



⑫ A **Terinzagelegging** ⑪ **8400728**

Nederland

⑲ NL

- ⑤4 **Digitale spraakcoder met basisband residucodering.**
- ⑤1 Int.Cl⁴: G10L9/18, G10L9/14, H04B 14/04.
- ⑦1 Aanvrager: N.V. Philips' Gloeilampenfabrieken te Eindhoven.
- ⑦4 Gem.: Ir. P.J.P.G. Simons c.s.
Internationaal Octrooibureau B.V.
Prof. Holstlaan 6
5656 AA Eindhoven.

- ②1 Aanvraag Nr. 8400728.
- ②2 Ingediend 7 maart 1984.
- ③2 --
- ③3 --
- ③1 --
- ⑥2 --

- ④3 Ter inzage gelegd 1 oktober 1985.

De aan dit blad gehechte stukken zijn een afdruk van de oorspronkelijk ingediende beschrijving met conclusie(s) en eventuele tekening(en).

N.V. Philips' Gloeilampenfabrieken te Eindhoven

"Digitale spraakcoder met basisband residucodering"

(A). Achtergrond van de uitvinding

De uitvinding heeft betrekking op een digitale spraakcoder met een zender en ontvanger voor de transmissie van in segmenten opgedeelde digitale spraaksignalen, waarbij de zender is voorzien van:

- een eerste LPC-analysator voor het in responsie op het digitale spraaksignaal van elk segment genereren van eerste predictieparameters die de omhullende van het segment-termijn spectrum van dit digitale spraaksignaal karakteriseren,
 - een eerste adaptief invers filter voor het in responsie op het digitale spraaksignaal van elk segment en op de eerste predictieparameters genereren van een spraakband residusignaal dat correspondeert met de predictiefout van dit segment,
 - een decimerend filter voor het in responsie op het spraakband residusignaal genereren van een basisband residusignaal,
 - een codeer-en-multiplexcircuit voor het coderen van de eerste predictieparameters en de golfvorm van het basisband residusignaal en voor het in tijdmultiplex overdragen van de resulterende codesignalen;
- en waarbij de ontvanger is voorzien van:
- een demultiplex-en-decodeercircuit voor het separeren van de overgedragen codesignalen en voor het decoderen van de gesepareerde codesignalen tot de eerste predictieparameters en de golfvorm van het basisband residusignaal,
 - een interpolerende excitatiegenerator voor het in responsie op het basisband residusignaal genereren van een excitatiesignaal dat correspondeert met het spraakband residusignaal,
 - een eerste adaptief synthesefilter voor het in responsie op het excitatiesignaal en de eerste predictieparameters construeren van een replica van het digitale spraaksignaal.

Een dergelijke spraakcoder, die gebaseerd is op lineaire predictieve codering (LPC) als methode van spectrale analyse

8400728

is bekend uit het artikel van V.R. Viswanathan et al., "Design of a Robust Baseband LPC Coder for Speech Transmission over 9.6 kbit/s Noisy Channels", IEEE Trans. Commun., Vol. COM-30, no. 4, April 1982, pp. 663-673.

5 In dit type spraakcoder wordt het digitale spraak-sig-
naal gefilterd met behulp van een invers filter, waarvan de over-
drachtsfunctie $A(z)$ in z-transform notatie wordt gegeven door

$$A(z) = 1 - P(z) = 1 - \sum_{i=1}^p a(i) z^{-i}$$

10

waarbij $P(z)$ de overdrachtsfunctie is van een predictor gebaseerd
op een segment-termijn spectrale omhullende van het spraaksignaal,
de filtercoëfficiënten $a(i)$ met $1 \leq i \leq p$ de LPC-parameters zijn die
voor elk spraaksignaalsegment van bijvoorbeeld 20 ms opnieuw berekend
15 worden en p de LPC-orde is met gewoonlijk een waarde tussen 8 en 16.
Het spraakband residusignaal aan de uitgang van dit inverse filter
 $A(z)$ heeft een in het algemeen vlakke spectrale omhullende, die vlakker
wordt naar gelang de LPC-orde p hoger is. Dit spraakband residusignaal
wordt gebruikt als excitatiesignaal voor het (recursieve) synthese-
20 filter met dezelfde filtercoëfficiënten $a(i)$ en dus met een overdrachts-
functie $1/A(z)$. Aangezien dit synthesefilter $1/A(z)$ een maskerend
effect heeft op de quantiseringsruis van het spraakband residusignaal,
blijkt een codering van de golfvorm van dit residusignaal met 3 bits
per monster voldoende te zijn voor het verkrijgen van een zelfde spraak-
25 kwaliteit als in het geval van een golfvormcodering van het spraak-
signaal met behulp van een PCM-coder die voor telefonie is gestandaard-
iseerd, waarin de monsterfrequentie 8 kHz bedraagt en een codering met
8 bits per monster wordt gebruikt. De totale bitfrequentie benodigd voor
het coderen van het spraakband residusignaal en de LPC-parameters
30 is echter niet significant lager dan in het geval van een gestandaard-
iseerde PCM-coder, omdat het spraakband residusignaal nog steeds
dezelfde bandbreedte heeft als het spraakbandsignaal zelf.

De uit bovengenoemd artikel bekende spraakcoder maakt
nu gebruik van de in het algemeen vlakke gedaante van de spectrale
35 omhullende van het spraakband residusignaal om de in totaal benodigde
bitfrequentie te reduceren. Daartoe wordt het spraakband residusignaal
toegevoerd aan een digitaal laagdoorlaatfilter, waarin tevens een
verlaging van de monsterfrequentie (decimatie) met een factor N van

2 tot 8 wordt uitgevoerd. Om een bruikbaar excitatiesignaal voor het
synthesefilter $1/A(z)$ te verkrijgen, moet het ontbrekende hoog-
frequente deel van het spectrum worden teruggewonnen uit het beschik-
bare laagfrequente deel, de basisband, en moet verder een verhoging
5 van de monsterfrequentie (interpolatie) tot de oorspronkelijke waarde
worden uitgevoerd. Een excitatiesignaal met de bandbreedte van het
spraaksignaal zelf wordt in de bekende spraakcoder verkregen met
behulp van een spectrale vouwmethode. Bij spectrale vouwing bestaat
de interpolatie uit niets anders dan het invoegen van $N-1$ monsters
10 met de waarde nul na elk monster van het basisband residusignaal, waar-
bij N de decimatiefactor is. Bijgevolg bestaat het spectrum van het
excitatiesignaal uit een laagfrequent deel in de vorm van de behouden
basisband en een hoogfrequent deel in de vorm van vouwproducten van
de basisband rond de gedecimeerde monsterfrequentie en gehele veel-
15 vouden daarvan. Een voordeel van deze methode is dat een basisband
residusignaal met een vlakke spectrale omhullende met zekerheid
resulteert in een excitatiesignaal dat over de gehele spraakband even-
eens een vlakke spectrale omhullende heeft. Deze eigenschap komt
direct tot uiting in de goede spraakwaliteit die aldus wordt ver-
20 kregen, waarbij de "heesheid" - die typerend is voor de welbekende
niet-lineaire vervormingsmethoden ter verkrijging van een excitatie-
signaal met de bandbreedte van het spraaksignaal zelf - thans niet
aanwezig is.

Spectrale vouwing is dus een heel eenvoudige
25 methode, die echter een eigen probleem oplevert: de productie van
hoorbare "metaalachtige" achtergrondgeluiden die in de literatuur
bekend staan als "toonruis" (tonal noises) en die toenemen naarmate
de decimatiefactor N hoger is en naarmate de grondtoon (pitch) van
de spraak hoger is.

30 Met het oog op dit probleem wordt in de excitatie-
generator van de bekende spraakcoder een variant van de spectrale
vouwmethode toegepast, waarbij na de interpolatie de monsters van het
excitatiesignaal ook nog onderworpen worden aan een tijd-positie-
perturbatie. In het bijzonder wordt de tijdpositie van een monster
35 met een waarde ongelijk nul (dus een oorspronkelijk monster van
het basisband residusignaal voor de interpolatie) op willekeurige
wijze geperturbeerd en wel door het eenvoudig verwisselen van dit
monster met een naburig monster van de waarde nul indien de grootte

van dit monster beneden een vooraf bepaalde drempel blijft, waarbij de waarschijnlijkheid van perturbatie toeneemt naarmate de grootte van dit monster kleiner is. Enerzijds wordt nu het niet-geperturbeerde excitatiesignaal toegevoerd aan een laagdoorlaatfilter ter selectie
5 van de basisband en anderzijds wordt het geperturbeerde excitatiesignaal toegevoerd aan een hoogdoorlaatfilter ter selectie van het hoogfrequente deel boven de basisband, waarna de twee geselecteerde signalen wordt opgeteld voor het verkrijgen van het uiteindelijke excitatiesignaal. Volgens deze variant van de spectrale vouwmethode
10 wordt in wezen signaal-gecorrleerde ruis opgeteld bij het spectraal gebouwen basisband residusignaal. Perceptioneel blijkt deze additieve ruis weliswaar een maskerend effect te hebben op de "toonruis", maar ook wederom enige "heesheid" te introduceren. De toepassing van deze variant in de bekende spraakcoder betekent dus een aanzienlijke
15 extra complicatie bij de praktische implementatie, maar leidt niet tot een bevredigende oplossing van het probleem van de "toonruis" bij spectrale vouwing als methode voor het verkrijgen van een excitatiesignaal met dezelfde bandbreedte als het spraaksignaal.

