegeben.

Durch Faltung der geschätzten Teilbandimpulsantworten $c_{\mu}^{(r)}(k_r)$ mit den Teilbandanregungssignalen des fernen Teilnehmers $\underline{\chi}_{\mu}^{(r)}(k_r)$ werden die geschätzten Mikrophonsignale $\underline{\hat{y}}_{\mu}^{(r)}(k_r)$ gebildet:

$$\begin{aligned} \underline{\hat{y}}_{\mu}^{(r)}(k_r) &= \underline{\chi}_{\mu}^{(r)}(k_r) * \underline{c}_{\mu}^{(r)}(k_r) \\ &= \underline{\tilde{\chi}}_{\mu}^{(r)H}(k_r) \underline{\tilde{c}}_{\mu}^{(r)}(k_r) \end{aligned} \quad (3.1)$$

Der Index μ soll dabei die Teilbandnummer anzeigen. Durch Differenzbildung zwischen dem geschätzten und dem gemessenen Mikrophonsignal wird der Adaptionsfehler $\underline{e}_{\mu}^{(r)}(k_r)$ berechnet:

$$\underline{e}_{\mu}^{(r)}(k_r) = \underline{y}_{\mu}^{(r)}(k_r) - \underline{\hat{y}}_{\mu}^{(r)}(k_r) \quad (3.2)$$

Dieser Fehler setzt sich aus einem sog. ungestörten Fehler $\underline{\epsilon}_{\mu}^{(r)}(k_r)$ und den durch den lokalen Sprecher hervorgerufene Anteil $\underline{n}_{\mu}^{(r)}(k_r)$ zusammen:

$$\underline{e}_{\mu}^{(r)}(k_r) = \underline{\epsilon}_{\mu}^{(r)}(k_r) + \underline{n}_{\mu}^{(r)}(k_r). \quad (3.3)$$

Die Adaption erfolgt mittels einer Näherung des NLMS-Algorithmus

$$\underline{\tilde{c}}_{\mu}^{(r)}(k_r + 1) = \underline{\tilde{c}}_{\mu}^{(r)}(k_r) + F \left(\underline{\alpha}_{\mu}^{(r)}(k_r) \frac{\underline{\tilde{\chi}}_{\mu}^{(r)}(k_r) \underline{e}_{\mu}^{(r)*}(k_r)}{\underline{\tilde{\chi}}_{\mu}^{(r)H}(k_r) \underline{\tilde{\chi}}_{\mu}^{(r)}(k_r)} \right), \quad (3.4)$$

wobei mit $F(\chi)$ die bereits angesprochene Näherungsfunktion bezeichnet ist.

Die Koeffizienten der Teilbandechokompensatoren werden während des Betriebs der Freisprecheinrichtung mit den Adaptionsverfahren laufend an die Teilbandimpulsantworten des LRM-Systems angepaßt. Damit kann auch nach Systemänderungen eine Reduktion der akustischen Echos erreicht werden. Das Einstellkriterium für das verwendete Adaptionsverfahren ist die Minimierung des mittleren quadratischen Fehlers. Gemäß der Rechenvorschrift des NLMS-Algorithmus erfahren die Koeffizienten eine starke Änderung, wenn die Abtastwerte des kompensierten Signals $\underline{e}_{\mu}^{(r)}(k_r)$ des μ -ten Teilbandes groß sind. Andauernd große Werte $\underline{e}_{\mu}^{(r)}(k_r)$ können auf zwei Ursachen zurückgeführt werden:

1. Nach Änderungen im LRM-System sind die adaptiven Filter schlecht an die Raum-Impulsantwort angepaßt. Es findet dann keine oder eine nur geringe Reduktion der akustischen Echos statt – die unkompenzierten Echoanteile bewirken eine Vergrößerung der Signale $\underline{e}_{\mu}^{(r)}(k_r)$. Die Kompensatoren sollten in solchen Situationen möglichst

schnell angeglichen werden.

2. Eine Erhöhung des lokalen Anteils $n(k)$ – beispielsweise bei Aktivität des lokalen Sprechers – bewirkt ebenfalls eine Vergrößerung der Signal $e_{\mu}^{(r)}(k_r)$. Dieser Anteil ist für die Freisprecheinrichtung das zu übertragende Nutzsignal, für die adaptiven Filter stellt er jedoch ein Störung dar, die zu einer Fehleinstellung der Koeffizienten führen kann. In solchen Situationen sollten die Kompensatoren nicht oder nur wenig verstellt werden, damit der bereits erreichte Abgleich nicht wieder verschlechtert wird.

Es wurde bereits eine Schrittweitensteuerung vorgestellt, welche die beiden beschriebenen Gesprächssituationen bzw. Zustände der Kompensatoren berücksichtigt und die gestellten Forderungen an die Adaptionssteuerung erfüllt. Die Schrittweite im μ -ten Teilband sollte gemäß

$$\alpha_{\mu}^{(r)}(k_r) = \frac{E \left\{ \left| \underline{e}_{\mu}^{(r)}(k_r) \right|^2 \right\}}{E \left\{ \left| \underline{e}_{\mu}^{(r)}(k_r) \right|^2 \right\}} \quad (3.5)$$

eingestellt werden. Das gestörte Fehlersigna $\underline{e}_{\mu}^{(r)}(k_r)$ im Nenner der Gleichung 3.5 ist direkt meßbar – der Erwartungswert davon kann durch

$$E \left\{ \left| \underline{e}_{\mu}^{(r)}(k_r) \right|^2 \right\} \approx \overline{\left| \underline{e}_{\mu}^{(r)}(k_r - 1) \right|^2}. \quad (3.6)$$

abgeschätzt werden. Die rechte Seite der Näherung 3.6 soll dabei eine rekursive Glättung erster Ordnung bezeichnen:

$$\overline{\left| \underline{e}_{\mu}^{(r)}(k_r) \right|^2} = \beta_e \overline{\left| \underline{e}_{\mu}^{(r)}(k_r) \right|^2} + (1 - \beta_e) \overline{\left| \underline{e}_{\mu}^{(r)}(k_r - 1) \right|^2}. \quad (3.7)$$

Für die Abschätzung des Zählers wird ein Leistungsübertragungsfaktor $\tilde{p}_{\mu}^{(r)}(k_r)$ eingeführt. Dabei wird die Parallelschaltung aus LRM-System und Echokompensator einschließlich der Subtraktionsstelle in erster Näherung als einfaches Dämpfungsglied modelliert.

Die Größe dieser Dämpfung (Verhältnis von Anregungs- zu Fehlerleistung) wird durch den Leistungsübertragungsfaktor im Teilband

$$\tilde{p}_{\mu}^{(r)}(k_r) = \frac{\left| \underline{e}_{\mu}^{(r)}(k_r) \right|^2}{\left| \underline{x}_{\mu}^{(r)}(k_r) \right|^2} \quad \text{mit } k_r \in K_{ES,FT} \quad (3.8)$$

abgeschätzt. Das Modell setzt hierbei voraus, daß im LRM-System keine zusätzlichen Störungssignale – wie z. B. Aktivität des lokalen Sprechers – vorhanden sind. In Gleichung 3.8 wurde aus diesem Grund die Menge $K_{ES,FT}$ eingeführt. Diese Menge soll die Zeitpunkte, in welchen sich die Freisprecheinrichtung im Zustand Einzelsprechen des fernen Teilnehmers befindet, beinhalten.

Das in Gleichung 3.8 verwendete, geglättete quadratische Anregungssignal wird dabei analog zur geschätzten Fehlerleistung bestimmt:

$$\overline{\left| \underline{x}_{\mu}^{(r)}(k_r) \right|^2} = \beta_x \overline{\left| \underline{x}_{\mu}^{(r)}(k_r) \right|^2} + (1 - \beta_x) \overline{\left| \underline{x}_{\mu}^{(r)}(k_r - 1) \right|^2} \quad (3.9)$$

In Zuständen ohne Raumänderung wird sich der Leistungsübertragungsfaktor im Vergleich zu den (Kurzzeit-) Anregungsleistungen nur sehr langsam ändern. Zur Verbesserung der Varianz der obigen Schätzung können damit rekursive Glättungen mit großen Zeitkonstanten verwendet werden. Die Bezeichnung groß ist dabei im Verhältnis zu den Zeitkonstanten bei den Leistungsschätzungen zu sehen.

Bei Aktivität des lokalen Teilnehmers wird die Schätzung des Restechos stark gestört. In solchen Fällen sollte die Erneuerung der Schätzung des Leistungsübertragungsfaktors nicht vorgenommen werden – die zuletzt berechneten $\tilde{p}_{\mu}^{(r)}(k_r)$ werden beibehalten. Durch diese Maßnahme können Raumänderungen bei Aktivität des lokalen Sprechers nicht detektiert werden. Erst nach dem erneuten Erreichen des Zustands Einzelsprechen des fernen Teilnehmers werden die Leistungsübertragungsfaktoren in solchen Fällen angeglichen. Die Bestimmungsgleichung für die geglätteten Leistungsübertragungsfaktoren kann damit gemäß

$$\overline{\tilde{p}_{\mu}^{(r)}(k_r)} = \begin{cases} \beta_p \overline{\tilde{p}_{\mu}^{(r)}(k_r)} + (1 - \beta_p) \overline{\tilde{p}_{\mu}^{(r)}(k_r - 1)} & \text{für } k_r \in K_{ES,FT} \\ \overline{\tilde{p}_{\mu}^{(r)}(k_r - 1)} & \text{sonst} \end{cases} \quad (3.10)$$

angegeben werden. Die Schrittweiten $\alpha_{\mu}^{(r)}(k_r)$ können wie folgt angenähert werden:

$$\alpha_{\mu}^{(r)}(k_r) \approx \frac{|\overline{x_{\mu}^{(r)}(k_r)}|^2 \overline{p_{\mu}^{(r)}(k_r)}}{|\overline{e_{\mu}^{(r)}(k_r)}|^2} \quad (3.11)$$

Aus den bisherigen Überlegungen folgt, daß die Bestimmung der Leistungsübertragungsfaktoren in zwei Teile untergliedert werden kann. Zum einen muß eine effektive Berechnung der beiden Leistungsschätzungen bzw. der Divisionen dieser beiden Größen auf der zur Verfügung stehenden Hardware gefunden werden. Zum anderen müssen die Zeitpunkte, welche in der Menge $K_{ES,FT}$ enthalten sind, detektiert werden.

Für das erste Teilproblem wurden nichtlineare, rekursive Glättungen verwendet. Als Eingangssignale dieser Filter wurde die Summe aus dem Betrag des Realteils und dem Betrag des Imaginärteils der Teilbandsignale gewählt. Zur Vermeidung der Division wurden die Leistungsfaktoren logarithmisch berechnet – die Division kann somit durch eine Subtraktion ersetzt werden.

Für das zweite Teilproblem wurde ein sog. Korrelationsmaß $\xi(k_r)$ eingesetzt. Hierbei wird eine normierte Kreuzkorrelationsanalyse des Anregungssignals des fernen Teilnehmers und des Mikrophonsignals durchgeführt. Bei Einzelsprechen des fernen Teilnehmers sind die beiden Signale stark korreliert und das Korrelationsmaß liefert Werte $\xi(k_r) \approx 1$. Bei Aktivität des lokalen Gesprächsteilnehmers verringert sich die Korrelation und es werden Werte $\xi(k_r) < 1$ detektiert.

Zur Verdeutlichung der hier folgenden Überlegungen wurde die Steuerung mit den in **Fig. 7** dargestellten Eingangssignalen des fernen und des lokalen Gesprächsteilnehmers getestet.

Für beide Signale wurde in den Aktivitätsphasen weißes, gaußverteiltes Rauschen gewählt. Zu Beginn der Sequenz liegt "Einzelsprechen" des fernen Teilnehmers vor (Phase A_1). Die adaptiven Echokompensatoren können in dieser Phase abgleichen und erreichen nach etwa 3 bis 4 Sekunden ihren Endabgleich. Nach 7,5 Sekunden beginnt der lokale Teilnehmer den fernen zu unterbrechen (Gegensprechen, Bereich B_1) und übernimmt dann die Rolle des "Alleinsprechenden" (Bereich C). Nach 10,75 Sekunden kehrt sich die Situation um. Der ferne Teilnehmer unterbricht den lokalen (Gegensprechen, Bereich B_2) und "redet" schließlich allein weiter (Phase A_2).

Das Mikrophonsignal wird durch Faltung des Anregungssignals mit der bereits vorgestellten Impulsantwort eines Bürraumes (Länge 2044 Koeffizienten bei 8 kHz Abtastrate) und anschließender Addition des Signals des lokalen Sprechers gebildet.

In **Fig. 8** sind die mittleren Leistungen des Anregungs- und des Fehlersignals dargestellt. Die Adaption wurde mit der im folgenden beschriebenen Schrittweitensteuerung durchgeführt, wobei davon ausgegangen wird, daß die Korrelationsauswertungen nur in den Bereichen A_1 und A_2 Freigaben liefern. In der Abbildung ist deutlich zu erkennen, daß der im Laufe der Phase A_1 erreichte Abgleich von etwa 25 dB über die Bereiche des Gegensprechens und des Einzelsprechens des lokalen Teilnehmers gehalten werden kann.

Zur Bestimmung des Leistungsübertragungsfaktors im μ -ten Teilband müssen gemäß Gleichung 3.8 die mittleren Leistungen des Anregungssignals und des ungestörten Fehlersignals geschätzt werden. Um das Problem von Grenzyklen zu vermeiden, wäre bei direkter Ausführung der Glättung wie sie in Gleichung 3.7 bzw. in Gleichung 3.9 vorgeschlagen wurde, eine Rechnung in Doppelwort-Genauigkeit (32 Bit) notwendig. Um den damit verbundenen Speicherbedarf bzw. die benötigte Rechenleistung zu reduzieren, werden lediglich Betragsglättungen durchgeführt:

$$|\overline{\chi_{\mu}^{(r)}(k_r)}| = \beta_x |\overline{\chi_{\mu}^{(r)}(k_r)}| + (1 - \beta_x) |\overline{\chi_{\mu}^{(r)}(k_r - 1)}| \quad (3.12)$$

$$|\overline{e_{\mu}^{(r)}(k_r - 1)}| = \beta_e |\overline{e_{\mu}^{(r)}(k_r)}| + (1 - \beta_e) |\overline{e_{\mu}^{(r)}(k_r - 1)}| \quad (3.13)$$

Damit der kritische Fall der Aktivität des lokalen Teilnehmers bei Gegensprechen möglichst schnell erkannt werden kann, wurden bei der Glättung des Fehlersignals zwei unterschiedliche Zeitkonstanten (β_{er} und β_{ef}) für steigende und fallende Flanken eingeführt. Die Zeitkonstante β_e wird gemäß

$$\beta_e = \begin{cases} \beta_{er} & \text{falls } |\overline{e_{\mu}^{(r)}(k_r)}| > |\overline{e_{\mu}^{(r)}(k_r - 1)}| \\ \beta_{ef} & \text{sonst} \end{cases} \quad (3.14)$$

mit $0 < \beta_{er} < \beta_{ef} < 1$

gebildet. Die so erhaltene Schätzung verliert durch die Wahl von zwei unterschiedlichen Zeitkonstanten ihre Erwartungstreue. Aus diesem Grund werden im Stand der Technik Korrekturfaktoren eingeführt. Hier soll ein anderer Weg eingeschlagen werden. Die Schätzung der Anregungsleistung erfolgt mit den gleichen Zeitkonstanten wie die Schätzung der Fehlerleistung:

$$\beta_x = \begin{cases} \beta_{\mathcal{X}} & \text{falls } |\underline{x}_\mu^{(r)}(k_r)| > \overline{|\underline{x}_\mu^{(r)}(k_r - 1)|} \\ \beta_{\mathcal{X}'} & \text{sonst} \end{cases} \quad (3.15)$$

mit $\beta_{\mathcal{X}'} = \beta_{er}$ und $\beta_{\mathcal{X}} = \beta_{ef}$.

