



NORGE

(12) PATENT

(19) NO

(11) 301356

(13) B1

(51) Int Cl⁶ G 10 L 9/14

Patentstyret

(21) Søknadsnr	891724	(86) Int. inng. dag og søknadsnummer	26.08.88, PCT/GB88/00708
(22) Inng. dag	26.04.89	(85) Videreføringsdag	26.04.89
(24) Løpedag	26.08.88	(30) Prioritet	28.08.87, GB, 8720389
(41) Alm. tilgj.	26.04.89		15.09.87, GB, 8721667
(45) Meddelt dato	13.10.97		

(73) Patenthaver	British Telecommunications plc, 81 Newgate Street, London EC1A 7AJ, England, GB
(72) Oppfinner	Daniel Kenneth Freeman, Ipswich, Suffolk, England, GB Ivan Boyd, Ipswich, Suffolk, England, GB
(74) Fullmektig	Bryn & Aarflot AS, 0104 OSLO

(54) Benevnelse **Talekoding**

(56) Anførte publikasjoner EP A 195487
Proceedings of the ICASSP 86, Tokyo, 7-11 april 1986 Vol 1, IEEE, Hernandez-Gomez
Proceedings of the ICASSP 87, Dallas, Texas, 4-6 april 1987 Vol 3, IEEE, Lin

(57) Sammendrag **Tale analyseres for å utlede parametrene for et syntese-filter og parametrene for en passende eksitering, utvalgt fra en kodebok med eksiterings-grupper. Utvelgelsen av kodebok-oppføringen forenkles ved å bestemme en enkelt puls-eksitering (f.eks. ved å benytte konvensjonelle "flerpuls"-eksiteringsteknikker) og benytte denne pulsens posisjon til å gjøre området for kodebok-søket smalere. Kodebok-oppføringene kan pålegges den begrensning at enkelte oppføringer er rotasjonsmessig forskjøvne versjoner av andre oppføringer.**

En vanlig teknikk for talekoding er den såkalte "LPC-koding", hvor et inngangs-talesignal deles opp i tidsintervaller i en kodeinnretning, og hvert intervall blir analysert for å bestemme parametre i et syntese-filter med respons som representerer signalets frekvensspektrum i løpet av det aktuelle intervallet. Parametrene overføres til en dekodingsinnretning, hvor de periodisk oppdaterer parametrene for et syntese-filter som, når det mates med et passende eksiterings-signal, frembringer et syntetisk tale-utgangssignal som er en tilnærming av det opprinnelige inngangssignalet.

Det er klart at kodeinnretningen også må overføre til dekoderen informasjon angående egenskapene til den eksitering som skal benyttes. Et antall muligheter er foreslått for å oppnå dette, og disse muligheter faller i to hovedkategorier, som følger:

(i) Resterende eksitert lineær prediktiv koding (RELP), hvor inngangssignalet ledes igjennom et filter som er det inverse av syntese-filtret, for å frembringe et rest-signal som kan kvantiseres og sendes (eventuelt etter filtrering) og benyttes som eksitering, eller kan analyseres, f.eks. for å tilveiebringe stemmings- og tonehøyde-parametre for overføring til en eksiterings-generator i dekoderen.

(ii) Analyse ved hjelp av syntese-metoder hvor en eksitering utledes på en slik måte at ved passasje gjennom syntese-filtret blir forskjellen mellom det oppnådde utgangssignalet og inngangs-talen minimalisert. Innen denne kategori foreligger det to klart forskjellige tilnæringsmåter: Den ene er en flerpuls-eksitering (MP-LPC) hvor en tidsramme eller tidsgruppe som tilsvarende et antall tale-samplere, inneholder et noe mindre og begrenset antall eksiteringspulser som har kodete amplituder og posisjoner. En modifisert form for flerpuls-eksitering er beskrevet i europeisk patentsøknad EP-0195487A, hvor eksiteringspulsene er begrenset til å ligge på ett av en rekke regelmessige gittere med forskjellig fase; gitter-posisjonen kan oppnås ved å rette det inn med den første pulsen som skal bestemmes. Posisjonen for den største enkelte puls i en gruppe benyttes til å bestemme et sett av