20 (B). Samenvatting van de uitvinding

De uitvinding beoogt een digitale spraakcoder van het in de aanhef van paragraaf (A) vermelde type te verschaffen die het optreden van "toonruis" op effectieve wijze tegengaat en die leidt
25 tot een relatief eenvoudige praktische implementatie.

De digitale spraakcoder volgens de uitvinding heeft het kenmerk, dat de zender verder is voorzien van:

- een tweede LPC-analysator voor het in responsie op het spraakband residusignaal van het eerste adaptieve inverse filter genereren
30 van tweede predictieparameters die de fijnstructuur van het kortetermijn spectrum van dit spraakband residusignaal karakteriseren,
- een tweede adaptief invers filter voor het in responsie op het spraakband residusignaal en op de tweede predictieparameters genereren van een gemodificeerd spraakband residusignaal dat wordt
35 toegevoerd aan het decimerende filter;

het codeer-en-multiplexcircuit in de zender en het demultiplex-en-decodeercircuit in de ontvanger zijn ingericht voor het verwerken van zowel de eerste als de tweede predictieparameters; en

de ontvanger verder is voorzien van:

- een tweede adaptief synthesefilter voor het in responsie op het excitatiesignaal van de interpolerende excitatiegenerator en de tweede predictieparameters construeren van een gemodificeerd excitatie-
5 signaal dat wordt toegevoerd aan het eerste adaptieve synthesefilter.

De maatregelen volgens de uitvinding berusten op het inzicht dat de overwegend in periodieke (stemhebbende) spraakfragmenten optredende "toonruis" in wezen wordt veroorzaakt door de
10 niet-harmonische relatie tussen de spraakfrequentiecomponenten van de verschillende spectraal gevouwen versies van het basisband residusignaal, maar dat voor niet-periodieke (stemloze) spraakfragmenten geen perceptioneel ongewenste effecten worden veroorzaakt door de spectrale vouwing. In de spraakcoder volgens de uitvinding wordt het
15 spraakband residusignaal ontdaan van mogelijk periodiciteit en dus van harmonisch gelegen spraakfrequentiecomponenten met behulp van een tweede adaptief invers filter. Bijgevolg worden zowel de decimatie in de zender als de door simpele interpolatie bewerkstelligde spectrale vouwing in de ontvanger uitgevoerd op signalen die altijd een gepronon-
20 ceerd niet-periodiek karakter hebben, zodat het ontstaan van "toonruis" op effectieve wijze wordt tegengegaan. Pas na uitvoering van de spectrale vouwing wordt de gewenste periodiciteit wederom geïntroduceerd in het spraakband excitatiesignaal met behulp van een tweede adaptief synthesefilter dat de tegenhanger is van het tweede adaptieve inverse
25 filter.

In verband met de maatregelen volgens de uitvinding wordt nog gewezen op het feit dat de bekende spraakcoder gebruik maakt van adaptieve predictieve codering (APC) voor de transmissie van het basisband residusignaal, vergelijk Fig. 6 van het in paragraaf
30 (A) vermelde artikel. De APC-coder gebruikt een ruis-tegenkoppelconfiguratie en bevat een ingangsfILTER in de vorm van een adaptief invers filter, waarvan de adaptie wordt uitgevoerd in responsie op de locatie en de waarde van de maximale autocorrelatiecoëfficiënt van het ingangssignaal voor vertragingen groter dan 2 ms, en de APC-
35 decoder bevat een adaptief synthesefilter dat de tegenhanger is van het adaptieve inverse filter in de PAC-coder. Ofschoon het ingangssignaal van de APC-coder ontdaan wordt van mogelijke periodiciteit die wederom wordt geïntroduceerd in het uitgangssignaal van de APC-

decoder, wordt het ontstaan van "toonruis" in de bekende spraakcoder niet tegengegaan door deze maatregelen. Immers de herintroductie van de periodiciteit geschiedt voorafgaand aan de interpolatie en bijgevolg produceert de spectrale vouwing "toonruis" die niet ver-
5 wijderd, maar slechts gemaskerd wordt door de verdere maatregelen in de bekende spraakcoder, waarbij bovendien enige "heesheid" als neveneffect optreedt. Het is daarom wezenlijk voor de onderhavige uitvinding dat de tweede adaptieve inverse filtering plaats vindt voorafgaand aan de decimatie en de corresponderende tweede adaptieve
10 synthesefiltering plaats vindt na de door simpele interpolatie bewerkstelligde spectrale vouwing.

(C). Korte beschrijving van de tekeningen

15 Bijzonderheden en voordelen van de spraakcoder volgens de uitvinding zullen thans worden toegelicht in de navolgende beschrijving van een uitvoeringsvoorbeeld aan de hand van de bijgevoegde tekeningen. Daarin toont:

Fig. 1 een blokschema van een digitale spraakcoder
20 volgens de uitvinding,

Fig. 2 een tweetal frequentiediagrammen ter toelichting van de spectrale vouwmethode,

Fig. 3, Fig. 4 en Fig. 5 een aantal amplitude-spectra en een autocorrelatiefunctie van signalen op verschillende
25 punten van de spraakcoder in Fig. 1 die alle betrekking hebben op een zelfde segment van het spraaksignaal.

(D). Beschrijving van een uitvoeringsvoorbeeld

30 In Fig. 1 is een functioneel blokschema weergegeven van een digitale spraakcoder met een zender 1 en een ontvanger 2 voor de transmissie van een digitaal spraaksignaal over een kanaal 3, waarvan de transmissiecapaciteit significant lager is dan de waarde van 64 kbit/s van een standaard PCM-kanaal voor telefonie.

35 Dit digitale spraaksignaal representeert een analoog spraaksignaal dat afkomstig is van een bron 4 met een microfoon of een andere electro-acoustische transducent en dat is begrensd tot een spraakband van 0-4 kHz met behulp van een laagdoorlaatfilter 5.

Dit analoge spraaksignaal wordt bemonsterd met een monsterfrequentie van 8 kHz en omgezet in een voor gebruik in zender 1 geschikte digitale code met behulp van een analoog-digitaalconverter 6 die tevens een opdeling uitvoert van dit digitale spraaksignaal in over-
 5 lappende segmenten van 30 ms (240 monsters) die elke 20 ms worden vernieuwd. In zender 1 wordt dit digitale spraaksignaal verwerkt tot een signaal dat kan worden overgedragen over kanaal 3 naar ontvanger 2 en daarin kan worden verwerkt tot een replica van dit digitale spraaksignaal. Met behulp van een digitaal-analoogconverter 7 wordt het
 10 gerepliceerde digitale spraaksignaal omgezet in een analoog spraaksignaal dat na begrenzing tot de spraakband van 0-4 kHz in een laag-doorlaatfilter 8 wordt toegevoerd aan een reproductiecircuit 9 met een luidspreker of een andere electro-acoustische transducent.

De in Fig. 1 weergegeven spraakcoder behoort tot
 15 de klasse van hybride coders die in de literatuur bekend staan als RELP-coders (Residual-Excited Linear Prediction). Eerst wordt nu de basisvorm van een RELP-coder beschreven aan de hand van Fig. 1.

In zender 1 worden de segmenten van het digitale spraaksignaal toegevoerd aan een LPC-analysator 10, waarin elke 20 ms
 20 de LPC-parameters van een spraaksegment van 30 ms op bekende wijze worden berekend, bijvoorbeeld op de basis van de autocorrelatiemethode of de covariantiemethode van lineaire predictie (vergelijk R.W. Schafer, J.D. Markel, "Speech Analysis", IEEE Press, New York, 1978, pp. 124-143). Het digitale spraaksignaal wordt eveneens
 25 toegevoerd aan een adaptief filter 11 met een predictor 12 en een aftrekker 13. Predictor 12 is een transversaal filter, waarvan de coëfficiënten $a(i)$ met $1 \leq i \leq p$ de in analysator 10 berekende LPC-parameters zijn, waarbij de LPC-orde p gewoonlijk een waarde tussen 8 en 16 heeft. In z -transform notatie wordt de overdrachtsfunctie $P(z)$
 30 van predictor 12 gegeven door:

$$P(z) = \sum_{i=1}^p a(i) z^{-i} \quad (1)$$

en de overdrachtsfunctie $A(z)$ van filter 11 door:

$$A(z) = 1 - P(z) \quad (2)$$

De LPC-parameters $a(i)$ worden zo bepaald dat het uitgangssignaal van filter 11, het spraakband (predictie) residusignaal, een zo vlak mogelijke segment-termijn (30 ms) spectrale omhullende bezit. Filter 11 staat daarom in de literatuur bekend als invers filter.