Durch die anschließende Division der beiden Größen kann auf den Korrekturfaktor verzichtet werden. Die Betragsbildungen wurden durch die aufwandgünstigeren Abschätzungen

$$|\underline{x}_\mu^{(r)}(k_r)| \approx \left| \operatorname{Re}\{\underline{x}_\mu^{(r)}(k_r)\} \right| + \left| \operatorname{Im}\{\underline{x}_\mu^{(r)}(k_r)\} \right| \quad (3.16)$$

$$|\underline{e}_\mu^{(r)}(k_r)| \approx \left| \operatorname{Re}\{\underline{e}_\mu^{(r)}(k_r)\} \right| + \left| \operatorname{Im}\{\underline{e}_\mu^{(r)}(k_r)\} \right| \quad (3.17)$$

angenähert. Auch hier kann wieder ein Korrekturterm durch die Divisionsbildung weggelassen werden. Wie bereits im vorigen Abschnitt erwähnt, werden die Leistungsübertragungsfaktoren nur logarithmisch bestimmt – die Division wird dadurch auf zwei Logarithmierungen und eine Subtraktion zurückgeführt. Die Leistungsübertragungsfaktoren werden somit gemäß

$$p_{\mu,\log}^{(r)}(k_r) = \operatorname{LOG} \left\{ \overline{|\underline{e}_\mu^{(r)}(k_r)|} \right\} - \operatorname{LOG} \left\{ \overline{|\underline{x}_\mu^{(r)}(k_r)|} \right\} \quad (3.18)$$

und

$$\overline{p_{\mu,\log}^{(r)}(k_r)} = \begin{cases} \beta_p p_{\mu,\log}^{(r)}(k_r) + (1 - \beta_p) \overline{p_{\mu,\log}^{(r)}(k_r - 1)} & \text{für } k_r \in K_{ES,FT} \\ \overline{p_{\mu,\log}^{(r)}(k_r) - 1} & \text{sonst} \end{cases} \quad (3.19)$$

geschätzt. Mit $\operatorname{LOG} \{ \dots \}$ wird dabei die Logarithmierung bezeichnet. Die Zeitkonstante β_p , wurde ebenfalls unterschiedlich für steigende und fallende Flanken gewählt. Hiermit soll dem nicht kompensierbaren Teil der Systemlaufzeit (künstliche Verzögerung des Mikrophonsignals) gerecht werden. Durch diese Laufzeit fällt die Signalleistung des Anregungssignals früher ab als die des Fehlersignals – ohne Korrektur dieses Vorgangs würde die Schätzung eine Absenkung des Schätzwertes nach jeder Anregungsphase durchführen. Zusätzlich werden bei Detektion von Gegensprechen die Zeitkonstanten erhöht. Der verwendete Gegensprechdetektor ist weiter unten beschrieben. Die Bestimmungsgleichung für die Zeitkonstante β_p lautet:

$$\beta_p = \begin{cases} \beta_{pr,ES} & \text{falls } |\underline{p}_\mu^{(r)}(k_r)| > \overline{|\underline{p}^{(r)}(k_r - 1)|} \text{ und } k_r \notin K_{GS} \\ \beta_{pr,GS} & \text{falls } |\underline{p}_\mu^{(r)}(k_r)| > \overline{|\underline{p}^{(r)}(k_r - 1)|} \text{ und } k_r \in K_{GS} \\ \beta_{pf,ES} & \text{falls } |\underline{p}_\mu^{(r)}(k_r)| \leq \overline{|\underline{p}^{(r)}(k_r - 1)|} \text{ und } k_r \notin K_{GS} \\ \beta_{pf,GS} & \text{sonst} \end{cases}$$

$$\text{mit } 0 < \beta_{pr,GS} < \beta_{pf,GS} < 1 \text{ und } 0 < \beta_{pr,ES} < \beta_{pf,ES} < 1. \quad (3.20)$$

Mit K_{GS} sollen dabei die Zeitpunkte, in welchen der oben beschriebene Detektor Gegensprechen erkennt, bezeichnet werden. Die Menge $K_{ES,FT}$ bezeichnet die Zeitpunkte, in welchen das Korrelationsmaß Einzelsprechen des fernen Teilnehmers erkennt.

Vergleiche zwischen diesen Näherungen und der exakten Berechnung nach Gleichung 3.10 ergaben Abweichungen bei Sprachanregung von weniger als 2 dB. Für die Verwendung innerhalb der Schrittweitensteuerung reicht dies aus, somit wurde dieses Schätzverfahren für den Leistungsübertragungsfaktor verwendet.

In **Fig. 9** ist der geschätzte Leistungsübertragungsfaktor im ersten Band $p_{1,\log}^{(r)}(k_r)$ dargestellt. Seine Schätzung wird in den Bereichen B_1 , C und B_2 nicht erneuert, da hier vom Korrelationsmaß keine Freigaben geliefert werden. Im Vergleich mit **Fig. 8** ist eine gute Übereinstimmung des Soll- und des Schätzwertes zu erkennen. Als Sollwert ist hierbei die Leistungsdifferenz zwischen Anregung und Fehler zu sehen. Sowohl der Verlauf als auch der auf **Fig. 8** zu erkennende Endwert von etwa 26–30 dB wird in der Schätzung gut nachgebildet.

Aus den bisher berechneten Größen können die Schrittweiten $\alpha_{\mu}^{(r)}(k_r)$ in den einzelnen Bändern gemäß

$$\alpha_{\mu}^{(r)}(k_r) = \begin{cases} \hat{\alpha}_{\mu}^{(r)}(k_r) & , falls \quad |x_{\mu}^{(r)}(k_r)| > |x|_{\min, \mu} \\ 0 & , sonst \end{cases} \quad (3.21)$$

mit

$$\hat{\alpha}_{\mu}^{(r)}(k_r) = LIN \left\{ LOG \left\{ |x_{\mu}^{(r)}(k_r)| \right\} + p_{\mu, \log}^{(r)}(k_r) - LOG \left\{ |e_{\mu}^{(r)}(k_r)| \right\} \right\} \quad (3.22)$$

bestimmt werden. Mit $LIN \{ \dots \}$ ist dabei die Linearisierung bezeichnet. Falls die Anregungsleistung eine Grenze $|x|_{\min, \mu}$ unterschreitet, wird davon ausgegangen, daß die Anregung lediglich aus Hintergrundgeräusch besteht und die Adaption wird angehalten.

In **Fig. 10** ist die Schrittweite im ersten Teilband logarithmisch dargestellt. In Phasen des Einzelsprechens des fernen Teilnehmers (A_1 und A_2) ist die Schrittweite etwa 1 – in Phasen des Einzelsprechens des lokalen Teilnehmers (B_1 und B_2) kann aus **Fig. 8** eine Differenz von gestörter zu ungestörter Fehlerleistung von etwa 26 bis 30 dB ermittelt werden. Die Schrittweite liegt demnach auch in den Gegensprechphasen im erwarteten Bereich (ca. -27 dB).

Für die oben vorgestellte Schrittweitensteuerung wird eine Schätzung des Leistungsübertragungsfaktors benötigt. Diese Schätzung sollte nur bei Einzelsprechen des fernen Teilnehmers erneuert werden. In Gleichung 3.19 wurde aus diesem Grund die Menge $K_{ES, FT}$ eingeführt, welche die Zeitpunkte beinhalten soll, in denen das gewünschte Einzelsprechen vorliegt. Durch die starke rekursive Glättung führen kurzzeitige Fehlentscheidungen bei der Auswahl der Zeitpunkte zu keinen großen Fehlschätzungen der Übertragungsfaktoren.

Der angestrebte Detektor sollte zwischen Einzelsprechen und Gegensprechen unabhängig von Raumänderungen und auch unabhängig von der Leistung der Eingangssignale entscheiden können. Es wird ein Korrelationsmaß verwendet – ein Detektor, welcher die obigen Anforderungen erfüllt. Hierbei wird die Kreuzkorrelation zwischen dem Lautsprecher-signal und dem Mikrophonsignal in einer normierten Form ausgewertet.

Für die Auswertung werden die beiden Signale mit Schätzfenstern (Rechteckfunktionen) der Länge L_1 multipliziert. Die so erhaltenen endlichen Signalfolgen werden gemäß

$$\hat{\xi}(k) = \frac{\left| \sum_{n=0}^{L_1-1} x(k-n)y(k-n) \right|}{\sum_{n=0}^{L_1-1} |x(k-n)y(k-n)|} \quad (3.23)$$

ausgewertet. Bei stark korrelierten Signalen wird ein Maximum der oben beschriebenen Auswertung erreicht, wenn die Schätzfenster gerade um die Laufzeit des LRM-Systems zueinander verschoben sind. Da diese Laufzeit unbekannt und auch veränderlich ist (z. B. durch Verschieben des Lautsprechers oder des Mikrophons), wird das Maximum aus einer Folge von L_2 Auswertungen weiterverarbeitet. Die einzelnen Auswertungen verwenden dann ein um l Takte verzögertes Anregungssignal $x(k-l)$. Die Bestimmungsgleichung erweitert sich zu:

$$\begin{aligned} \tilde{\xi}(k) &= Max \left\{ \hat{\xi}(k, l) \right\} \\ &= Max \left\{ \frac{\left| \sum_{n=0}^{L_1-1} x(k-n-l)y(k-n) \right|}{\sum_{n=0}^{L_1-1} |x(k-n-l)y(k-n)|} \right\} \\ & \quad \text{mit } l \in \{0 \dots L_2 - 1\}. \end{aligned} \quad (3.24)$$

Die Zähler und Nenner der obigen Gleichung müssen dabei in Doppelwort-Genauigkeit (32 Bit) ausgewertet werden. Um den Rechenaufwand zu verringern, werden die einzelnen Korrelationsmaße $\hat{\xi}(k, l)$ rekursiv berechnet:

$$\hat{\xi}(k,l) = \frac{|Z(k,l)|}{N(k,l)}$$

$$= \frac{|Z(k-1,l) + x(k-l)y(k) - x(k-l-L_1+1)y(k-L_1+1)|}{N(k-1,l) + |x(k-l)y(k)| - |x(k-l-L_1+1)y(k-L_1+1)|}$$

(3.25)

Eine Freigabe wird dann gesetzt, wenn das Maximum aus den bestimmten Korrelationsmaßen größer als ein Grenzwert ξ_0 ist. Um eine Division von zwei 32-Bit-Werten zu vermeiden, wird der Grenzwert ξ_0 durch eine endliche Summe aus nichtpositiven Zweierpotenzen

$$\xi_0 \approx \sum_{n=0}^{N_\xi} a_n 2^{-n}$$

mit $a_n \in \{0,1\}$

(3.26)

angenähert. Der Schwellwertvergleich kann dann auf eine Summation von rechtsverschobenen Nennerwerten und einen Vergleich zurückgeführt werden:

$$\sum_{n=0}^{N_\xi} a_n 2^{-n} N(k,l) \lessdot Z(k,l).$$

(3.27)

Um den Rechenaufwand weiter zu reduzieren, wurden die Auswertungen nur im leistungsstärksten, ersten Teilband und dort auch nur mit den Realteilen der komplexwertigen Signale durchgeführt. In diesem Band ist bei Sprachanregung mit dem größten Signal-Geräusch-Abstand zu rechnen, was die Zuverlässigkeit der Detektorergebnisse verbessern sollte. Durch diese Maßnahme werden durch die Unterabtastung die Berechnungen nur alle r Abtastakte durchgeführt werden. Der Zeitpunkt k_r wird dann in die Menge $K_{ES,FT}$ aufgenommen, falls einer der L_2 Vergleiche ein Korrelationsmaß größer als ξ_0 ergibt.

Entsprechend der ITU-Empfehlung G. 167 kann die durch die Freisprecheinrichtung zu erbringende Echodämpfung in Gegensprechsituationen um 15 dB verringert werden. Aus diesem Grund wurde ein Gegensprechdetektor gemäß den folgenden Überlegungen entwickelt. Gleichzeitig kann dieser Detektor dazu verwendet werden, die Schätzungen in der Schrittweitensteuerung bei auftretendem Gegensprechen "vorsichtiger" einzustellen.

Die Detektion von Gegensprechen wird in zwei Schritten durchgeführt. In einer ersten Stufe wird überprüft, ob der ferne Sprecher aktiv ist. Hierzu wird zum einen das betragsglättete Anregungssignal des fernen Teilnehmers mit einer Schwelle $|x|_1$ verglichen – zum anderen wird überprüft, ob der Pegelwaagenalgorithmus Anregung des fernen Teilnehmers erkannt hat. Der zweite Vergleich ist immer dann notwendig, wenn die Pegelwaage große Dämpfungswerte einbringt (z. B. nach Raumänderungen). In solchen Situationen kann der Empfangspfad stark bedämpft sein. Hier würde der Vergleich mit dem geglätteten Eingangssignal kein zuverlässiges Ergebnis liefern. Anregung des fernen Teilnehmers ($A_{fe} = 1$) wird demnach immer dann angenommen, wenn entweder der Leistungsvergleich oder der Pegelwaagendetektor (Variable SR = 1) dies erkennen:

$$A_{fe} = \begin{cases} 1, & \text{falls } \overline{|x(k)|} > |x|_1 \text{ oder } SR = 1 \\ 0, & \text{sonst} \end{cases}$$

(3.28)

Das betragsglättete Anregungssignal wird dabei analog zu den in der Schrittweitensteuerung beschriebenen rekursiven, nichtlinearen Glättungen berechnet. Zu beachten ist hier allerdings, daß durch die höhere Abtastrate größere Zeitkonstanten verwendet werden müssen und dadurch Grenzyklen auftreten können. Eine Rechnung in Doppelwort-Genauigkeit (32 Bit) ist deshalb erforderlich:

$$\overline{|x(k)|} = \beta_{xg} \overline{|x(k-N)|} + (1 - \beta_{xg}) \overline{|x(k-1)|}$$

(3.29)

Die Zeitkonstante β_{xg} wird dabei wie folgt gewählt:

$$\beta_{\chi g} = \begin{cases} \beta_{\chi gr} & , falls |\chi(k-N)| > |\chi(k-1)| \\ \beta_{\chi gs} & , sonst \end{cases} \quad (3.30)$$

mit $0 < \beta_{\chi gr} < \beta_{\chi gs} < 1$.

Die Verzögerung von N Takten wurde eingeführt, um bei den Vergleichen in der zweiten Detektorstufe die Laufzeit des Analyse-Synthese-Systems wieder auszugleichen. Es ist hierzu kein zusätzlicher Speicher notwendig, da das Analysefilter ohnehin die letzten N Signalwerte des Eingangssignals speichert.

In einer zweiten Stufe wird festgestellt, ob auch der lokale Gesprächsteilnehmer aktiv ist. Hierzu wird ein Vergleich zwischen der Leistung des geschätzten, ungestörten Fehlers und des meßbaren, gestörten Fehlers durchgeführt. Die Leistungsschätzungen werden wieder auf Betragsglättungen bzw. die Bestimmung eines Leistungsübertragungsfaktors zurückgeführt. Die Glättung des Fehlersignals wird gemäß

$$|e(k)| = \beta_g |e(k)| + (1 - \beta_g) |e(k-1)| \quad (3.31)$$

durchgeführt. Die Zeitkonstante β_g wird wie folgt gewählt:

$$\beta_{eg} = \begin{cases} \beta_{egr} & , falls |e(k)| > |e(k-1)| \\ \beta_{egf} & , sonst \end{cases} \quad (3.32)$$

mit $\beta_{egr} = \beta_{\chi gr}$ und $\beta_{egf} = \beta_{\chi gs}$.