posisjoner for hvilke pulsene så beregnes; posisjonen for den største enkelte pulsen benyttes ikke til å forskyve en eksisterende eksitering. Den andre tilnæringsmåten er en stokastisk koding eller en kode-eksitert lineær prediksjon (CELP). Koderen og dekodeeren har hver for seg en lagret liste over standardiserte grupper av eksiteringer. For hver tale-gruppe identifiseres den oppføring i kodeboken som gir syntetisk tale med best tilnærkelse til den faktiske talen når signalet går gjennom syntese-filteret, og et tilknyttet kodeord sendes til dekodeeren, som så kan gjenfinne samme oppføring i sin egen lagrede liste. Slike kodebøker kan utarbeides ved bruk av generering av vilkårlige sekvenser, men en annen variant er den såkalte "spredte vektorer"-kodeboken ("sparse vector"), hvor en gruppe eller ramme bare inneholder et lite antall pulser (eksempelvis 4 eller 5 pulser av 32 mulige posisjoner innen en gruppe). En CELP-koder kan typisk ha en kodebok med 1024 oppføringer.

CELP-kodere er beskrevet i Proceedings of the ECASSP 87, International Conference on Acoustics, Speech and Signal Processing, Dallas, Texas, 6.-9. april 1987, bind 3, side 1354-1357, IEEE, New York, US; D. Lin: "Speech coding using efficient pseudo-stochastic block codes", og også i Proceedings of the ICASSP 86, International Conference on Acoustics, Speech and Signal Processing, Tokyo, 7.-11. april 1986, bind 1, side 469-472, IEEE, New York, US; L. A. Hernandez-Gomez et al: "On the behaviour of reduced complexity code-excited linear prediction" (CELP)". Lin beskriver et tre-strukturert søk for å velge den ønskede kodebok-oppføring, hvor noen oppføringer kan være rotasjonsforskjøvne versjoner av andre oppføringer. Forslaget ifølge Hernandez-Gomez et al innebærer søking i kodeboken ved bruk av et første kriterium for å oppnå en undermengde av oppføringene, og så søking ved bruk av et andre kriterium for å identifisere den ønskede oppføring innen denne undermengden.

I samsvar med foreliggende oppfinnelse er det tilveiebrakt en tale-kodeinnretning som omfatter:

midler anordnet for under drift å generere, ut fra

suksessive tidsgruppe-perioder i inngangs-talesignaler, filterinformasjon som definerer suksessive representasjoner av et syntese-filters respons, og for å utmate filterinformasjonen;

midler anordnet for under drift å motta, for hver enkelt av suksessive tidsgruppe-perioder av talen, inngangstalesignalene og den respektive filterinformasjon, og for å generere eksiteringsinformasjon, omfattende:

(a) et datalager som definerer flere eksiteringsgrupper som hver består av flere pulser;

(b) midler for å utvelge en eksiteringsgruppe fra de nevnte flere grupper og rotasjonsmessig forskjøvnede versjoner av gruppene. Tale-kodeinnretningen ifølge oppfinnelsen kjennetegnes ved at utvelgesmidlene er anordnet for å

- (i) bestemme, i en flerhet av enkeltpuls-grupper som hver består av en enkelt puls i en forskjellige posisjon, hvilken gruppe som tilfredsstillere det kriterium at den når den påtrykkes inngangen på et filter med respons definert av filterinformasjonen, vil frembringe en gruppe med syntetisk tale som ligner mest på gruppen med inngangs-tale;
- (ii) bestemme hvilken av de nevnte flere lagrede grupper som, når den forskyves rotasjonsmessig i en grad som utledes fra den bestemte pulsens posisjon, tilfredsstillere kriteriet; og
- (iii) generere data som identifiserer lager-oppføringen og graden eller størrelsen av dens eventuelle rotasjonsmessige forskyvning.

Andre, valgfrie trekk ved oppfinnelsen er definert i de uselvstendige kravene tilknyttet det vedføyde patentkrav 1.

Noen utførelser av oppfinnelsen skal nå beskrives i eksempelvis form med henvisning til de medfølgende tegningene, hvor

fig. 1 viser den roterende puls-forskyvning som benyttes i foreliggende oppfinnelse,

fig. 2 er et blokkdiagram over en form av en tale-koder ifølge foreliggende oppfinnelse, og

fig. 3 er et blokkdiagram over en passende dekode.