5 In de basisvorm van een RELP-coder worden de LPC-parameters $a(i)$ en de golfvorm van het spraakband residusignaal van zender 1 naar ontvanger 2 overgedragen. In ontvanger 2 wordt het overgedragen spraakband residusignaal gebruikt als excitatiesignaal voor een adaptief synthesefilter 14 met een predictor 15 en een
10 opteller 16 in een recursieve configuratie. Predictor 15 is eveneens een transversaal filter met als coëfficiënten de overgedragen LPC-parameters $a(i)$, zodat de overdrachtsfunctie van predictor 15 eveneens wordt gegeven door formule (1) en de overdrachtsfunctie van synthesefilter 14 door:

$$15 \quad 1/[1 - P(z)] = 1/A(z) \quad (3)$$

In het hier veronderstelde ideale geval van volkomen vervormingsvrije transmissie en volkomen stationaire spraaksignalen zijn de beide filters 11 en 14 exact invers ten opzichte van elkaar, zodat het
20 oorspronkelijke digitale spraaksignaal aan de ingang van zender 1 wordt herwonnen aan de uitgang van synthesefilter 14 in de ontvanger. Aangezien spraaksignalen slechts als lokaal stationair beschouwd mogen worden en dus de LPC-parameters $a(i)$ voor beide predictoren 12, 15 elke 20 ms vernieuwd worden, is deze veronderstelling slechts
25 in eerste benadering geldig, maar ook dan blijkt bij volkomen vervormingsvrije transmissie geen perceptioneel verschil te bestaan tussen het oorspronkelijke analoge spraaksignaal aan de uitgang van filter 5 in zender 1 en het gerepliceerde analoge spraaksignaal aan de uitgang van filter 8 in ontvanger 2.

30 In de praktijk vereist de digitale transmissie van de LPC-parameters $a(i)$ en de golfvorm van het spraakband residusignaal een quantisering en een codering. Daartoe bevat zender 1 een codeer- en-multiplexcircuit 17 met een parametercoder 18, een adaptieve golfvormcoder 19 en een multiplexer 20 voor het combineren van de resul-
35 terende codesignalen tot een tijdmultiplexsignaal. Ontvanger 2 bevat een corresponderend demultiplex-en-decodeercircuit 21 met een demultiplexer 22 voor het separeren van de in tijdmultiplex overgedragen codesignalen, een parametercodecoer 23 en een adaptieve golfvorm-

decoder 24.

Zoals bekend, verdient het de voorkeur om voor de transmissie van de LPC-parameters $a(i)$ gebruik te maken van "log area ratio" (LAR) coëfficiënten $g(i)$, die verkregen worden door de LPC-parameters $a(i)$ eerst om te zetten in reflectiecoëfficiënten $k(i)$ en dan de volgende logaritmische transformatie toe te passen:

$$g(i) = \log \left[\frac{1 + k(i)}{1 - k(i)} \right], \quad 1 \leq i \leq p \quad (4)$$

Deze LAR-coëfficiënten $g(i)$ worden uniform gequantiseerd en elke 20 ms gecodeerd, waarbij het totale aantal bits op optimale wijze wordt toegewezen aan de verschillende LAR-coëfficiënten $g(i)$ volgens een bekende methode voor het minimaliseren van de maximale spectrale fout in het gerepliceerde digitale spraaksignaal (vergelijk V.R. Viswanathan, J. Makhoul, "Quantization Properties of Transmission Parameters in Linear Predictive Systems", IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-23, No. 3, June 1975, pp. 309-321). Wanneer in parametercoder 18 elke 20 ms bijvoorbeeld 64 bits beschikbaar zijn voor de transmissie van 16 LPC-parameters $a(i)$ en de LPC-orde dus $p = 16$ is, dan wordt de volgende bittoewijzing voor de LAR-coëfficiënten $g(1) - g(16)$ gebruikt: 6 bits voor $g(1)$, $g(2)$; 5 bits voor $g(3)$, $g(4)$; 4 bits voor $g(5) - g(10)$; 3 bits voor $g(11) - g(16)$. Door voor de LAR-coëfficiënten benodigde transmissiecapaciteit van kanaal 3 bedraagt dan 3,2 kbit/s. Aangezien predictor 15 van synthesefilter 14 in ontvanger 2 gebruik maakt van LPC-parameters $a(i)$ die zijn verkregen uit gequantiseerde LAR-coëfficiënten $g(i)$ met behulp van parameterdecoder 23, dient predictor 12 van invers filter 11 in zender 1 gebruik te maken van dezelfde gequantiseerde waarden van de LPC-parameters $a(i)$.

Voor de transmissie van het spraakband residusignaal kan in principe elk van de bekende golfvormcodeermethodes gebruikt worden. In Fig. 1 is gekozen voor een eenvoudige adaptieve PCM-methode, waarbij in zender 1 de maximum amplitude D van het spraakband residusignaal voor elk interval van 20 ms wordt bepaald met behulp van een maximum detector 25 en waarbij adaptieve PCM-coder 19 de monsters van het spraakband residusignaal uniform quantiseert in een bereik $(-D, +D)$. Aangezien synthesefilter 14 een maskerend effect heeft op de quantiseringsruis, is in PCM-coder 19 een codering met 3 bits per monster voldoende voor het verkrijgen van een zelfde spraak-

8400728

5 kwaliteit als in het geval van de voor openbare telefonie reeds
jaren gestandaardiseerde (logaritmische) PCM met een codering van
8 bits per monster. De maximum amplitude D wordt in een dynamiek-
bereik van 64 dB logaritmisch gecodeerd met 6 bits in parametercoder
18. Na decodering in parameterdecoder 23 wordt deze maximum amplitude
D in ontvanger 2 gebruikt voor de besturing van adaptieve PCM-
decoder 24. De voor het spraakband residusignaal benodigde capaciteit
van transmissiekanaal 3 bedraagt dan 24,3 kbit/s.

Bij het multiplexen van de codesignalen voor de
16 LAR-coëfficiënten (3,2 kbit/s) en voor het spraakband residu-
signaal (24,3 kbit/s) worden door multiplexer 20 nog 2 bits toegevoegd
aan het 20 ms frame van het tijdmultiplexsignaal voor de synchronisatie
van demultiplexer 22, zodat de beschreven basisvorm van een RELP-
coder een transmissiekanaal 3 vereist met in totaal een capaciteit
15 van 27,6 kbit/s. Deze waarde betekent weliswaar een belangrijke
verbetering ten opzichte van de waarde van 64 kbit/s voor de ge-
standaardiseerde PCM, maar in vergelijking met adaptieve differentiële
PCM (ADPCM) die thans in overweging is als mogelijke nieuwe standaard
voor openbare telefonie en die slechts een transmissiecapaciteit
20 van 32 kbit/s vereist, kan deze verbetering niet als significant worden
beschouwd.

Uit het beschreven voorbeeld blijkt duidelijk dat in
de basisvorm van een RELP-coder verreweg het grootste deel (88%)
van de capaciteit van kanaal 3 wordt gebruikt voor de transmissie van
25 een residusignaal in de spraakband van 0-4 kHz, dat wil zeggen met
een bandbreedte gelijk aan die van het over te dragen spraaksignaal
zelf. Een belangrijke reductie van deze transmissiecapaciteit kan
nu worden bereikt door gebruik te maken van het feit dat dit spraak-
band residusignaal een in het algemeen vlakke spectrale omhullende
30 heeft.