Für die Schätzung der ungestörten Fehlerleistung wird ein (Gesamtband-) Leistungsübertragungsfaktor $p_{EK}(k)$ bestimmt:

$$p_{EK}(k) = \text{LOG} \{ |e(k)| \} - \text{LOG} \{ |x(k)| \}. \quad (3.33)$$

Um die Varianz der Schätzung zu verbessern, wird auch diese Größe rekursiv geglättet. Da die Bestimmung des Übertragungsfaktors lediglich aus geglätteten Größen besteht, wird sie nur unterabgetastet ausgeführt:

$$\overline{p_{EK}^{(r)}(k_r)} = \begin{cases} \beta_{pg} p_{EK}(rk_r) + (1 - \beta_{pg}) \overline{p_{EK}^{(r)}(k_r - 1)} & \text{für } k_r \in K_{ES,FT} \\ \overline{p_{EK}^{(r)}(k_r - 1)} & \text{sonst} \end{cases}$$

mit $0 < \beta_{pg} < 1$.

(3.34)

Zur Detektion der Anregung des lokalen Teilnehmers ($A_{lo} = 1$) wird die Differenz aus der gemessenen und der geschätzten Fehlerleistung bestimmt. Um Fehlentscheidung zu vermeiden wurde eine zusätzliche Sicherheitsschwelle p_{GS} eingeführt. Der Detektor erkennt Anregung des lokalen Teilnehmers, wenn die gemessene Fehlerleistung um mindestens p_{GS} dB größer ist als die aus der Anregungsleistung und dem Leistungsübertragungsfaktor geschätzte Fehlerleistung. Auch dieser Vergleich wird unterabgetastet ausgeführt:

$$A_{lo(r)} = \begin{cases} 1 & , falls \text{LOG} \{ |e(rk_r)| \} - p_{GS} > \text{LOG} \{ |x(rk_r)| \} + \overline{p_{EK}^{(r)}(k_r)} \\ 0 & , sonst \end{cases}$$

(3.35)

Der Detektor erkennt Gegensprechen, wenn die UND-Verknüpfung der Variablen A_{fe} und A_{lo} den Wert eins ergibt. In diesen Fällen kann die Restdämpfung, welche durch die Pegelwaage eingebracht wird, um $p_{GSmax} = 15$ dB verringert werden. Die Verringerung der Dämpfungsanforderung erfolgt tiefpaßgeglättet. Die Zeitkonstante für die steigende Flanke β_{Gsr} sollte möglichst klein sein, um den Beginn einer Sprachpassage nicht abzuschneiden. Die Zeitkonstante für die fallende Flanke β_{Gsf} sollte größer als die Anstiegskonstante gewählt werden, damit die Dämpfungsabsenkung $\overline{p_{GS}^{(r)}(k_r)}$ in kurzen Sprachpausen nicht vollständig zurückgenommen wird. In **Fig. 11** ist dieser Zusammenhang dargestellt. Die geglättete Dämpfungsabsenkung wird wie folgt bestimmt:

$$\overline{p_{GS}^{(r)}(k_r)} = \begin{cases} \beta_{GSr} p_{GS\max} + (1 - \beta_{GSr}) \overline{p_{GS}^{(r)}(k_r - 1)}, & \text{falls } A_{fe} = A_{lo} = 1 \\ (1 - \beta_{GSr}) \overline{p_{GS}^{(r)}(k_r - 1)}, & \text{sonst} \end{cases} \quad (3.36)$$

5

Der Zeitpunkt k_r wird in die Menge K_{gs} aufgenommen, falls die Dämpfungsabsenkung über einem vorbestimmten Wert liegt. Ein beispielhafter Verlauf der Dämpfungsabsenkung ist in **Fig. 11** dargestellt.

Die Gesamtdämpfung der Pegelwaage, welche durch die ITU-T-Empfehlung G. 167 vorgeschrieben ist, kann um die Dämpfung des Gesamtsystems aus Raum und Echokompensator abgesenkt werden. Selbst im Falle abgeschalteter Echo-

10
kompensation erfolgt durch die beschriebene Steuerung eine Schätzung des Übertragungsfaktors der akustischen Strecke vom Lautsprecher zum Mikrofon einschließlich der analogen Verstärkungen. Hierdurch kann auf unterschiedliche Lautsprecher- bzw. verschiedene (analoge) Mikrofonverstärkungen reagiert und die Gesamtdämpfung entsprechend den geforderten Werten (digital) angepaßt werden. Im Gegensprechfall kann die Gesamtdämpfung ebenfalls gemäß der ITU-T-Empfehlung G. 167 auf einen geringeren Wert gesetzt werden. Auch hierfür wurde ein Detektor und eine entsprechende Übergabegröße vorgestellt bzw. definiert. Die Pegelwaagengesamtdämpfung $D_{PW}(k)$ wird damit (zunächst noch ohne Berücksichtigung der Postfilterung) nach folgendem Verfahren gesteuert:

15

$$D_{PW}(k) = D_0 - D_{EK}(k) - D_{GS}(k). \quad (3.37)$$

20

Alle Größen der obigen Gleichung liegen entsprechend den Anforderungen des ARCOFI-Pegelwaagen-Verfahrens in logarithmischer Form vor. D_0 ist dabei die geforderte Maximaldämpfung (z. B. 45 dB). Die Dämpfung des Echokompensators $D_{EK}(k)$ wird durch die Berechnungsform

$$D_{EK}(k) = \begin{cases} \overline{p_{EK}^{(r)}(\frac{k}{r})} & \text{falls } k = ir \\ D_{EK}(k-1) & \text{sonst} \end{cases} \quad (3.38)$$

25

mit $i \in \mathbb{Z}$

$$(3.39)$$

30

bestimmt. Analog dazu kann die Gegensprechabsenkung $D_{GS}(k)$ mit

$$D_{GS}(k) = \begin{cases} \overline{p_{GS}^{(r)}(\frac{k}{r})} & \text{falls } k = ir \\ D_{EK}(k-1) & \text{sonst} \end{cases} \quad (3.40)$$

35

mit $i \in \mathbb{Z}$

$$(3.41)$$

angegeben werden.

40

In der Echtzeitrealisierung des Echokompensationsverfahrens zeigt sich, daß die adaptiven Filter den Anteil des fernen Sprechers niemals vollständig aus dem Mikrophonsignal herausrechnen können. Dies kann viele verschiedene Ursachen haben, drei davon sind hier exemplarisch angeführt:

- a) Die Raumimpulsantworten sind im allgemeinen länger als die Echokompensatoren, wodurch ein Restfehler übrig bleibt.
- b) Die Festkommaarithmetik des verwendeten DSP's wirkt sich begrenzend auf den Endabgleich der Filter aus.
- c) Bei Raumänderungen führt der NLMS-Algorithmus die adaptiven Filter nur mit einer endlichen Geschwindigkeit nach – bis zum erneuten Erreichen des Endabgleichs sind Echos wieder stärker wahrnehmbar.

50

Das Fehlersignal $e(k)$ enthält somit neben dem Anteil des lokalen Sprechers $n(k)$ auch noch den nicht kompensierten Anteil des fernen Sprechers, der bereits in den vorherigen Teilen dieser Beschreibung als "ungestörter" Fehler $\epsilon(k)$ bezeichnet wurde. Für den fernen Teilnehmer ist das Signal $n(k)$ der Nutzanteil des Signals $e(k)$ – das Signal $\epsilon(k)$ ist aus dieser Sicht die Störung.

Im folgenden wird gezeigt, wie eine Nachfilterung des Signals $e(k)$ – zur Dämpfung der "Störung" $\epsilon(k)$ – basierend auf einem Wiener-Filter-Ansatz mit der Schrittweitensteuerung für die Teilbandechokompensatoren verknüpft werden kann. Hierzu wird ein Transversalfilter der Ordnung $M - 1$ im Anschluß an die Synthesefilterung eingefügt. Der Parameter M ist dabei gleichzeitig die Bandanzahl der Filterbank. Die Koeffizienten werden in der Teilbandebene bestimmt und mit einer inversen DFT in den Zeitbereich transformiert. Die Koeffizientenbestimmung ist durch mehrere Glättungen mit einer Trägheit und damit einer Laufzeit behaftet. Durch die zwischen der Koeffizientenbestimmung und -verwendung liegende, maximalphasig entworfene Synthesefilterung kann diese Laufzeit zumindest zum Teil wieder ausgeglichen werden. Die Nachfilterung erfolgt hierbei im Zeitbereich und frequenzselektiv.

55

Bei der Herleitung ergeben sich einfache Steuergrößen, mit denen der "Einfluß" des Wiener-Filters abhängig von der Kompensationsleistung der adaptiven Filter gesteuert werden kann. Auch die durch diese Maßnahme eingefügte Dämpfung kann mit geringem Aufwand geschätzt und der Pegelwaage "mitgeteilt" werden.

60

Im folgenden wird sich zeigen, daß die Bestimmung der Koeffizienten des Wiener-Filters sich auf die Berechnung von $M/2 + 1$ Subtraktionen, einer (vereinfachten) inversen Fourier-Transformation der Länge M und einigen rekursiven Glättungen zurückführen läßt. Sowohl die Subtraktionen als auch die inverse FFT und die Glättungen sind dabei nur alle r

65

Abtastwerte auszuführen. Der Berechnungsaufwand ist damit im Vergleich zu den übrigen Komponenten der Freisprech-einrichtung sehr gering!

Gemäß Fig. 12 wird das Filter $\bar{g}(k)$ **30** hinter der Synthese plaziert. Die Ordnung des Filters betrage $M - 1$, es müssen also M Koeffizienten eingestellt werden. Das Filter **30** soll gemäß dem Wiener-Ansatz das "gestörte" Signal $e(k)$ optimal von der "Störung" $\varepsilon(k)$ befreien. Der Frequenzgang eines solchen Filters lautet:

$$G_{opt}(\Omega) = \frac{S_{en}(\Omega)}{S_{ee}(\Omega)}$$

Für das Signal $e(k)$ gilt:

$$e(k) = \varepsilon(k) + n(k).$$

Der Filterfrequenzgang kann damit zu

$$G_{opt}(\Omega) = \frac{S_{en}(\Omega) + S_{nn}(\Omega)}{S_{ee}(\Omega) + S_{en}(\Omega) + S_{ne}(\Omega) + S_{nn}(\Omega)}$$

umgeformt werden. Die Signale des fernen und des lokalen Teilnehmers ($n(k)$ bzw. $\varepsilon(k)$) werden als unkorreliert vorausgesetzt. Bedingt durch die Hochpaßfilterung des Leitungseingangs- und des Mikrophonsignals wird weiter Mittelwertfreiheit der Signale $n(k)$ und $\varepsilon(k)$ angenommen. Der Frequenzgang vereinfacht sich dadurch zu:

$$\begin{aligned} G_{opt}(\Omega) &= \frac{S_{nn}(\Omega)}{S_{ee}(\Omega) + S_{nn}(\Omega)} \\ &= 1 - \frac{S_{\varepsilon\varepsilon}(\Omega)}{S_{\varepsilon\varepsilon}(\Omega) + S_{nn}(\Omega)} \\ &= 1 - \frac{S_{\varepsilon\varepsilon}(\Omega)}{S_{ee}(\Omega)} \end{aligned}$$

Da das Filter $\bar{g}(k)$ die Ordnung $M - 1$ haben und durch inverse Fourier-Transformation aus dem Frequenzgang $G_{opt}(\Omega)$ bestimmt werden soll, müssen M Stützstellen des Frequenzgangs bestimmt werden. Für die Frequenzen

$$\Omega_{\mu} = \mu \frac{2\pi}{M} \text{ mit } \mu \in \{0 \dots M - 1\}$$

ergibt sich:

$$G_{opt}(\Omega_{\mu}) = 1 - \frac{S_{\varepsilon\varepsilon}(\Omega_{\mu})}{S_{ee}(\Omega_{\mu})}.$$

Die Frequenzen Ω_{μ} stellen aber neben den Stützstellen im Frequenzbereich auch gleichzeitig die Bandmitten der zuvor beschriebenen Bandpässe bei der Teilbandzerlegung dar. Bei der Schätzung der Größe $S_{\varepsilon\varepsilon}(\Omega_{\mu})/S_{ee}(\Omega_{\mu})$ kann somit auf entsprechende Größen in den einzelnen Teilbändern zurückgegriffen werden. $G_{opt}(\Omega_{\mu})$ kann durch

$$G_{opt}(\Omega_{\mu}) = 1 - \frac{E\left\{ \left| \underline{\varepsilon}_{\mu}(k) \right|^2 \right\}}{E\left\{ \left| \underline{e}_{\mu}(k) \right|^2 \right\}}$$

angenähert werden. Da bei der Herleitung des Wiener-Filters Stationarität der Eingangssignale vorausgesetzt wurde, dies aber bei Sprache nur für kurze Passagen angenommen werden kann, sollten die Leistungsdichtespektren durch entsprechende Kurzzeitleistungsschätzwerte im jeweiligen Frequenzbereich ersetzt werden. Somit gelten für die Schätzung der Quotienten

$$\frac{S_{\varepsilon\varepsilon}(\mu \frac{2\pi}{M}, k)}{S_{ee}(\mu \frac{2\pi}{M}, k)} \text{ mit } \mu \in \{0 \dots M - 1\}$$

die gleichen Voraussetzungen wie für die Schätzung der Schrittweiten in den jeweiligen Bändern. Die DFT-Transformierte des Filters $\bar{g}(k)$ könnte daher gemäß

$$\vec{G}^{(r)}(k) = 1 - \vec{\alpha}^{(r)}(k)$$

bestimmt werden. Die hochgestellten "(r)" sollen dabei auf die Unterabtastebene hinweisen. $\vec{G}^{(r)}(k)$ bzw. $\vec{\alpha}_{(1)}^{(r)}(k)$ än-

den sich somit nur alle r Abtastschritte. Im bevorzugten Ausführungsbeispiel wurde r = 13 gewählt. Es wurde gezeigt, daß die komplexen Bänder nur für $\mu = 1 \dots M/2 - 1$ berechnet werden müssen – die Bänder $\mu = M/2 - 1 \dots M - 1$ können durch komplexe Konjugation ermittelt werden. Da die Schrittweiten α aber reellwertig sind, kann der Vektor $\bar{\alpha}^{(r)}(k)$ wie folgt gebildet werden

$$\bar{\alpha}^{(r)}(k) = (\alpha_o^{(r)}(k), \alpha_1^{(r)}(k), \dots, \alpha_7^{(r)}(k), 1, \alpha_7^{(r)}(k), \dots, \alpha_1^{(r)}(k))^T$$

Da die Teilbandzerlegung den Bereich des letzten Teilbandes (bei 8 kHz Abtastrate 3750 Hz – 4000 Hz) herausfiltert, soll dieser Bereich im verwendeten Wiener-Filter ebenfalls undurchlässig sein, wodurch sich die Wahl von $G_8^{(r)}(k) = 0$ bzw. $\alpha_8^{(r)}(k) = 1$ ergibt.

In der praktischen Anwendung dieses Verfahrens zeigt sich, daß ein leicht modifizierter Ansatz zu besseren Ergebnissen führt. Analog zu bekannten Verfahren der Geräuschreduktion werden die geschätzten Stützstellen des Filterfrequenzganges zeitlich geglättet, sowie mit einem sog. Überschätzungsfaktor β und einer Maximaldämpfung $G_{\min}(k)$ versehen. Die zeitliche Glättung wird auf die Schrittweiten angewendet und erfolgt mit einem IIR-Filter erster Ordnung mit zwei verschiedenen Zeitkonstanten für steigende (γ_r) und fallenden (γ_f) Flanken:

$$\overline{\alpha}_\mu^{(r)}(k) = (1 - \gamma_{r,f})\alpha_\mu^{(r)}(k) + \gamma_{r,f}\overline{\alpha}_\mu^{(r)}(k - 1)$$

mit $\mu \in \left\{0 \dots \left(\frac{M}{2} - 1\right)\right\}$.

Bei einer linearen Glättung ($\gamma_r = \gamma_f$) würde die Dämpfung bei Beginn einer Sprachpassage des fernen Teilnehmers zunächst langsam und dann immer schneller eingebracht. Am Ende der Sprachpassage würde die Dämpfung dann zunächst schnell und dann immer langsamer reduziert. Um diesen Zusammenhang zu verdeutlichen, ist in Fig. 13 ein beispielhafter Verlauf des Terms $(1 - \alpha(k))$ in einem der Teilbänder dargestellt. Zu Beginn soll eine Sprachpause des fernen Sprechers vorliegen, der Term $(1 - \alpha(k))$ ist dementsprechend gleich Eins. Mit dem Einsetzen der Sprachpassage werde die Schrittweite $\alpha(k)$ auf einen Wert nahe bei Eins gesetzt – zur Vereinfachung bleibe die Schrittweite bis zum Ende der Sprachsequenz auf diesem Wert, anschließend wird die Schrittweite wieder auf Null gesetzt. Zur Verdeutlichung der Größe der eingefügten Dämpfung (es wird vereinfachend davon ausgegangen, daß in allen Bändern der gleiche Verlauf vorliegt) sind die Stellen, an denen die Kurve mit der geglätteten Schrittweite die Werte $(1 - 1/2)$, $(1 - 1/4)$ und $(1 - 1/8)$ erreicht, gekennzeichnet. Diese Werte entsprechen dann einer Dämpfung von 6 dB, 12 dB bzw. 18 dB. Im unteren Teil der Fig. 13 ist der mit zwei unterschiedlichen Zeitkonstanten geglättete Term $\bar{\alpha}(k)$ dargestellt. Zu Beginn der Sprachpassage wird die Dämpfung hier schnell eingefügt – am Ende erfolgt eine langsamere Reduktion der eingebrachten Dämpfung.

Der in der Implementierung verwendete Vektor $\bar{\alpha}^{(r)}(k)$ setzt sich somit aus den geglätteten Schrittweiten zusammen:

$$\bar{\alpha}^{(r)}(k) = (\overline{\alpha}_o^{(r)}(k), \overline{\alpha}_1^{(r)}(k), \dots, \overline{\alpha}_7^{(r)}(k), 1, \overline{\alpha}_7^{(r)}(k), \dots, \overline{\alpha}_1^{(r)}(k))^T$$

Der Filterfrequenzgang wird dann gemäß

$$G_n^{(r)}(k) = \text{Max} \left\{ \left(1 - \beta \overline{\alpha}_n^{(r)}(k) \right), G_{\min}(k) \right\}$$

(4.1)

mit $n \in \{0 \dots (M - 1)\}$

geschätzt. Der Überschätzungsfaktor β beschleunigt bei einer Wahl größer als eins das Einbringen der Dämpfung und er vergrößert die Dämpfung. Für β wird vorzugsweise ein Wert zwischen 1.0 und 3.0 gewählt.

Durch den Parameter $G_{\min}(k)$ können die Spektralschätzwerte des Filters nach unten begrenzt werden. Wird dieser Parameter beispielsweise zu Null gewählt, so könnte durch das Filter das Ausgangssignal zu Null gesetzt werden. Wird $G_{\min}(k) = 1$ gesetzt, so erfährt das Ausgangssignal keine Änderung. Mit dem Parameter $G_{\min}(k)$ kann somit der "Einfluß" des Wiener-Filters gesteuert werden. In Echtzeitversuchen zeigte sich, daß es sinnvoll ist, die Steuerung dieses Parameters mit dem Abgleichzustand der Echokompensatoren zu verknüpfen. Zu Beginn eines Abgleichvorgangs ist die Dämpfung, welche durch die Echokompensatoren erreicht wird, noch sehr gering. Hier sollte das Wiener-Filter stark eingreifen und große Dämpfungen (z. B. bis zu 45 dB gemäß den ITU-Empfehlungen) einbringen können. Ist in dem Raum, in dem sich die Freisprecheinrichtung befindet, starkes Hintergrundgeräusch vorhanden, so werden durch das Wiener-Filter die Echos zwar unterdrückt, der ferne Teilnehmer nimmt dann aber eine Art Modulation des Hintergrundgeräusches wahr. In seinen Sprachpausen wird das Geräusch ungedämpft übertragen, während er spricht, erfährt es eine (z. B. 45 dB große) Dämpfung.

Zu Beginn eines Abgleichvorgangs sind solche "Effekte" tolerabel, zumal "herkömmliche" Verfahren wie die Pegelwaage ähnliches bewirken. Mit zunehmendem Abgleich der Kompensatoren sollte dieser Effekt aber verringert werden. Auch hier liefert die Schrittweitensteuerung eine geeignete Steuergröße – den geschätzten Leistungsübertragungsfaktor $D_{EK}(k)$. Die Einstellung des Parameter $G_{\min}(k)$ erfolgt daher gemäß:

$$G_{\min}(k) = \text{LIN} \{ \text{Max} \{ 0, (G_{\max, \log} - D_{EK}(k) - D_{GS}(k)) \} \}. \quad (4.2)$$

Mit "LIN" wird dabei die bereits in der Schrittweitensteuerung verwendete Linearisierung von logarithmischen Größen bezeichnet. Mit dem Parameter $G_{\max, \log}$ kann die maximale Einfügedämpfung (z. B. 45 dB) eingestellt werden. Dieser Festwert wird dann um die Dämpfung $D_{EK}(k)$, welche die Echokompensatoren im Mittel leisten, sowie um die Gegensprechabsenkung $D_{GS}(k)$ reduziert. Die Größen $D_{EK}(k)$ und $D_{GS}(k)$ liegen dabei in der gleichen logarithmischen Form wie die Konstante $G_{\max, \log}$ vor. Die Begrenzung der errechneten Größe auf 0 dB dient der Anpassung an die Linearisierung.

Damit sind alle Steuergrößen zur Einstellung des Wiener-Filters und die Filterkoeffizienten im Teilbandbereich bestimmt. Die so erhaltenen Spektralschätzwerte des Filters werden mit Hilfe einer inversen DFT so in den Zeitbereich transformiert, daß ein phasenlineares Filter entsteht. Hierbei kann von der Tatsache, daß die Systemfunktion sowohl reellwertig als auch symmetrisch ist, Gebrauch gemacht werden und der Aufwand der IDFT auf etwa ein Viertel reduziert werden.

Die Dämpfung $D_w(k)$ des Signals $e(k)$ durch das Wiener-Filter wird analog zur Dämpfung der Echokompensatoren und der Dämpfungsreduktion bei Gegensprechen über eine Schittstelle der Pegelwaage mitgeteilt. Die Dämpfung wird dabei durch den Mittelwert über alle zu übertragenden Frequenzbereiche angenähert:

$$D_w^{(r)}(k) = LOG\left(\frac{1}{8} \sum_{\mu=0}^7 \left(1 - \overline{\alpha_{\mu}^{(r)}}(k)\right)\right).$$

Mit "LOG" wird dabei die bereits in der Schrittweitensteuerung verwendete Normierung bzw. Logarithmierung bezeichnet. Sie sorgt für die schnittstellenspezifische Kommunikation mit der Pegelwaage. Die Division durch 8 wird durch Rechtsschieben um 3 Bit erreicht. Bevor die Dämpfung dann endgültig an die Pegelwaage übergeben wird, erfolgt eine rekursive, nichtlineare Glättung:

$$\overline{D_w^{(r)}}(k) = \beta_{r,f} \overline{D_w^{(r)}}(k-1) + (1 - \beta_{r,f}) D_w^{(r)}(k).$$

Die Verwendung von unterschiedlichen Zeitkonstanten für steigende und fallende Flanken bewirkt, daß die Schätzung "vorsichtiger" wird. Wird durch das Wiener-Filter Dämpfung eingefügt, so verringert die Pegelwaage ihre Dämpfung langsamer. Kurzzeitig wird das Fehlersignal damit mehr als die geforderten 45 dB gedämpft. Verringert das Wiener-Filter umgekehrt seine Dämpfung, fügt die Pegelwaage sehr schnell die restliche Dämpfung ein. Durch die zeitliche Verzögerung durch die Synthesefilterung kann es auch hier zu einer kurzzeitigen Gesamtdämpfung von mehr als der eingestellten Obergrenze (z. B. 45 dB) kommen.

Zur Verdeutlichung der bisherigen Überlegungen wurde die im Abschnitt der Schrittweitensteuerung beschriebene Simulation wiederholt – diesmal aber erweitert mit dem oben vorgestellten Wiener-Filter. Als Raumimpulsantwort wurde die gemessene Raumimpulsantwort eines Büroraumes mit etwa 300 ms Nachhallzeit verwendet. Als Anregungen wurde sowohl auf der fernen als auch auf der lokalen Teilnehmerseite weißes Rauschen gemäß **Fig. 14** eingespeist.

Um den Einfluß des Wiener-Filters deutlich darzustellen, wurde die Maximaldämpfung $G_{\max, \log}$ zu 60 dB gewählt. Im Bereich A_1 findet der Anfangsabgleichvorgang der Kompensatoren statt. Zu Beginn dieses Bereiches sind die Kompensatoren noch nicht abgeglichen – am Ende wurde in allen Bändern der Endabgleichzustand erreicht. Da in dieser Phase kein Gegensprechen stattfindet, sollte das Wiener-Filter die Differenz zwischen 60 dB und der Dämpfung, welche durch die Echokompensatoren erreicht wird, einfügen. Hierzu ist im Bereich A_1 der Koeffizient

$$G_1^{(r)}(k) = Max\left\{\left(1 - \overline{\beta \alpha_1^{(r)}}(k)\right), G_{\min, 1}(k)\right\}$$

im Teilband **1** (250–750 Hz bei 8 kHz Abtastrate) zusammen mit dem Anregungs- und Fehlersignal vor dem Wiener-Filter in **Fig. 15** dargestellt. Zu erkennen ist hierbei zunächst der Einschwingvorgang des Wiener-Filters. Bedingt durch die Trägheit der Tiefpaßglättungen wird die Dämpfung nicht sofort eingefügt – dieser Effekt wird durch die Transformation in den Zeitbereich und die dazwischenliegende Synthesefilterung zum Teil wieder ausgeglichen. Im Gesamtbandsignal (s. **Fig. 18**) werden dadurch immerhin schon zu Beginn der Aktivität des fernen Sprechers 25 dB Dämpfung eingefügt. Nach etwa 200 ms hat sich die Dämpfung dann bereits auf ihren Endwert von 60 dB erhöht. Mit zunehmendem Abgleichen des Kompensators verringert sich die Dämpfung durch das Wiener-Filter im Band **1** und erreicht erwartungsgemäß einen Endwert von etwa 30 dB (60 dB Maximalbegrenzung – 30 dB Echodämpfung durch den Kompensator). Da das Wiener-Filter erst nach der Synthese eingefügt wurde, können die Verläufe der Anregung, des Fehlers, der Schrittweite und des Leistungsübertragungsfaktors im Band **1** aus den **Fig. 9** und **10** entnommen werden.

Für den Fall des Einzelsprechens des fernen Gesprächsteilnehmers (Bereich A_1 und A_2) ist somit die Maximalgrenze der einzufügenden Dämpfung $G_{\min}(k)$ die bestimmende Größe. Entsprechend dem Ansatz des Filters soll das Gesamtsignal $e(k)$ von seiner Störung $\epsilon(k)$ getrennt werden. Da der lokale Teilnehmer – das Nutzsignal in $e(k)$ – jedoch nicht aktiv ist, besteht das Gesamtsignal lediglich aus der Störung. Würde die Begrenzung bei der Bestimmung der Koeffizienten $G_1^{(r)}(k)$ weggelassen, so würden diese Koeffizienten zu Null gesetzt und die Störung damit eliminiert.

In **Fig. 16** ist zur Verdeutlichung dieses Zusammenhangs die Dämpfung, welche durch das Wiener-Filter eingefügt wird, im Band **1** dargestellt. Der anfängliche Wert von etwa 60 dB wird durch die eingestellte Maximaldämpfung $G_{\max, \log}$ bestimmt. Die zu Beginn der Simulation mit Nullvektoren initialisierten Kompensatoren gleichen im Verlauf der Phase A_1 ab und reduzieren damit die Obergrenze der einzufügenden Dämpfung auf etwa 30 dB. In der nun folgenden Gegensprechphase B_1 wird diese Obergrenze durch den Gegensprechdetektor nochmals um 15 dB auf nun noch etwa 15 dB verringert. Da die Leistung des lokalen Sprechers aber deutlich über der des Restechos liegt, wird diese Grenze nicht erreicht. Gemäß dem gewählten Einstellalgorithmus wird somit in der Gegensprechphase B_1 fast keine Dämpfung eingefügt. Die bestimmende Größe in der Gegensprechphase ist das Leistungsverhältnis des Signals des lokalen Spre-

chers und des Restechos des fernen Sprechers. Die Leistung des Restechos hängt zum einen von der Anregungsleistung des fernen Teilnehmers und zum anderen vom Abgleichzustand der Kompensatoren ab. Je besser diese abgeglichen sind, um so geringer wird der Einfluß des Wiener-Filters in diesen Passagen sein.

In der folgenden Gesprächssituation C hat der lokale Teilnehmer das Wort übernommen. Die Schrittweiten werden in diesen Situationen zu Null gesetzt, wodurch das Wiener-Filter zu einer Durchschaltung wird. Die Passagen B₂ und A₂ sind analog zu den eben beschriebenen Phasen zu sehen.

Da die Schätzung der Dämpfung, welche durch das Wiener-Filter eingefügt wird, mit unterschiedlichen Zeitkonstanten ausgeführt wird, kommt es in bestimmten Phasen zu einer zu "vorsichtigen" Schätzung. Um diesen Sachverhalt zu verdeutlichen sind in Fig. 17 die Verläufe der Schätzung der Dämpfungen durch die Echokompensatoren und durch das Wiener-Filter, sowie die Absenkung im Gegensprechfall aufgetragen. Die Summe dieser drei Größen wird der Pegelwaage übergeben und ist im unteren Teil der Fig. 17 dargestellt. Diese Schätzung kann mit den wirklichen Signalverläufen der Anregung und des Fehlers im Gesamtband in Fig. 18 verglichen werden. In den Bereichen B₁ und B₂ erkennt der Gegensprechdetektor die Aktivität der beiden Teilnehmerseiten und erhöht die Dämpfungsübergabe um 15 dB. Diese Erhöhung wird mit einer kurzen Zeitkonstante eingefügt und am Ende der Gegensprechphase langsam wieder herausgenommen. Diese Maßnahme wurde zur Überbrückung kurzer Sprachpausen eingeführt. Gleichzeitig wird mit dem Einsetzen des Gegensprechens die Schrittweite reduziert und das Wiener-Filter verringert seine Dämpfung. In den Passagen ohne Anregung (Bereich C) des fernen Teilnehmers wird die Schrittweite zu Null gesetzt – das Wiener-Filter wirkt dadurch lediglich als Verzögerungsglied.

Das bisher vorgestellte Verfahren wurde für die endgültige Implementierung allerdings noch einmal leicht modifiziert – hierdurch konnte der Rechenaufwand noch einmal gesenkt werden, ohne merkliche Qualitätseinbußen zu erhalten.