De daartoe gebruikte methode is bekend (vergelijk het in
paragraaf (A) vermelde artikel) en bestaat uit het selecteren van een
basisband van bijvoorbeeld 0-1 kHz van het spraakband residusignaal
aan de uitgang van invers filter 11 in zender 1 en het corresponderend
35 verlagen van de monsterfrequenties van 8 kHz met een decimatiefactor
 $N = 4$ tot een monsterfrequentie van 2 kHz. In de praktijk worden beide
signaalbewerkingen gecombineerd uitgevoerd in een digitaal decimerend
laagdoorlaatfilter 26. Het aldus verkregen basisband residusignaal

8400728

wordt toegevoerd aan adaptieve PCM-coder 19 en daarin op dezelfde wijze gecodeerd als het spraakband residusignaal bij de basisvorm van de RELP-coder. Dank zij de decimatie van de monsterfrequentie tot een waarde van 2 kHz is echter de voor het basisband residusignaal
5 benodigde transmissiecapaciteit van kanaal 3 aanzienlijk lager en deze capaciteit bedraagt thans slechts 6,3 kbit/s. Bij ongewijzigde transmissie van de 16 LAR-coëfficiënten en de 2 framesynchronisatiebits vereist deze basisbandversie van een RELP-coder dan een transmissie- kanaal 3 met in totaal een capaciteit van 9,6 kbit/s, een waarde die
10 wel beschouwd kan worden als significant lager dan de capaciteit van 64 kbit/s nodig voor een standaard PCM-kanaal.

Om in ontvanger 2 een bruikbaar excitatiesignaal voor synthesefilter 14 te herkrijgen, moet het ontbrekende hoogfrequente deel in de band van 1-4 kHz worden teruggewonnen uit het beschikbare
15 overgedragen basisband residusignaal en moet verder de gedecimeerde monsterfrequentie van 2 kHz worden verhoogd met een factor $N = 4$ tot de oorspronkelijke waarde van 8 kHz. Daartoe wordt in ontvanger 2 gebruik gemaakt van een spectrale vouwmethode, waarbij de excitatie- signaalgenerator die deze beide signaalbewerkingen gecombineerd
20 uitvoert, uit niets anders bestaat dan een simpele interpolator 27 die na elk monster van het overgedragen basisband residusignaal $N-1 = 3$ monsters met de waarde nul invoegt. Bijgevolg heeft het excitatie- signaal aan de uitgang van interpolator 27 niet alleen de oorspronke- lijke monsterfrequentie van 8 kHz, maar ook een spectrum waarvan het
25 laagfrequente deel wordt gevormd door de behouden basisband van 0-1 kHz en het hoogfrequente deel boven 1 kHz wordt gevormd door de vouwproducten van deze basisband rond de gedecimeerde monsterfrequentie van 2 kHz en gehele veelvouden daarvan. Een belangrijk voordeel van deze spectrale vouwmethoden is dat het excitatiesignaal een in het
30 algemeen vlakke spectrale omhullende heeft over de gehele spraakband van 0-4 kHz. Deze eigenschap komt direct tot uiting in de goede kwaliteit van de aldus verkregen analoge spraaksignalen, waarbij de "heesheid" die typerend is voor niet-lineaire vervormingsmethoden ter verkrijging van een geschikt excitatiesignaal, thans niet aanwezig
35 is.

De spectrale vouwing blijkt echter hoorbare "metaalachtige" achtergrondgeluiden op te leveren die bekend staan als "toonruis" en die toenemen naarmate de decimatiefactor N hoger is en naarmate

de grondtoon (pitch) van de spraak hoger is.

Uit uitvoerige onderzoeken naar de oorzaken van deze "toonruis" heeft Aanvrager het inzicht gewonnen dat de overwegend in periodieke (stemhebbende) spraakfragmenten optredende "toonruis" in wezen wordt veroorzaakt door de niet-harmonische relatie tussen de spraakfrequentiecomponenten van de verschillende spectraal gevouwen versies van het basisband residusignaal. Voor niet-periodieke (stemloze) spraakfragmenten veroorzaakt de spectrale vouwing daarentegen geen perceptueel ongewenste effecten. De verstoring van de harmonische relatie door spectrale vouwing wordt geïllustreerd in Fig. 2. Daarin toont frequentiediagram a een voorbeeld van het spectrum van een periodiek spraakband residusignaal met een gestippeld weergegeven vlakke spectrale omhullende en met een grondtoon (pitch) van 300 Hz. Selectie van de basisband van 0-1 kHz en de daarin gelegen componenten bij 300, 600 en 900 Hz met behulp van decimerend laagdoorlaatfilter en spectrale vouwing met behulp van interpolator 27 resulteert dan in een excitatiesignaal met een in frequentiediagram b weergegeven spectrum. Weliswaar heeft het excitatiesignaal in frequentiediagram b eveneens een vlakke spectrale omhullende, maar de componenten van de spectraal gevouwen versies in de respectieve banden van 1-2 kHz, 2-3 kHz en 3-4 kHz bezitten niet langer een harmonische relatie, zowel onderling als ook ten opzichte van de componenten in de (behouden) basisband van 0-1 kHz.

Het feit dat de "toonruis" blijkt toe te nemen met toenemende decimatiefactor N en toenemende grondtoonfrequentie, onderstreept dat juist de niet-harmonische uitbreiding van het basisband residusignaal (zelf wel harmonisch bij periodieke spraakfragmenten) in wezen verantwoordelijk geacht moet worden voor het optreden van de "toonruis", omdat een toenemende decimatiefactor en een toenemende grondtoonfrequentie in het algemeen gepaard gaan met een toenemende verstoring van de oorspronkelijk harmonische relatie tussen de componenten van een periodiek spraakband residusignaal.

Overeenkomstig de uitvinding wordt nu het spraakband residusignaal aan de uitgang van invers filter 11 in zender 1 ontdaan van mogelijke periodiciteit en dus van harmonisch gelegen componenten met behulp van een tweede adaptief invers filter 28 met een predictor 29 en een aftrekker 30. Predictor 29 is eveneens een transversaal filter, waarvan de coëfficiënten tweede LPC-parameters zijn die elke 20 ms

berekend worden in een tweede LPC-analysator 31 en die de fijnstructuur van het korte-termijn (20 ms) spectrum van het spraakband residusignaal karakteriseren. Zonder wezenlijk verlies in effectiviteit kan volstaan worden met een predictor 29, waarvan bijna alle coëfficiënten zijn ingesteld op de waarde nul en slechts zeer weinige coëfficiënten, of zelfs maar één coëfficiënt, een waarde ongelijk nul bezitten. Uit een oogpunt van eenvoud verdient dan een predictor 29 met één coëfficiënt de voorkeur, te meer daar het benutten van meer coëfficiënten, bijvoorbeeld 3 of 5, in slechts zeer marginale verbeteringen blijkt te resulteren. In het beschreven uitvoeringsvoorbeeld is predictor 29 dan ook een transversaal filter met slechts één coëfficiënt c en een overdrachtsfunctie $PP(z)$ die in z -transform notatie wordt gegeven door:

$$PP(z) = c z^{-M} \quad (5)$$

waar M het grondinterval van de periodiciteit is, uitgedrukt in het aantal monsters van het spraakband residusignaal. De beide tweede predictieparameters c en M worden verkregen met behulp van een simpele tweede LPC-analysator in de vorm van een autocorrelator 31 die de autocorrelatiefunctie $R(n)$ van elk interval van 20 ms van het spraakband residusignaal berekent voor vertragingen ("laags"), uitgedrukt in het aantal n van de monsters, groter dan de LPC-orde p van analysator 10, en die verder M bepaalt als de locatie van het maximum van $R(n)$ voor $n > p$ en c als de verhouding $R(M)/R(0)$. Dit tweede adaptieve inverse filter 28 heeft een overdrachtsfunctie $AA(z)$ die wordt gegeven door:

$$AA(z) = 1 - PP(z) = 1 - c z^{-M} \quad (6)$$

Aan de uitgang van filter 28 treedt dan een gemodificeerd spraakband residusignaal op dat zowel voor stemloze als ook voor stemhebbende spraakfragmenten een geprononceerd niet-periodiek karakter heeft. In ontvanger 2 wordt de gewenste periodiciteit pas na uitvoering van de spectrale vouwing in interpolator 27 wederom geïntroduceerd in het excitatiesignaal met behulp van een tweede adaptief synthesefilter 32, dat de tegenhanger is van tweede invers filter 28 in zender 1 en dat een predictor 33 en een opteller 34 in een recursieve configuratie bevat. De overdrachtsfunctie van predictor 33 wordt dus eveneens gegeven door formule (5) en de overdrachtsfunctie van dit tweede adaptieve synthesefilter 32 door:

$$1/[1 - PP(z)] = 1/AA(z) \quad (7)$$

Aan de uitgang van dit tweede adaptieve synthesefilter 32 treedt dan een gemodificeerd excitatiesignaal op met de gewenste harmonische
 5 relatie tussen de periodieke componenten over de gehele spraakband van 0-4 kHz, welk gemodificeerd excitatiesignaal wordt toegevoerd aan het eerste adaptieve synthesefilter 14. Dank zij deze maatregelen worden zowel de decimerende laagdoorlaatfiltering in zender 1 voor het verkrijgen van een basisband residusignaal als ook de door inter-
 10 polatie bewerkstelligde spectrale vouwing in ontvanger 2 voor het verkrijgen van een excitatiesignaal uitgevoerd op signalen die in wezen altijd vrij zijn van periodiciteit, zodat het ontstaan van "toonruis" bij spectrale vouwing op effectieve wijze wordt tegengegaan.