Nach einer schrittweitenabhängigen Bestimmung der Filterkoeffizienten im Teilbandbereich, wurde gemäß Gleichung 4.1 eine Obergrenze der Dämpfung bestimmt. Diese Obergrenze wurde in Abhängigkeit von der bereits erreichten Dämpfung, welche durch die Leistungsübertragungsfaktoren im jeweiligen Band bzw. durch die Gegensprechdämpfung gegeben ist, bestimmt. Beide Größen wurden in der Schrittweitenberechnung lediglich in logarithmischer Darstellung berechnet und gespeichert. Um die Größen in der Begrenzungsfunktion verwenden zu können, sind demnach acht Linearisierungen notwendig. Die Bestimmung der Maximalwerte würde damit mehr Rechenleistung benötigen als die gesamte restliche Koeffizientenberechnung. Aus diesem Grund wurde für alle Bänder eine einheitliche Obergrenze eingeführt. Diese wird ebenfalls gemäß Gleichung 4.1 berechnet, allerdings mit den Gesamtbandgrößen. Der Ressourcenbedarf der so erhaltenen Nachfilterung liegt bei deutlich unter 1 MIPS bei Verwendung von 16-Bit-Festkomma-Signalprozessoren.

Bei eingeschaltetem Wiener-Filter 30 kann die Gesamtdämpfung zusätzlich um die Dämpfung des Wiener-Filters 30 abgeschwächt werden. Der Maximalhub der Pegelwaage kann damit durch

$$D_{PW}(k) = D_0 - D_{EK}(k) - D_{GS}(k) - D_W(k) \quad (4.3)$$

angegeben werden. Die Größe D_W(k) wird dabei gemäß

$$D_W(k) = \begin{cases} \overline{D_W^{(r)}(k)} & \text{falls } k = ir \\ D_{EK}(k-1) & \text{sonst} \end{cases} \quad (4.4)$$

mit $i \in \mathbb{Z}$

bestimmt.

Patentansprüche

1. Verfahren zur Verbesserung der akustischen Rückhördämpfung in Freisprecheinrichtungen mit einer Pegelwaage (22) und einer frequenzselektiven steuerbaren Echokompensation (28) mit Teilbandverarbeitung, wobei das abgehende Signal nach der frequenzselektiven Echokompensation (28) einer Nachfilterung in einem weiteren frequenzselektiven Filter (30) mit Einstellalgorithmus gemäß einem Wiener-Ansatz unterworfen wird (Wiener-Filterung), **dadurch gekennzeichnet**, daß eine einzige Steuergröße (Schrittweitenvektor $\vec{\alpha}(k)$ sowohl für die Steuerung der frequenzselektiven Echokompensation, als auch für die Steuerung des weiteren Filters (30) verwendet wird.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß mehrere unterschiedliche Abtastraten verwendet werden.
3. Verfahren nach Anspruch 1 oder Anspruch 2, dadurch gekennzeichnet, daß sowohl bei der Echokompensation (28) als auch für das weitere Filter (30) adaptive Filter verwendet werden.
4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß die Echokompensation (28) mittels einer Filterbank in Frequenzteilbändern implementiert wird.
5. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß zur Steuerung der Adaption und der Schrittweite sowohl leistungsbasierende Schätzungen als auch korrelationsbasierende Analysen verwendet werden.
6. Verfahren nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, daß zur Schrittweitenbestimmung Leistungsübertragungsfaktoren in Teilbändern geschätzt werden.
7. Verfahren nach einem der Ansprüche 1 bis 6, dadurch gekennzeichnet, daß sowohl die Echokompensation (28) als auch der weitere Filter (30) Schätzwerte für die durch sie eingebrachte Echodämpfung liefern.
8. Verfahren nach Anspruch 7, dadurch gekennzeichnet, daß die Schätzwerte für die Dämpfung zur Steuerung der Dämpfung der Pegelwaage (22) verwendet werden.
9. Verfahren nach einem der Ansprüche 1 bis 8, dadurch gekennzeichnet, daß die gleichzeitige Aktivität beider Gesprächsteilnehmer (Gegensprechen) detektiert wird.
10. Verfahren nach Anspruch 9, dadurch gekennzeichnet, daß die Gesamtdämpfung der Pegelwaage im Gegen-

sprechfall reduziert wird.

Hierzu 17 Seite(n) Zeichnungen

5

10

15

20

25

30

35

40

45

50

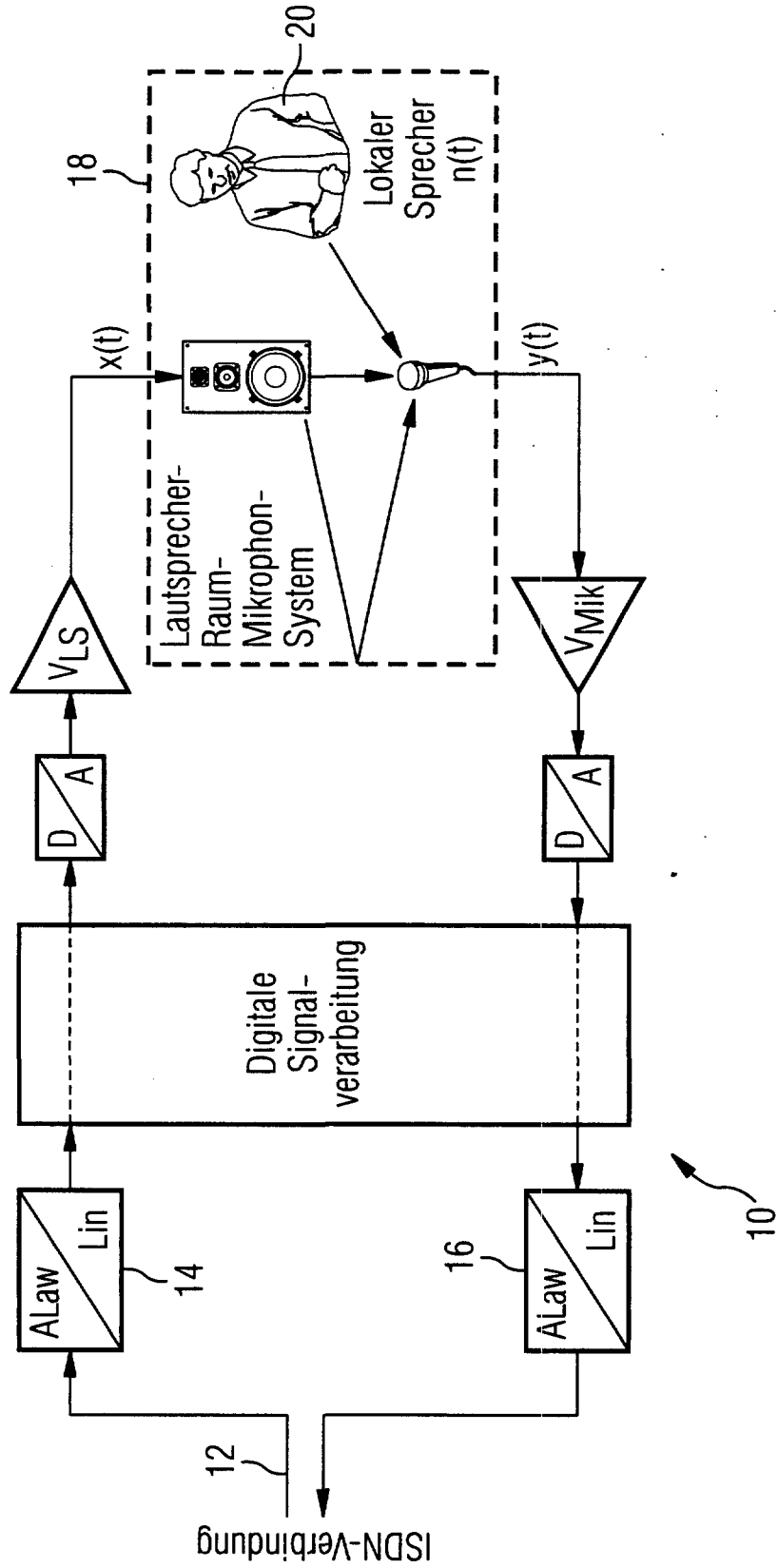
55

60

65

- Leerseite -

FIG 1



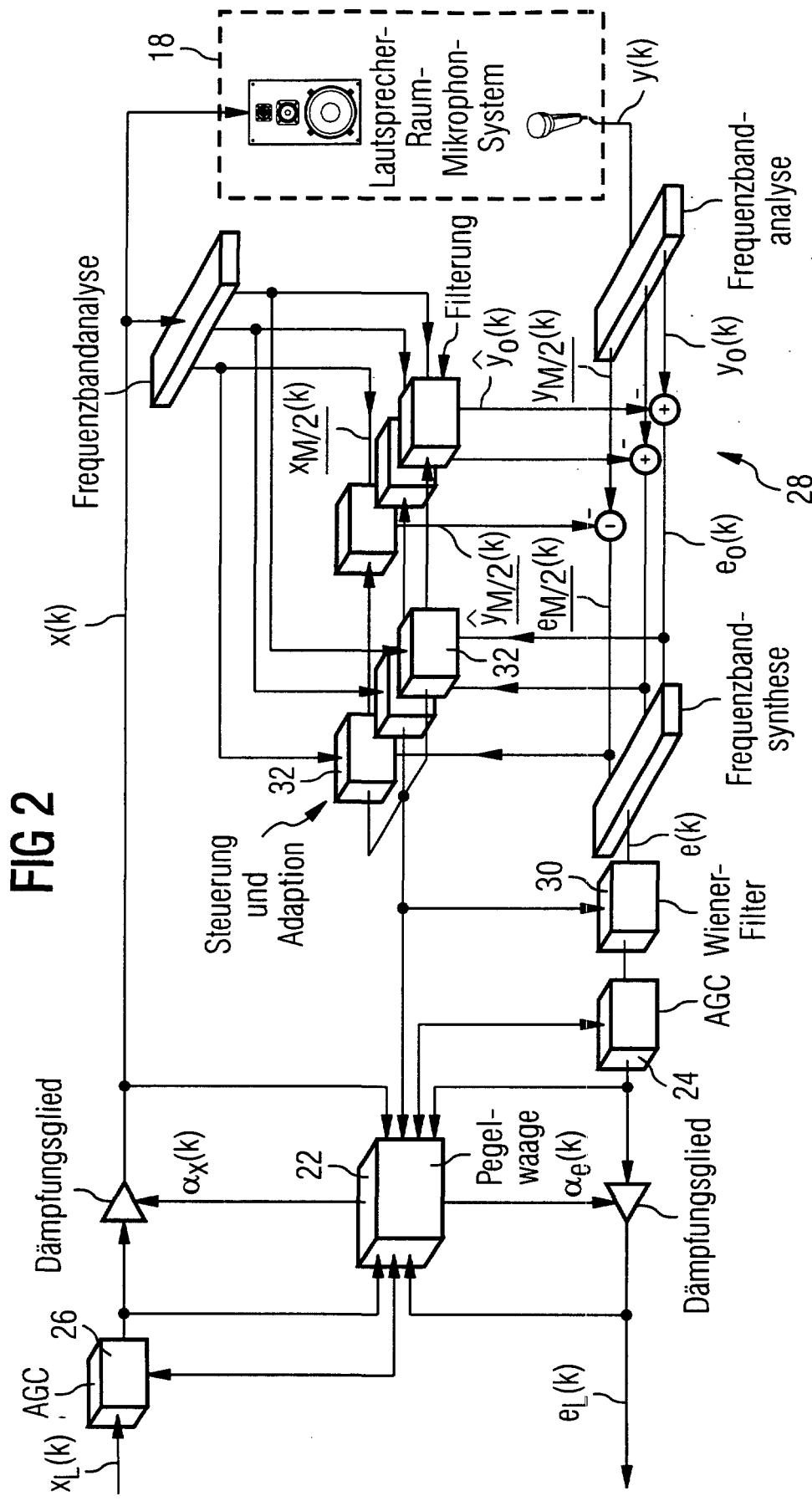


FIG 2

FIG 3

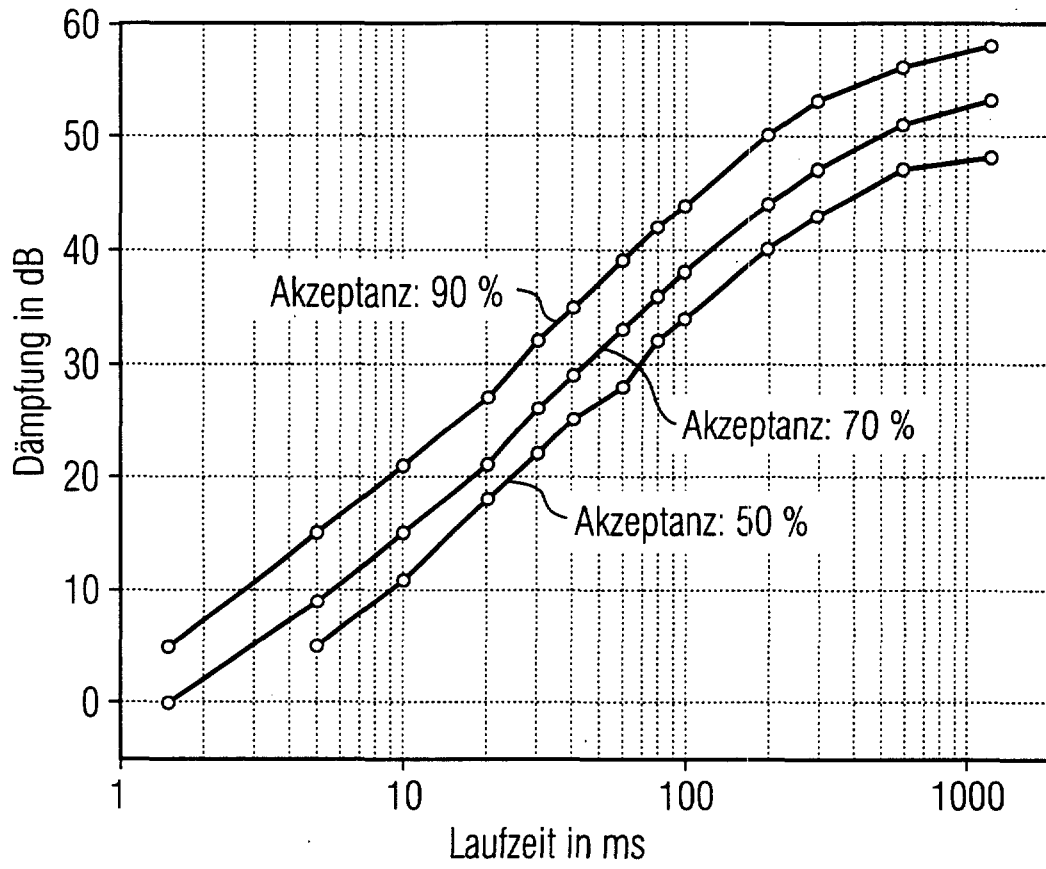


FIG 4

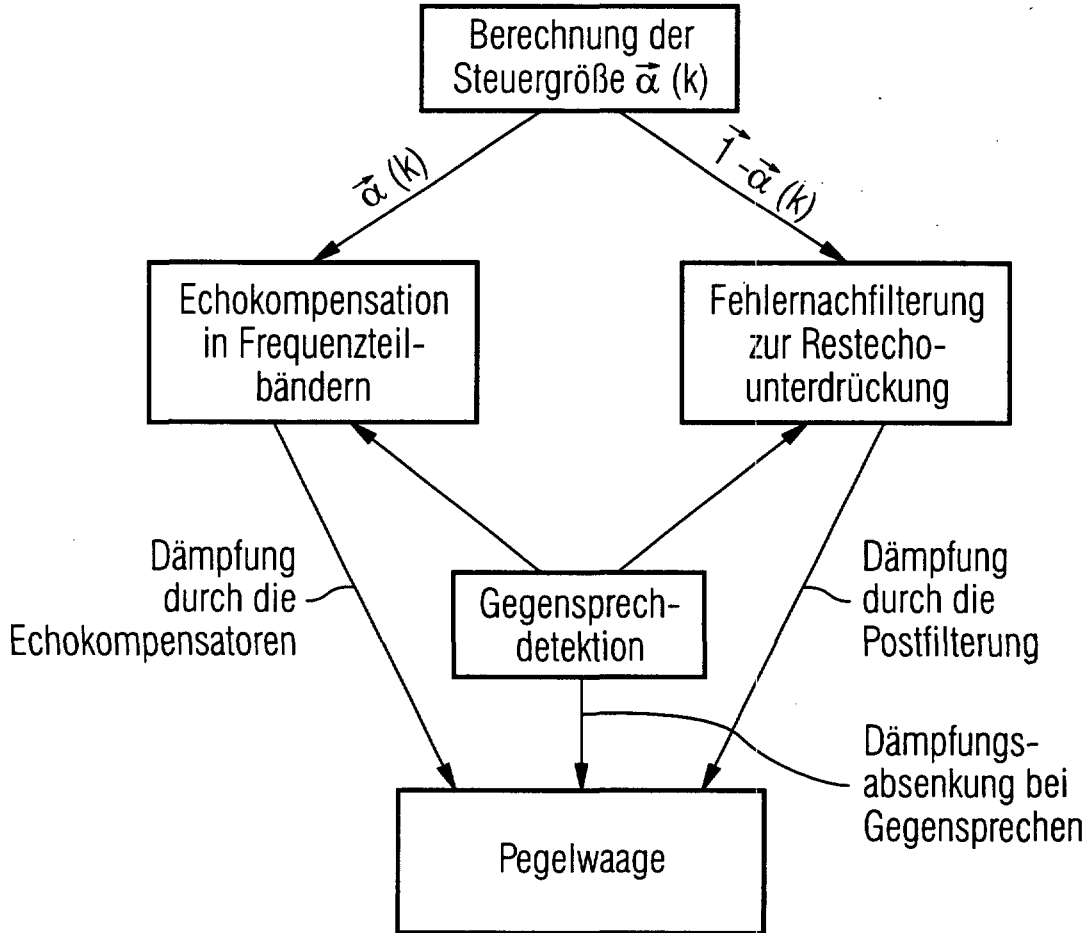


FIG 5

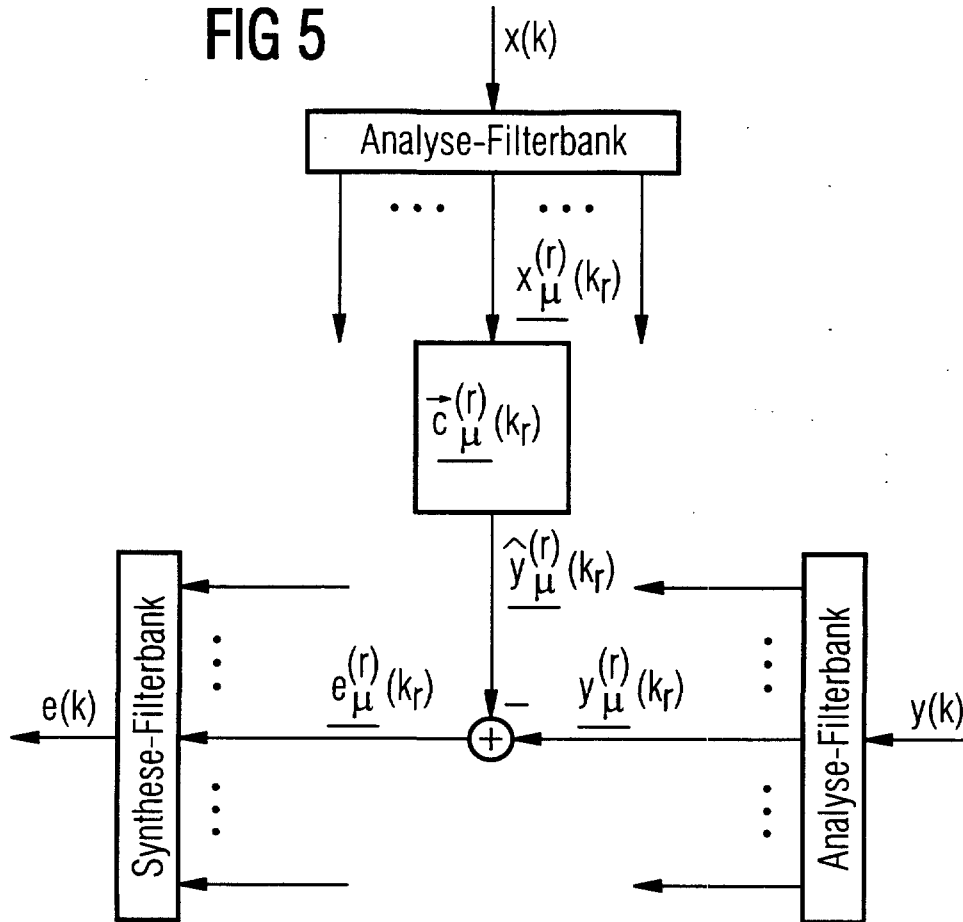
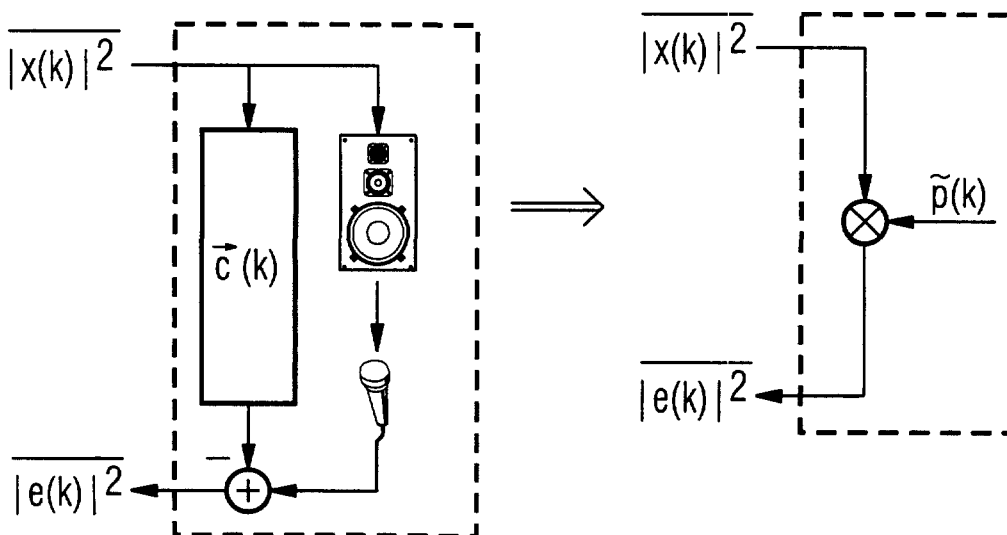
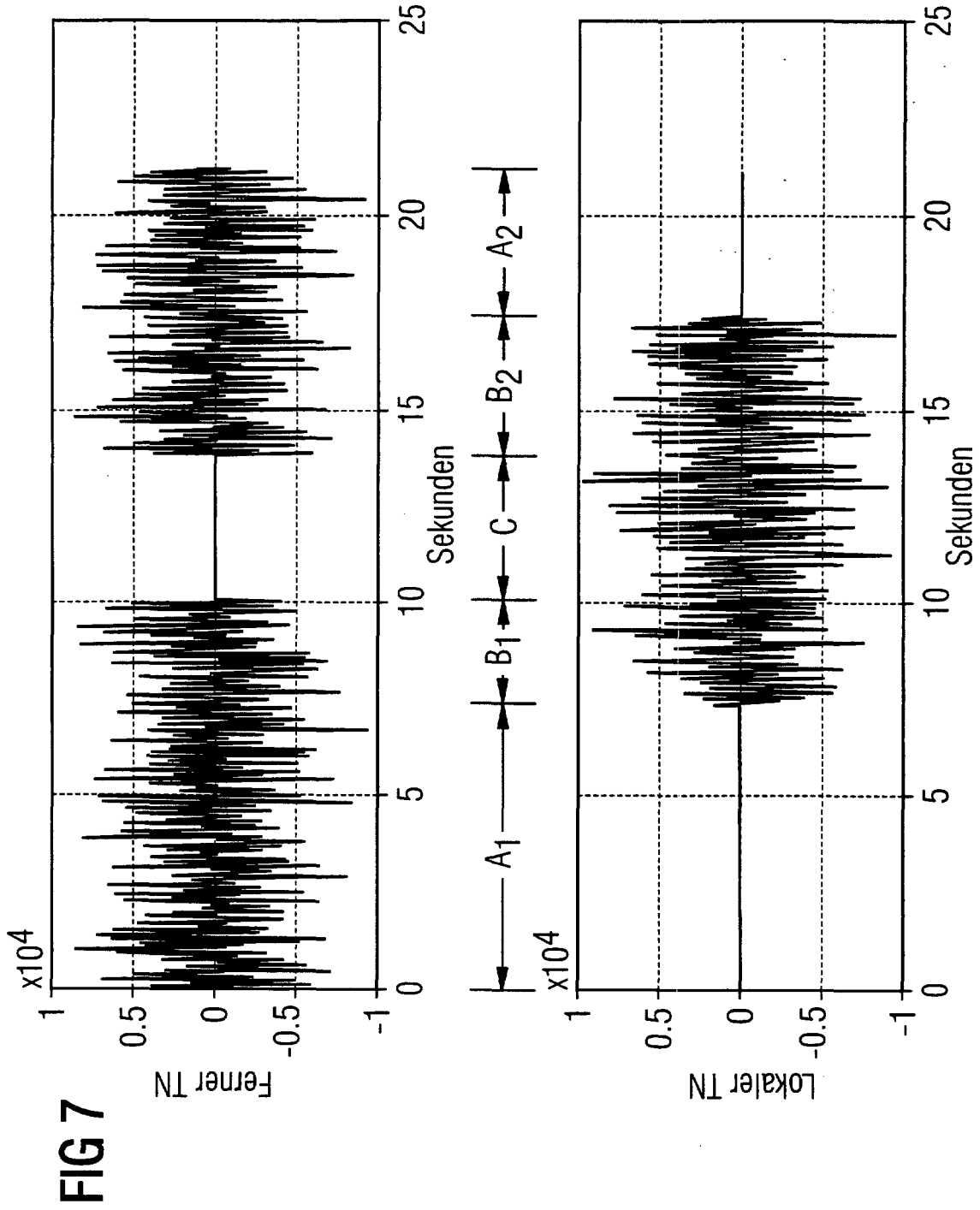