Voor niet-periodieke spraaksignalen zoals stemloze spraak-
 15 fragmenten of spraakpauzes, is de maximale autocorrelatiecoëfficiënt $R(M)$ zo laag en dus de waarde van predictieparameter $c = R(M)/R(0)$ zo klein dat het spraakband residusignaal het tweede inverse filter 28 nagenoeg ongemodificeerd doorloopt. Voor periodieke spraaksignalen zoals stemhebbende spraakfragmenten, wordt de periodiciteit van het
 20 spraakband residusignaal voornamelijk bepaald door de grondtoon (pitch). Nu hebben de hoogste grondtoonfrequenties die in spraak voorkomen altijd een waarde kleiner dan 500 Hz en dus een periode groter dan 2 ms, terwijl voor waarden beneden 100 Hz en dus grondtoonperiodes groter dan 10 ms, geen hoorbare "toonruis" wordt waargenomen. Wat betreft
 25 de praktische implementatie van autocorrelator 31 houdt dit in dat de autocorrelatiefunctie $R(n)$ slechts berekend behoeft te worden in het interval van 2 ms tot 10 ms en dus voor waarden n met $17 \leq n \leq 80$ bij een monsterfrequentie van 8 kHz, waardoor een aanzienlijke besparing aan rekeninspanning wordt bereikt. In het bijzonder wordt $R(n)$
 30 berekend volgens de formule:

$$R(n) = \sum_{r=0}^{159-n} b(r) \cdot b(r+n), \quad 17 \leq n \leq 80 \quad (8)$$

35 waarbij $b(r)$ met $r = 0, 1, 2, \dots, 159$ de monsters van het spraakband residusignaal voorstellen in het interval van 20 ms. De waarde van $R(n)$ voor $n = 0$, dus:

$$R(0) = \sum_{r=0}^{159} b^2(r) \quad (9)$$

8400728

wordt genormaliseerd op $R(0) = 2048$ zodat de predictieparameter c wordt gegeven door:

$$C = R(M)/2048 \quad (10)$$

- 5 Aangezien voor M geldt dat $17 \leq M \leq 80$, kan de waarde van M met 6 bits worden gecodeerd. In de praktijk kan worden volstaan met een quantisering van de waarde van c waarbij 4 bits worden gebruikt. Deze codering van de tweede predictieparameters c en M dient elke 20 ms te worden uitgevoerd, waartoe parametercoder 18 in zender 1 en parameterdecoder 10 23 in ontvanger 2 zo worden ingericht dat zowel de LPC-parameters $a(i)$ met $1 \leq i \leq p$ als ook de tweede predictieparameters c , M worden verwerkt. Aangezien predictor 33 van synthesefilter 32 in ontvanger 2 gebruik maakt van een gequantiseerde predictieparameter c , dient predictor 29 van invers filter 28 in zender 1 gebruik te maken van 15 dezelfde gequantiseerde waarde van c .

- Door de effectieve verwijdering van "toonruis" is het mogelijk om een lagere LPC-orde p toe te passen dan bij de in het voorgaande beschreven basisbandversie van een RELP-coder, waarbij $p = 16$ is. Wanneer bijvoorbeeld een LPC-orde $p = 12$ wordt gekozen, behoeven slechts 20 12 LAR-coëfficiënten $g(i)$ te worden overgedragen. Bij een zelfde totale capaciteit van 9,6 kbit/s voor transmissiekanaal 3, kan de aanvankelijk voor de transmissie van LAR-coëfficiënten $g(13)$ - $g(16)$ gereserveerde capaciteit van 600 bit/s worden benut voor de transmissie van de tweede predictieparameters c en M , waarvoor in het beschreven 25 voorbeeld een capaciteit van 500 bit/s nodig is. De resterende capaciteit van 100 bit/s kan dan worden benut om 2 extra bits toe te voeren aan het 20 ms frame van het tijdmultiplexsignaal voor de synchronisatie van demultiplexer 22, zodat thans in elk frame van 192 bits 4 bits worden gebruikt voor de framesynchronisatie, waardoor de betrouwbaar- 30 heid van de transmissie wordt vergroot.

- Ter verduidelijking van de werking van de digitale spraakcode- volgens de uitvinding tonen Fig. 3, Fig. 4 en Fig 5 een aantal amplitudespectra en een autocorrelatiefunctie van signalen op verschillende punten van de vocoder in Fig. 1 die alle betrekking hebben op een 35 zelfde stemhebbend spraaksegment van 30 ms. De langs de verticale as aangegeven dB-waarden zijn daarbij steeds gerelateerd aan een zelfde, overigens willekeurig gekozen referentiewaarde.

Diagram a in Fig. 3 toont het amplitudespectrum van het

spraaksegment aan de uitgang van analoog-digitaalomzetter 6 en diagram b toont het amplitudespectrum van het spraakband residusignaal aan de uitgang van eerste invers filter 11. Uit diagram b van Fig. 3 blijkt dat dit spraakband residusignaal een nagenoeg
5 vlakke spectrale omhullende heeft en dat een duidelijke periodiciteit aanwezig is die correspondeert met een grondtoon van ongeveer 195 Hz. Diagram c van Fig. 3 toont de autocorrelatiefunctie $R(n)$ van dit spraakband residusignaal die genormaliseerd is op een waarde $R(0) = 2048$ en die in autocorrelator 31 slechts berekend is in het deelinterval
10 van 2 ms tot 10 ms binnen het interval van 20 ms. De piek van $R(n)$ treedt op voor een waarde van 5,125 ms die correspondeert met een waarde $M = 41$ en een grondtoon van ongeveer 195 Hz, terwijl de coëfficiënt $c = R(M)/2048$ een waarde van ongeveer 0,882 heeft die wordt gequantiseerd tot een waarde $c = 0,875$. In Fig. 4 toont
15 diagram a het amplitudespectrum van het gemodificeerde spraakband residusignaal aan de uitgang van tweede invers filter 28, waarbij de waarden $M = 41$ en $c = 0,875$ zijn gebruikt in predictor 29. Uit een vergelijking van diagram a in Fig. 4 met diagram b in Fig. 3 blijkt duidelijk de onderdrukking van de periodiciteit die correspondeert
20 met de grondtoon van ongeveer 195 Hz. Diagram b in Fig. 4 toont het amplitude spectrum van het basisband residusignaal na de laagdoorlaat-filtering in filter 26 (maar voor de decimatie met een factor 4).

In Fig. 5 toont diagram a het amplitudespectrum van het excitatiesignaal aan de uitgang van interpolator 27 dat wordt ver-
25 kregen na het uitvoeren van de decimatie op het basisband residusignaal van diagram b in Fig. 4 en het vervolgens uitvoeren van de codering, de transmissie, de decodering en de interpolatie (door toevoeging van monsters met amplitude nul). Diagram b in Fig. 5 toont het amplitudespectrum van het gemodificeerde excitatiesignaal aan de
30 uitgang van tweede synthesesefilter 32, waaruit duidelijk blijkt dat de met de grondtoon van ongeveer 195 Hz corresponderende periodiciteit is herïntroduceerd en de correcte harmonische relatie aanwezig is over de gehele spraakband van 0-4 kHz. Tenslotte toont diagram c in Fig. 5 het amplitudespectrum van het gerepliceerde spraakfragment aan
35 de uitgang van eerste synthesesefilter 14.