FIG 6





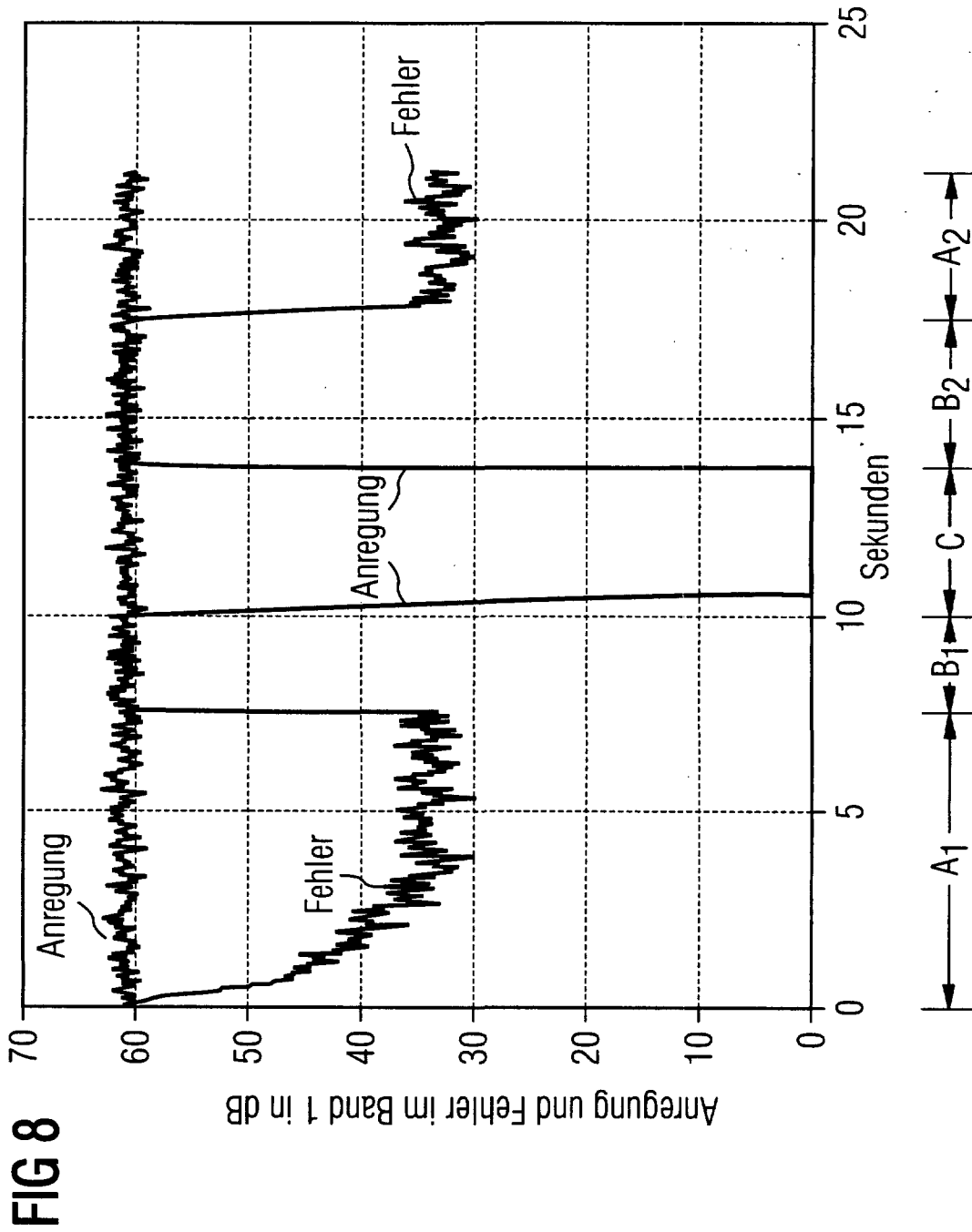


FIG 8

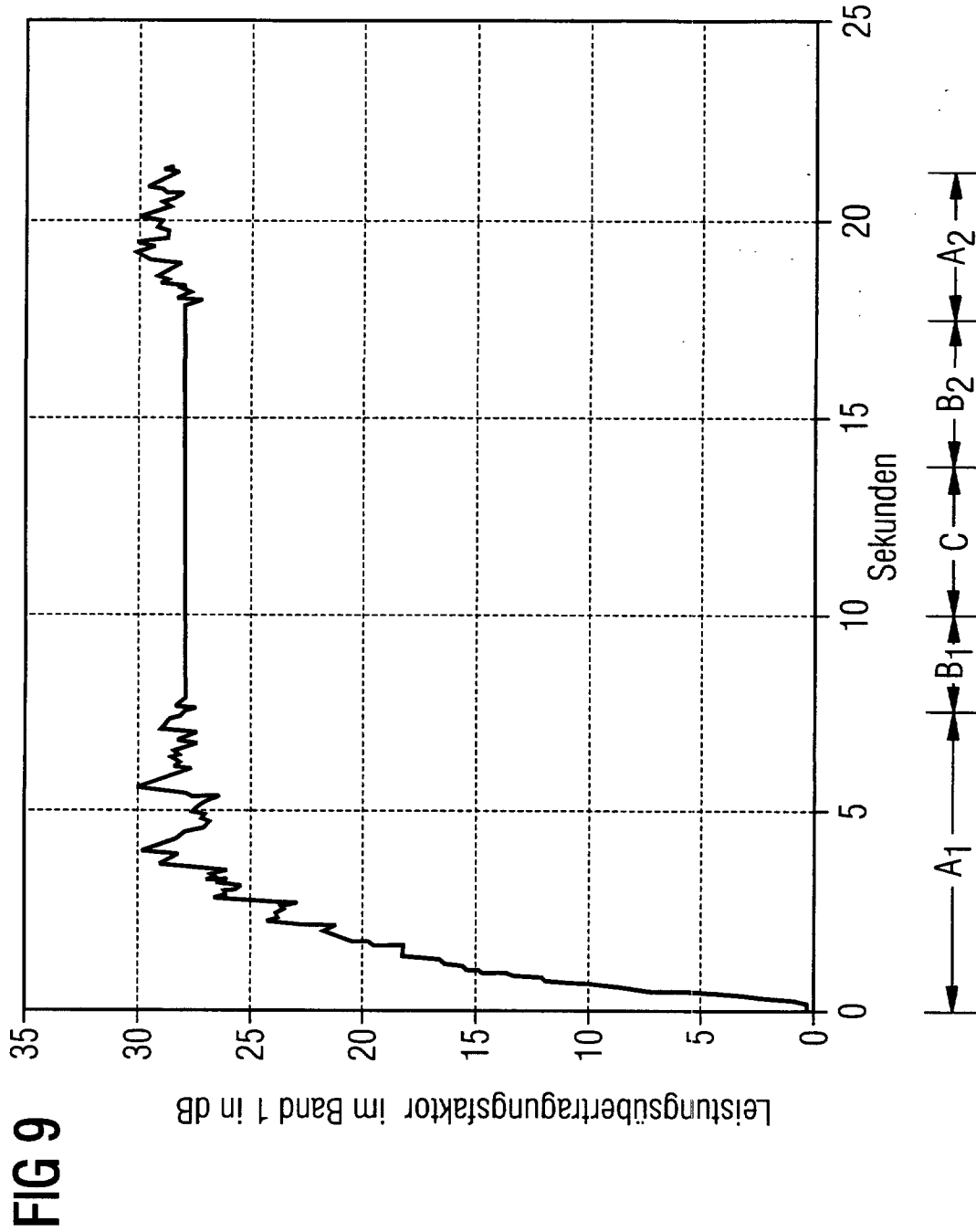


FIG 9

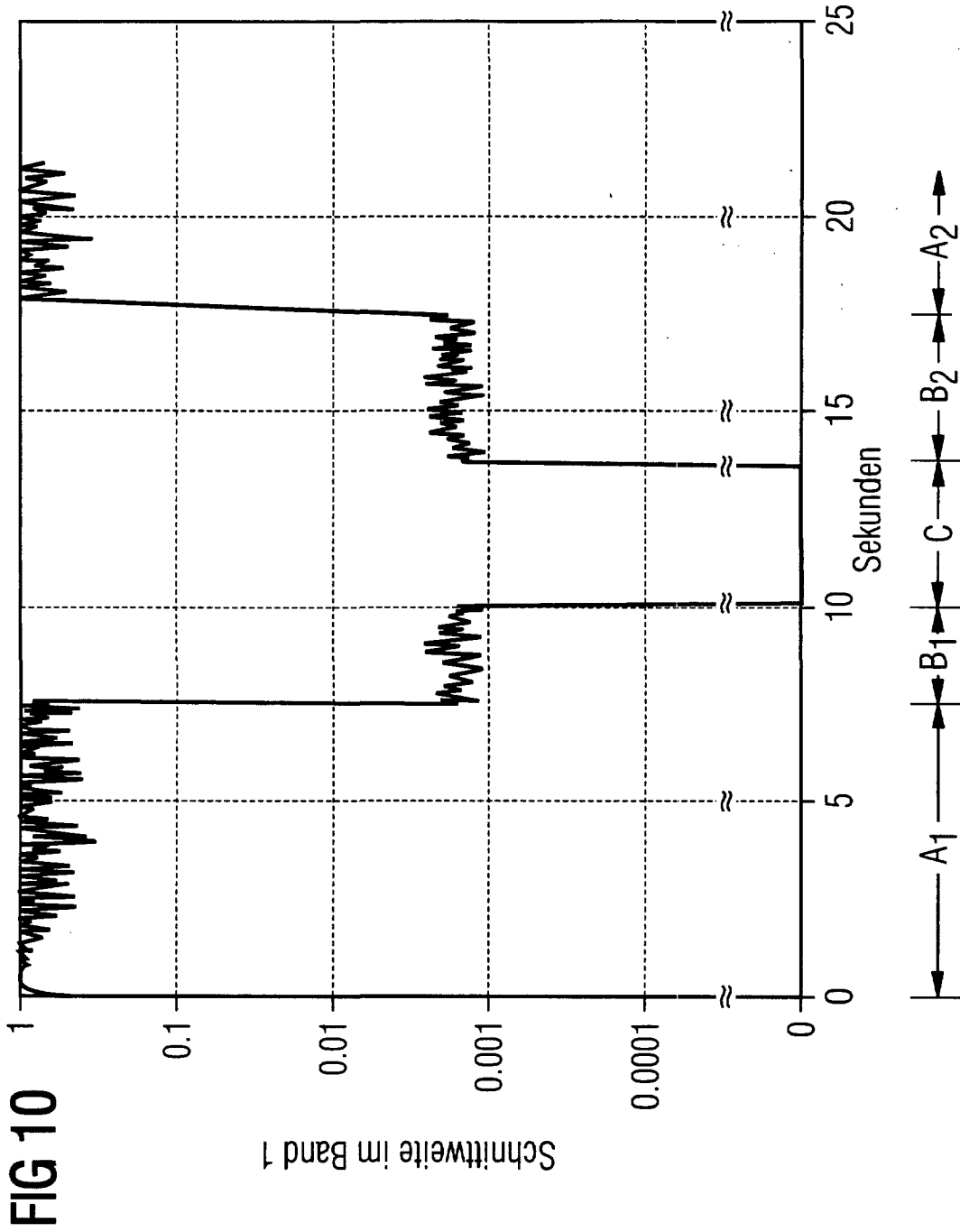


FIG 11

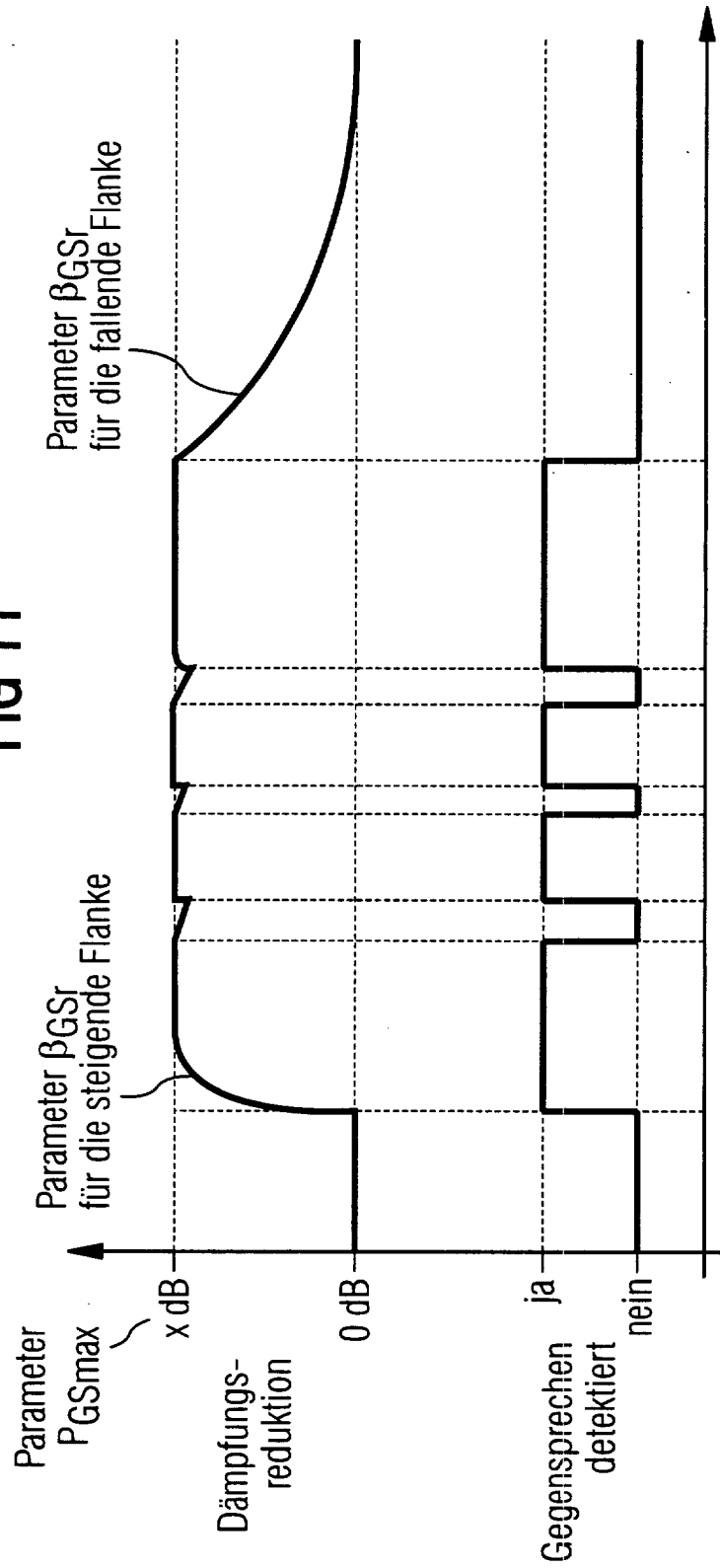


FIG 12

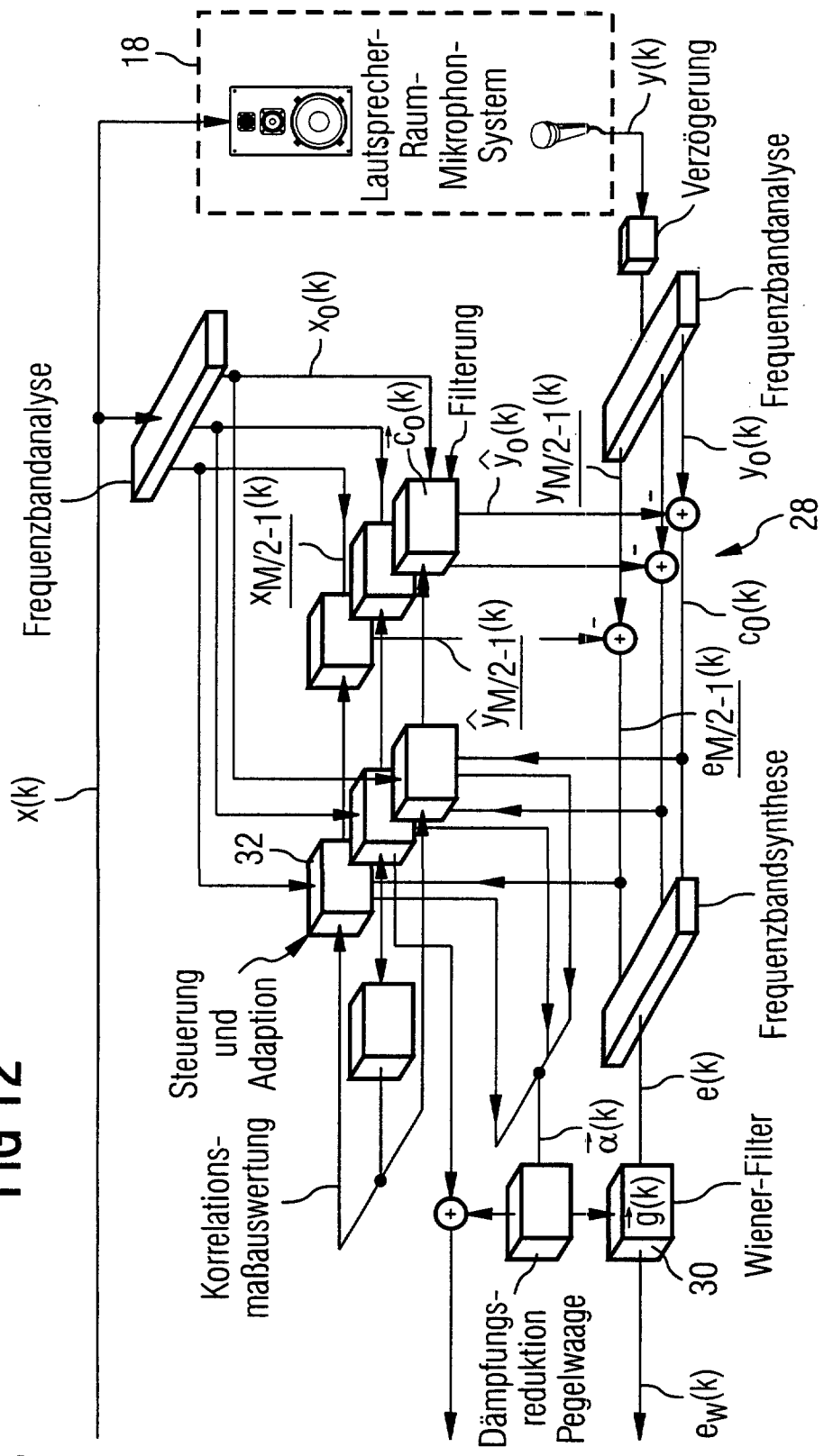
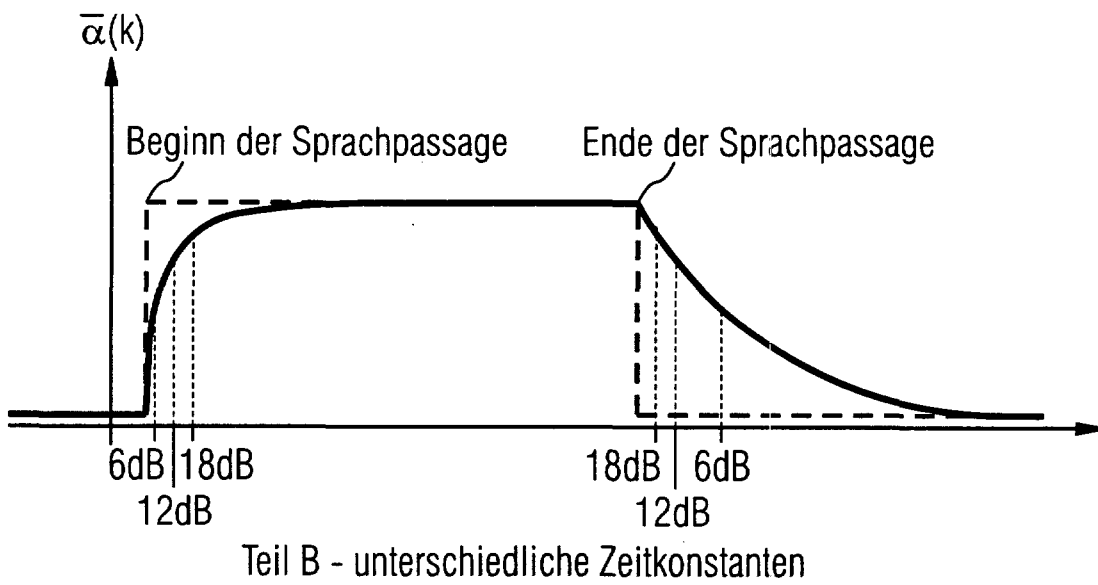
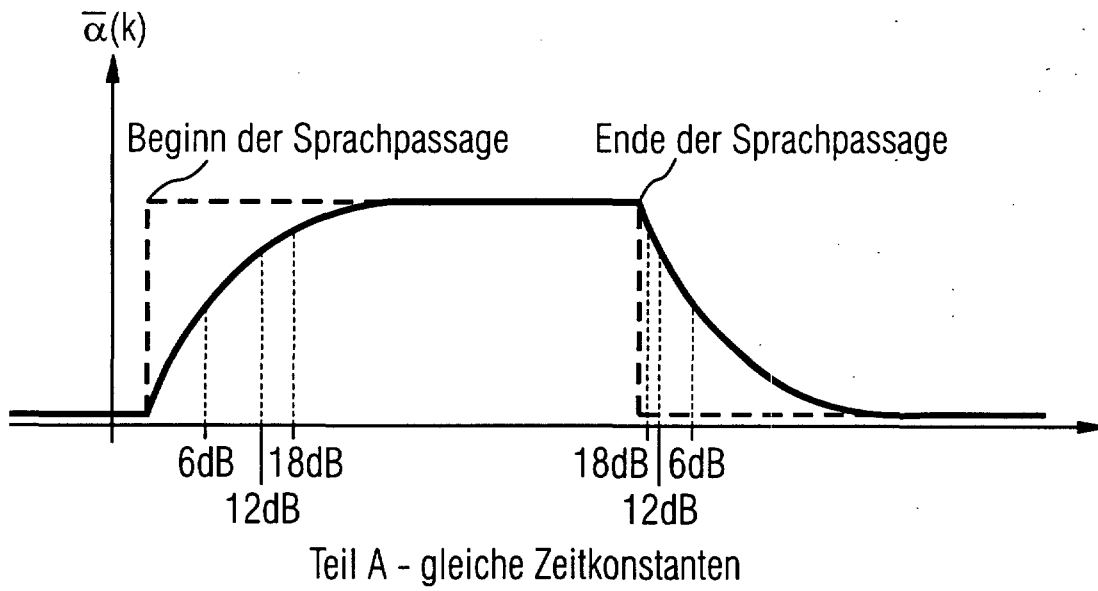