Door toepassing van de beschreven maatregelen is een basisbandversie van een RELP-coder verkregen die de volgende voordelen bezit:

8400728

- Het optreden van "toonruis" is op effectieve wijze tegengegaan.
- De basisband van het spraaksignaal behoeft niet afzonderlijk behandeld te worden omdat de onderhavige spraakcoder volledig transparant is voor de basisband. Uit formules (1)-(3) en (5)-(7) volgt immers dat voor de serieschakeling van respectievelijk eerste en tweede inverse filters 11, 28 en tweede en eerste synthesesefilters 32, 14 geldt:

$$A(z) \cdot AA(z) \cdot 1/AA(z) \cdot 1/A(z) = 1 \quad (11)$$

- onafhankelijk van de waarden van de predictieparameters $a(i)$, c en M .
- Tweede invers filter 28 heeft een reducerend effect op het dynamiekbereik van het over te dragen basisband residusignaal zodat dit minder gevoelig wordt voor quantisering.
- Ingeval van willekeurige bitfouten in transmissiekanaal 3 neemt de spraakwaliteit slechts geleidelijk af met toenemende bitfoutenkans tot aan een breekpunt, waarbij de verstaanbaarheid snel daalt voor groter bitfoutenkansen. Dit breekpunt ligt ongeveer bij een bitfoutenkans van 1%, maar door toepassing van foutencorrectietechnieken kan dit cijfer verbeterd worden ten koste van enige verhoging in bitfrequentie.
- Zender 1 en ontvanger 2 kunnen op eenvoudige wijze worden geïmplementeerd met behulp van een aantal gangbare digitale signaalprocessors, bijvoorbeeld van het type NEC μ DP 7720, in een bekende parallelconfiguratie waarbij de processors kunnen communiceren via een 8 bits brede databus. Via de seriële interfaces kunnen de processors communiceren met externe componenten zoals de analoog-digitaal en digitaal-analoogconverters 6, 8 en modems die deel uitmaken van transmissiekanaal 3. Verder is een input-output controller geassocieerd met elke processor voor het verkeer over de databus. De microprogramma's voor de controllers en de processors die nodig zijn voor het uitvoeren van de verschillende signaalprocessen die in het voorgaande zijn beschreven, kunnen door een gemiddelde deskundige worden opgesteld met gebruikmaking van de gebruikers-informatie die de fabrikant van de signaalprocessor levert. Om een goede indruk te geven van de complexiteit wordt vermeld dat het signaalprocessor type NEC μ DP 7720 een 28-pin behuizing heeft en ongeveer 1 Watt consumeert, en dat een input-output controller slechts enkele tien-

tallen logische poorten omvat.

5

10

15

20

25

30

35

8400728

1. Digitale spraakcoder met een zender en ontvanger voor de transmissie van in segmenten opgedeelde digitale spraaksignalen, waarbij de zender is voorzien van:
- een eerste LPC-analysator voor het in responsie op het digitale spraaksignaal van elk segment genereren van eerste predictieparameters die de omhullende van het segment-termijnspectrum van dit digitale spraaksignaal karakteriseren,
 - een eerste adaptief invers filter voor het in responsie op het digitale spraaksignaal van elk segment en op de eerste predictieparameters genereren van een spraakband residusignaal dat correspondeert met de predictiefout van dit segment,
 - een decimerend filter voor het in responsie op het spraakband residusignaal genereren van een basisband residusignaal,
 - een codeer-en-multiplexcircuit voor het coderen van de eerste predictieparameters en de golfvorm van het basisband residusignaal en voor het in tijdmultiplex overdragen van de resulterende codesignalen;
- en waarbij de ontvanger is voorzien van:
- een demultiplex-en-decodeercircuit voor het separeren van de overgedragen codesignalen en voor het decoderen van de gesepareerde codesignalen tot de eerste predictieparameters en de golfvorm van het basisband residusignaal,
 - een interpolerende excitatiegenerator voor het in responsie op het basisband residusignaal genereren van een excitatiesignaal dat correspondeert met het spraakband residusignaal,
 - een eerste adaptief synthesefilter voor het in responsie op het excitatiesignaal en de eerste predictieparameters construeren van een replica van het digitale spraaksignaal;
- met het kenmerk, dat de zender verder is voorzien van:
- een tweede LPC-analysator voor het in responsie op het spraakband residusignaal van het eerste adaptieve inverse filter genereren van tweede predictieparameters die de fijnstructuur van het kortetermijn spectrum van dit spraakband residusignaal karakteriseren,
 - een tweede adaptief invers filter voor het in responsie op het spraakband residusignaal en op de tweede predictieparameters genereren van een gemodificeerd spraakband residusignaal dat wordt toegevoerd aan het decimerende filter;

het codeer-en-multiplexcircuit in de zender en het demultiplex-en-decodeercircuit in de ontvanger zijn ingericht voor het verwerken van zowel

de eerste als de tweede predictieparameters; en
de ontvanger verder is voorzien van:

- een tweede adaptief synthesefilter voor het in responsie op het excitatiesignaal van de interpolerende excitatiegenerator en de tweede
5 predictieparameters construeren van een gemodificeerd excitatiesignaal
 dat wordt toegevoerd aan het eerste adaptieve synthesefilter.

2. Digitale spraakcoder volgens conclusie 1, met het kenmerk,
dat de tweede LPC-analysator wordt gevormd door een autocorrelator
voor het genereren van autocorrelatiecoëfficiënten van het spraakband
10 residusignaal en voor het selecteren van de locatie en de waarde van
 de maximale autocorrelatiecoëfficiënt voor vertragingen die groter zijn
 dan de met de orde van de eerste LPC-analysator corresponderende ver-
 traging.

3. Digitale spraakcoder volgens conclusie 2, met het kenmerk,
15 dat de autocorrelator is ingericht om slechts voor vertragingen in het
 tijdinterval tussen 2 ms en 10 ms autocorrelatiecoëfficiënten te
 genereren.

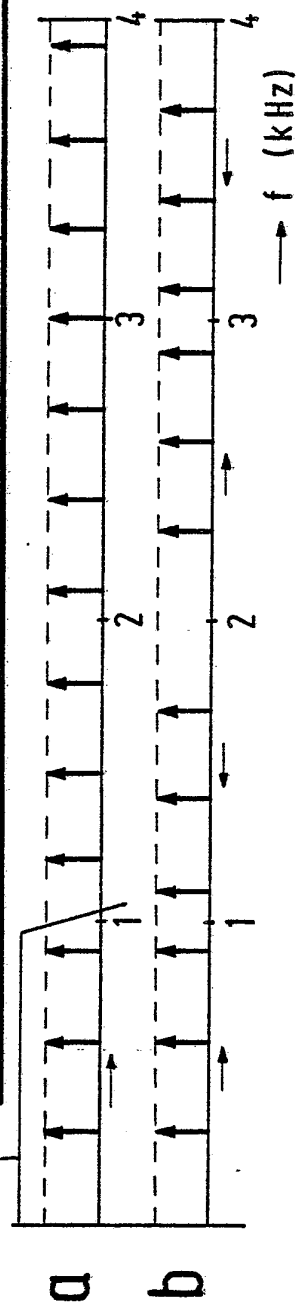
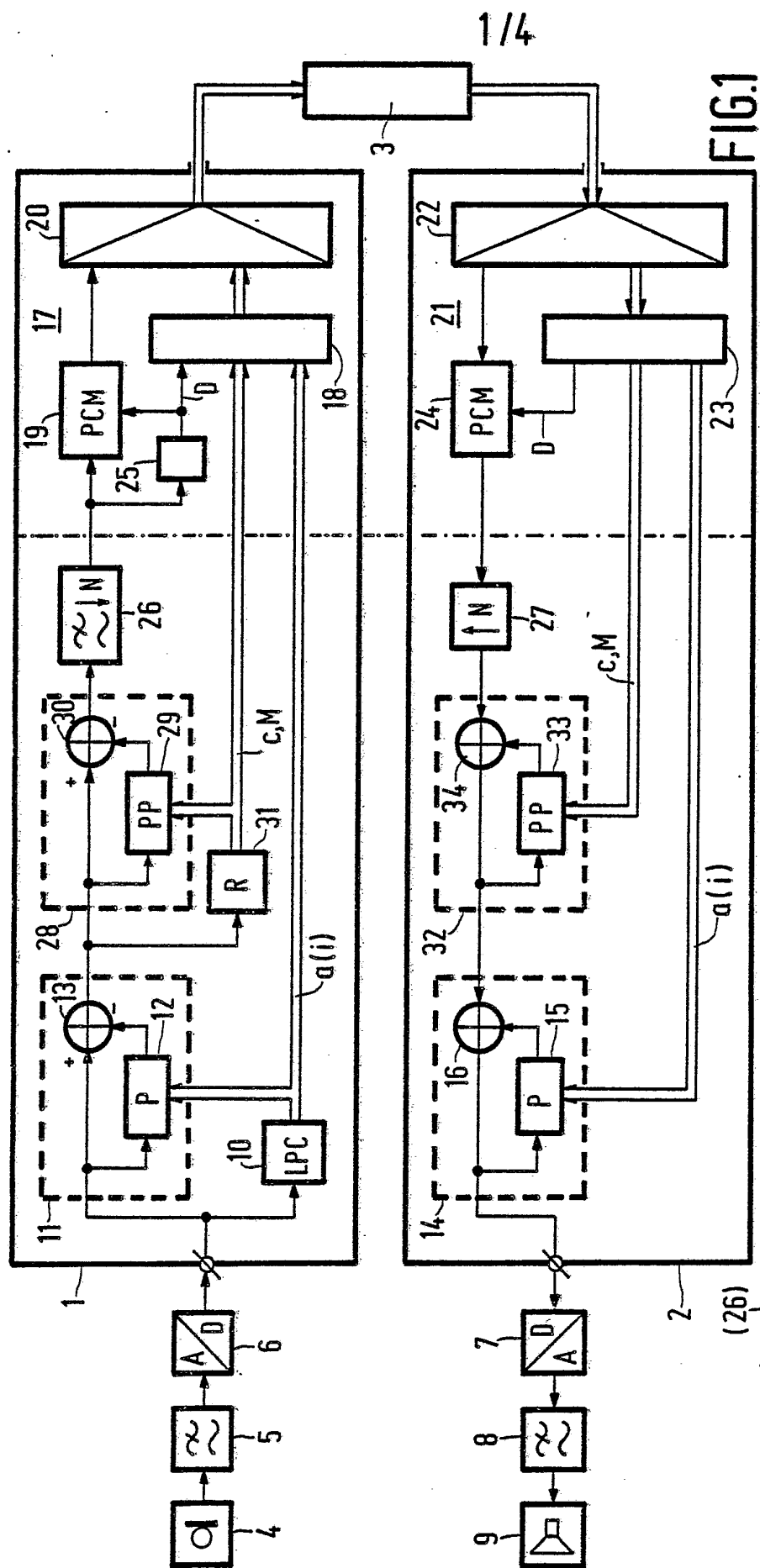
20

25

30

35

8400728



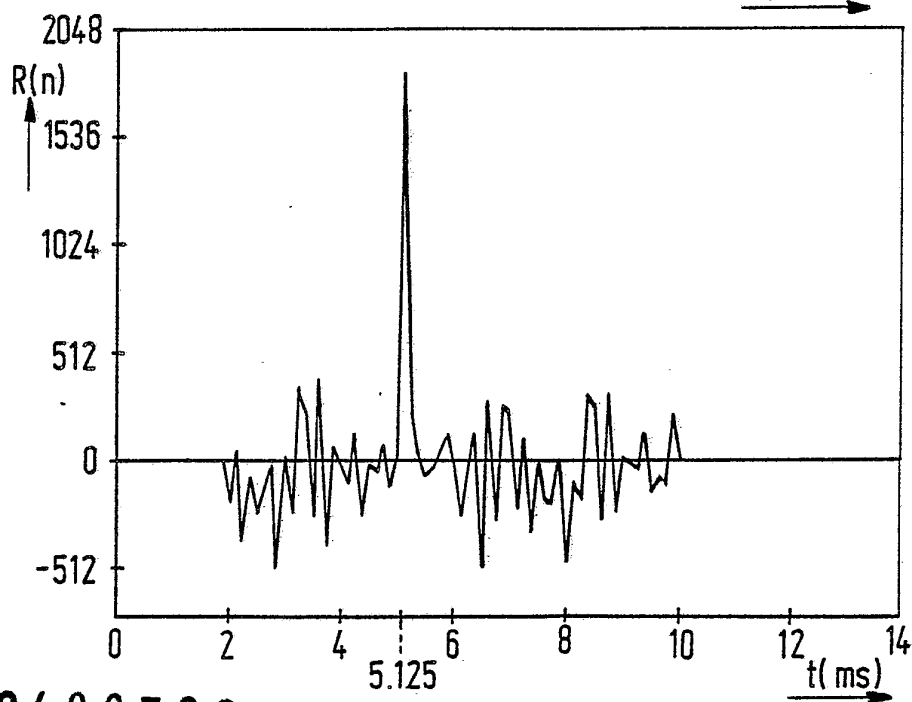
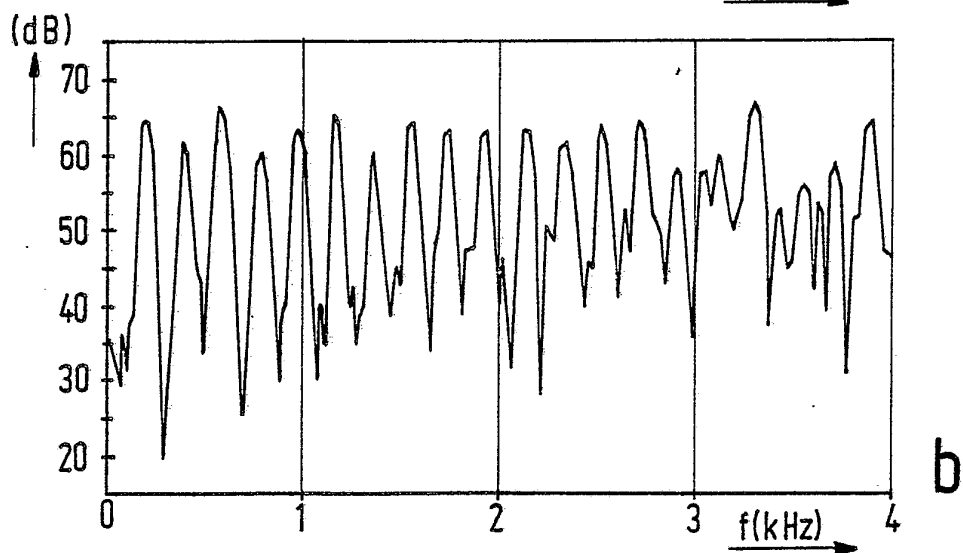
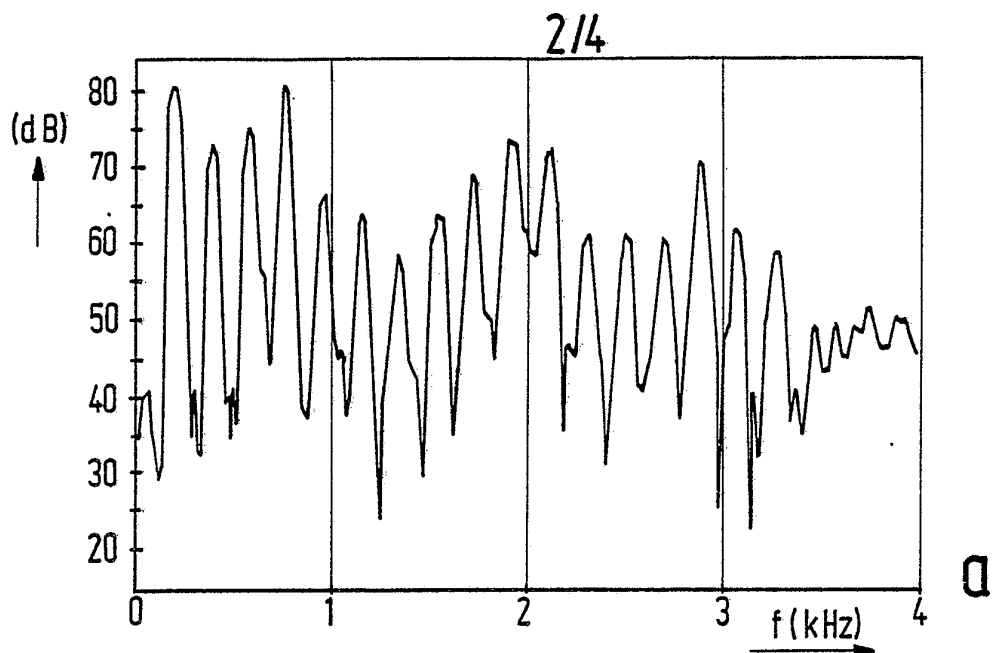
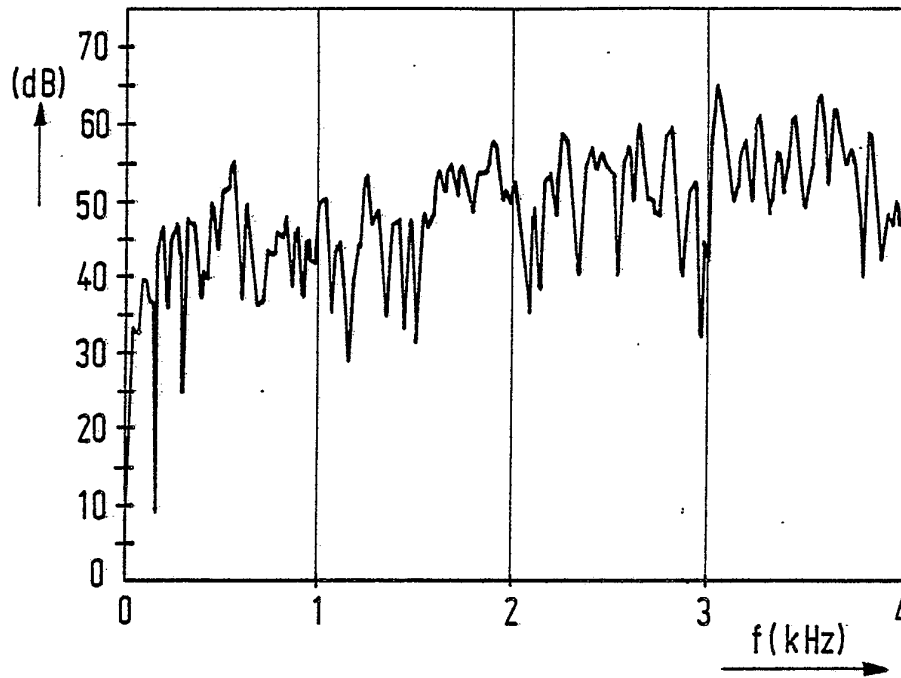