FIG 13



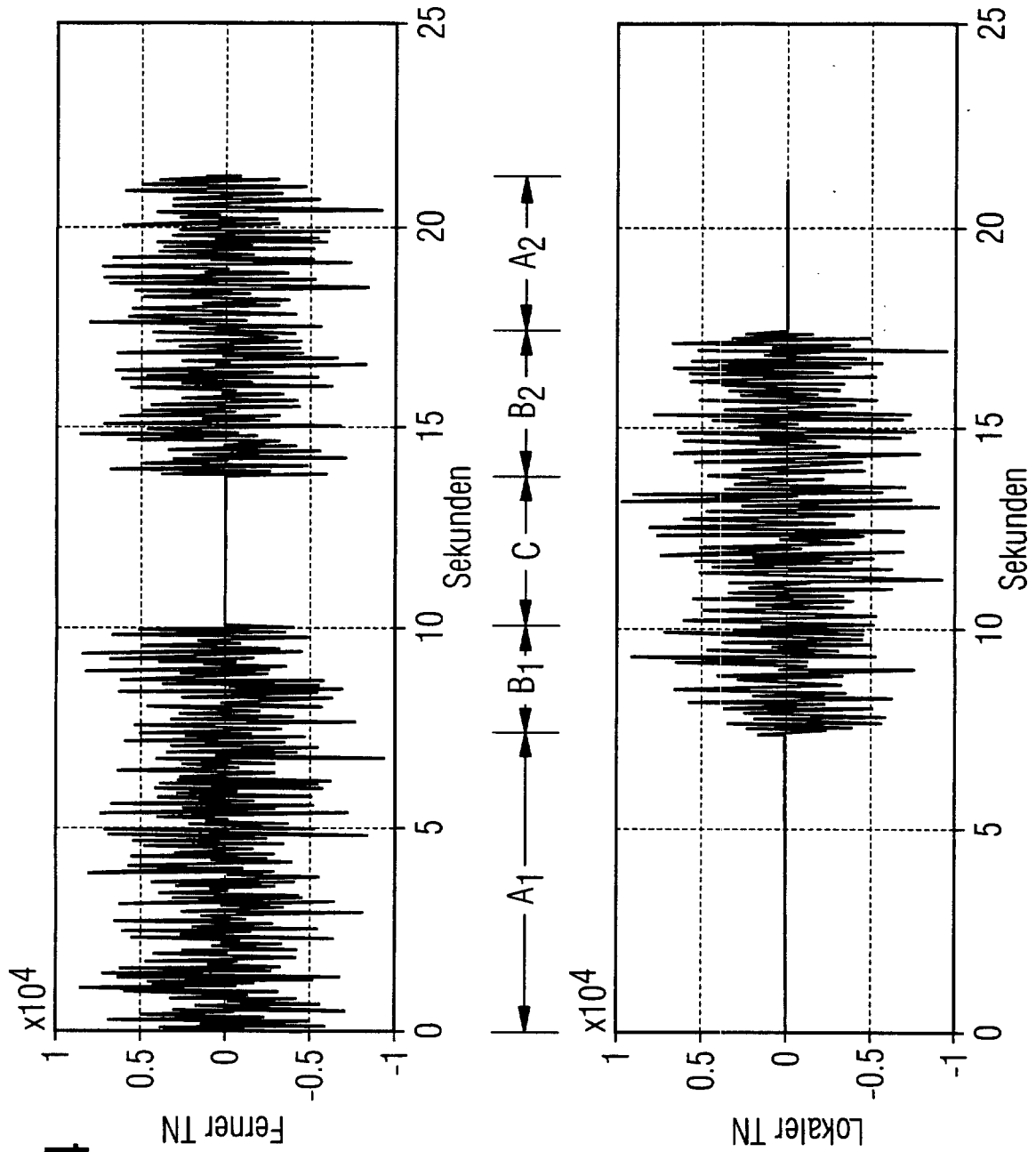
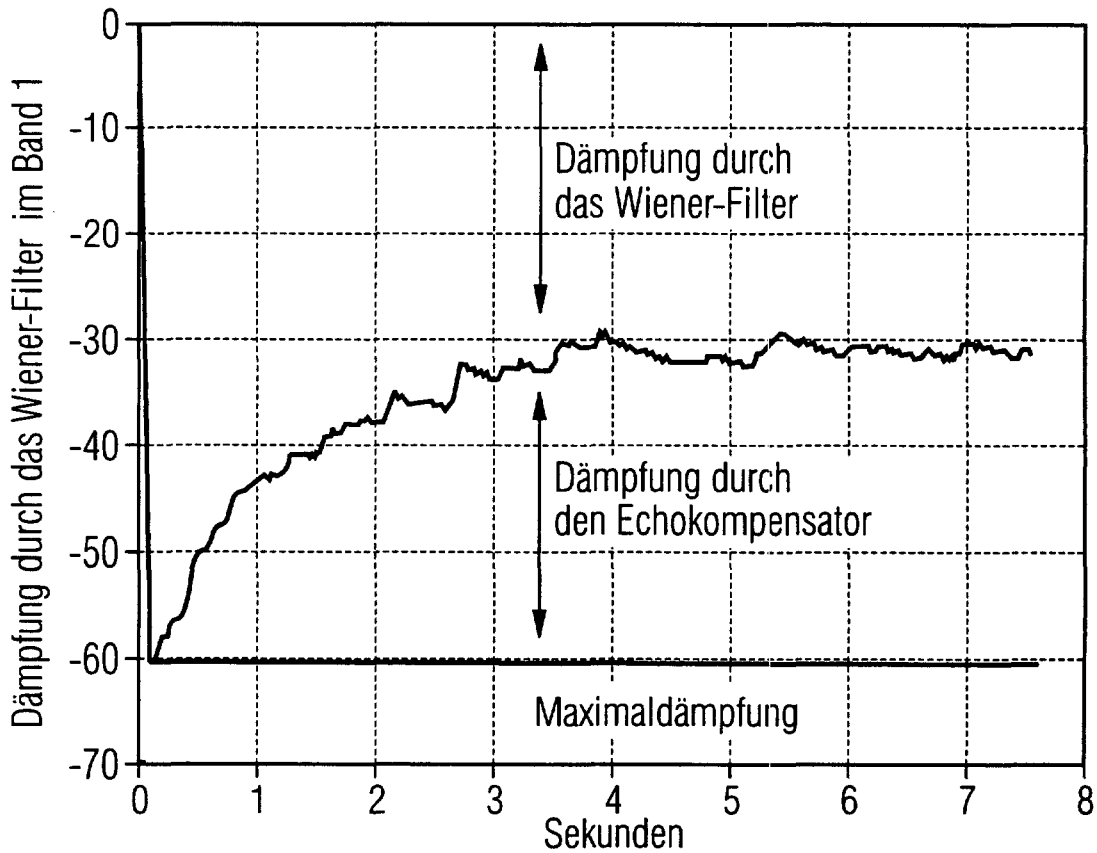
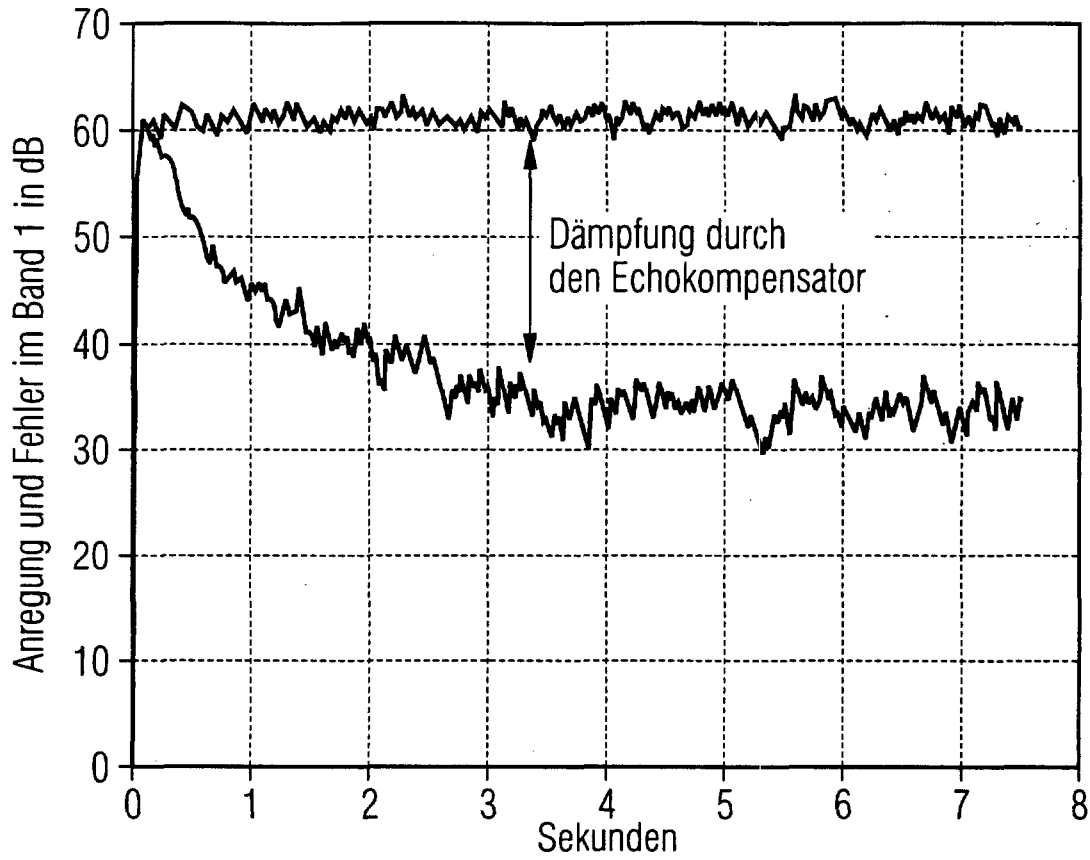


FIG 14

FIG 15



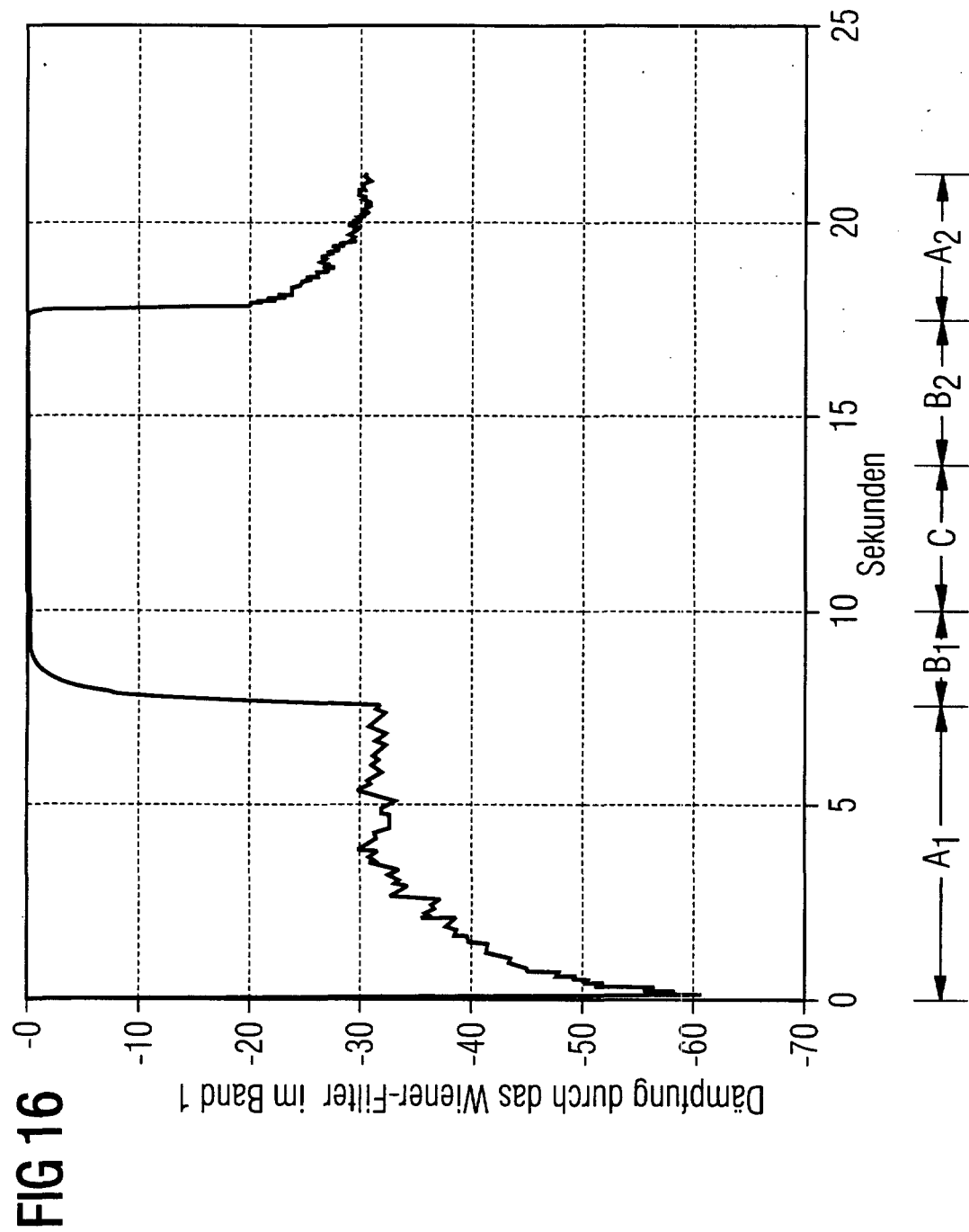


FIG 16

Dämpfung durch das Wiener-Filter im Band 1

FIG 17