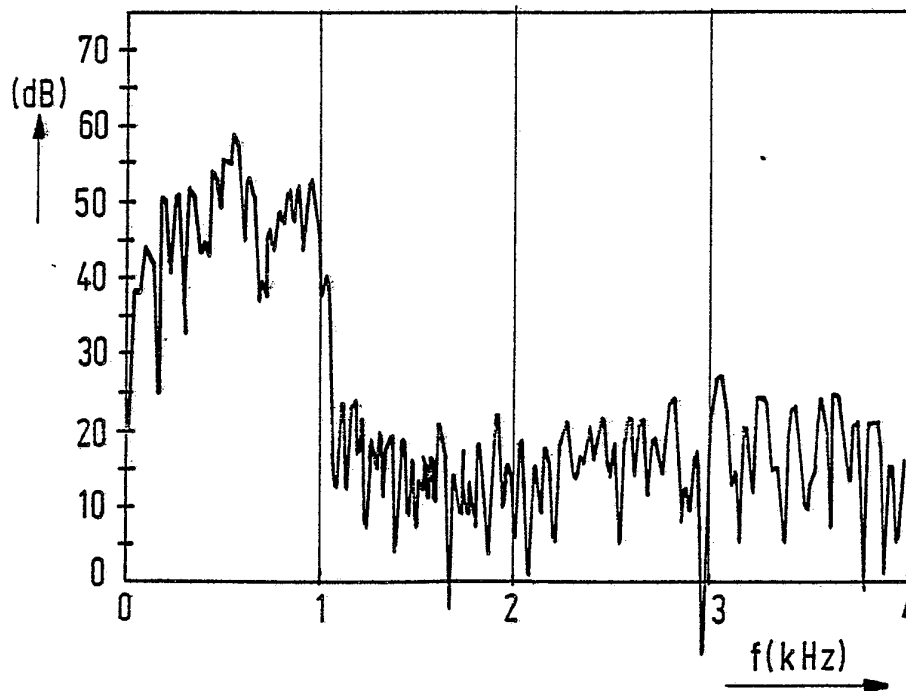
FIG.3

8400728

3/4



a



b

FIG.4

8400728

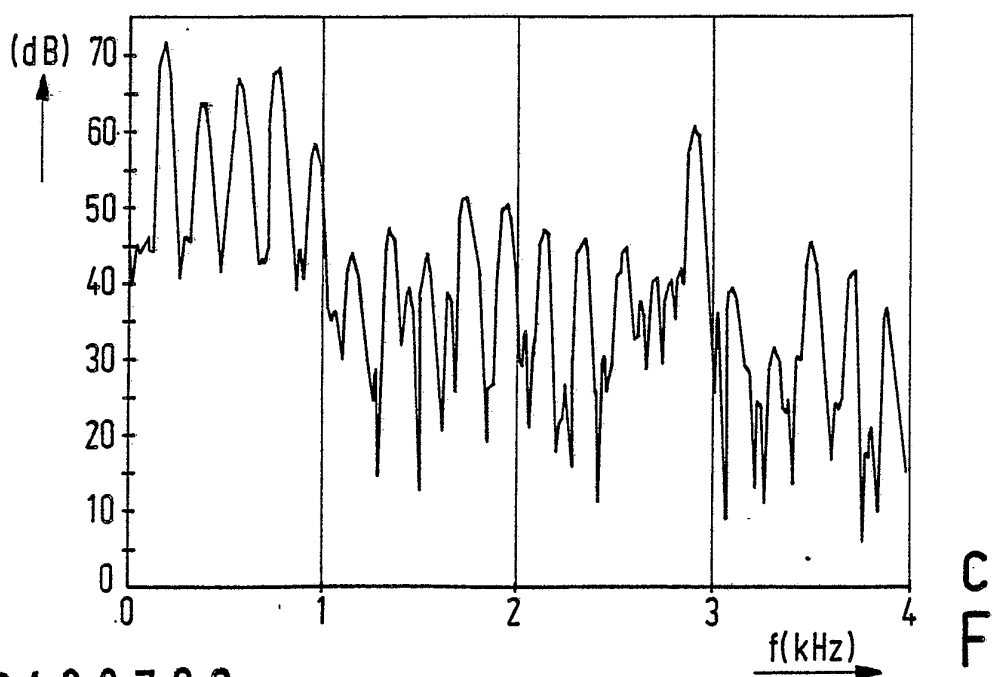
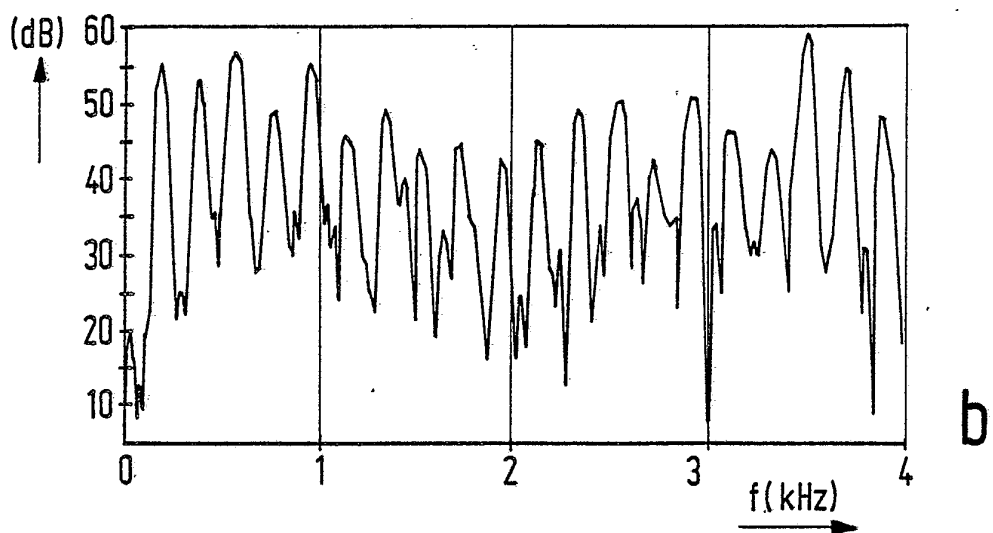
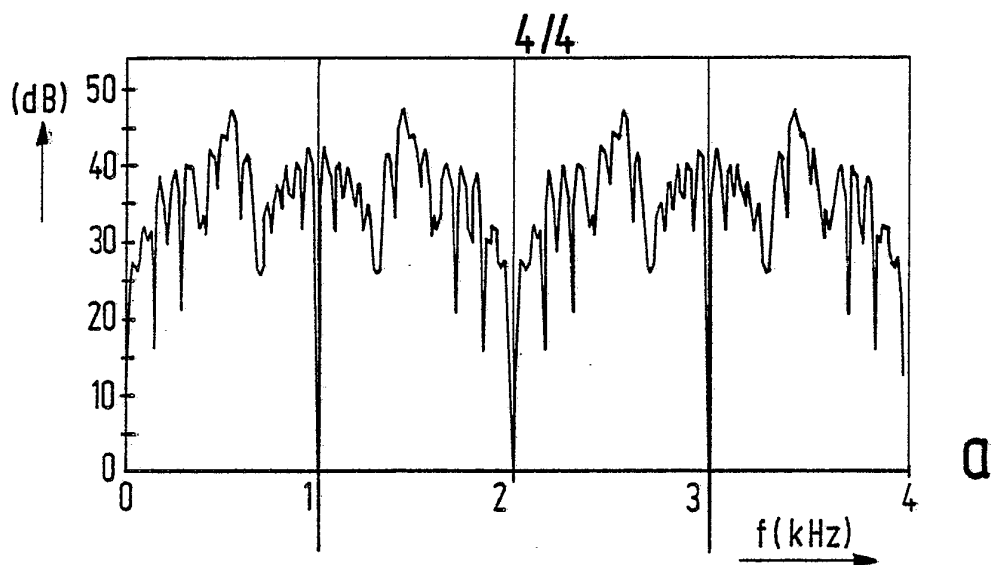


FIG. 5

8400728