Man vil forstå av det som allerede er sagt at flerpuls-kodere og CELP-kodere av typen "spredte vektorer" har de trekk felles at eksiteringen som benyttes, i begge tilfeller er en gruppe som inneholder et betydelig mindre antall pulser enn antallet tillatte posisjoner i gruppen.

Kodeinnretningen som nå skal beskrives, ligner på en CELP-innretning ved at den benytter en kodebok av "spredte vektorer"-type, som imidlertid er mye mindre enn den som vanligvis benyttes, muligens 32 eller 64 oppføringer. Hver oppføring representerer en eksitering fra hvilken det kan utledes andre elementer i en mengde eksiteringer som skiller seg fra den ene eksitering - og fra hverandre - bare ved en syklisk forskyvning. Tre slike elementer i mengden er vist i figurene 1a, 1b og 1c for en 32-posisjoners gruppe med 5 pulser, hvor man kan se at 1b kan dannes fra 1a ved å forskyve oppføringen syklisk mot venstre, og likeledes kan 1c dannes fra 1a. Graden av forskyvning indikeres i figuren av en to-spisset pil. Syklisk forskyvning betyr at pulsene som forskyves ut på venstre side, går rundt og kommer inn på nytt fra høyre side. Den oppføring som representerer mengden er lagret med den største pulsen i posisjon 1, dvs. som vist på fig. 1d. Størrelsen av den største pulsen trenger ikke å lagres dersom de andre normaliseres ved hjelp av denne.

Dersom antallet kodebok-oppføringer er 32, kan den valgte eksitering representeres ved hjelp av et 5-biters kodeord som identifiserer oppføringen, samt ytterligere 5 biter som angir antallet forskyvninger fra den lagrede posisjonen (dersom alle 32 mulige forskyvninger er tillatt).

Fig. 2 er et blokkdiagram over en tale-kodeinnretning. Talesignaler mottatt på inngangen 1 omvandles til sampler ved hjelp av en samplings-innretning 2, og omgjøres så til digital form i en analog/digital-omvandler 3. En analyse-enhet 4 beregner, for hver påfølgende gruppe av sampler, koeffisientene i et syntese-filter med respons som svarer til talens spektral-innhold. Utledning av LPC-koeffisienter er i og for seg velkjent, og skal ikke beskrives videre her.

Koeffisientene leveres til en utgangs-multiplekser 5, og også til et lokalt syntese-filter 6. Filtrets oppdateringshastighet kan typisk være en gang i løpet av 20 ms.

Kodeinnretningen har også et kodebok-lager 7 som inneholder de 32 kodebok-oppføringer som er omtalt ovenfor. Det er ikke av vesentlig betydning for foreliggende oppfinnelse hvordan disse oppføringene er lagret, men det antas at hver oppføring (for en 5-pulsers eksitering i en gruppe med 32-samplers periode) inneholder posisjonene i gruppen og amplitudene for de fire pulsene etter den første. Når denne informasjonen leses ut fra lageret, leveres den til en eksiteringsgenerator 8 som tilveiebringer en faktisk eksiteringsgruppe - dvs. 32 verdier (av hvilke 27 selvfølgelig er lik 0). Utgangssignalet fra eksiterings-generatoren 8 leveres via en styrbar forskyvningsenhet 9 til inngangen på syntese-filteret 6. Utgangssignalet fra filtret sammenlignes i en subtraherer 10 med inngangs-talesamplene som leveres via en buffer 11 (slik at et antall sammenligninger kan foretas mellom en 32-samplers talegruppe og forskjellige filtrerte eksiteringer).

For å sikre passende forskyvnings-verdi lånes visse teknikker fra flerpuls-koding. I flerpuls-koding er en vanlig metode for å utlede pulsens posisjoner og amplituder en iterativ metode, hvor en puls som minimaliserer feilen mellom den syntetiske og den faktiske talen, beregnes, og en ytterligere puls finnes så, som i kombinasjon med den første, minimaliserer feilen, osv. Analyse av statistikk vedrørende pulser av MP-LPC type viser at den første pulsen som utledes, vanligvis har størst amplitude.

Denne utførelsen av oppfinnelsen gjør nytte av denne teknikk ved å utføre et flerpuls-søk for å finne plasseringen av bare denne første pulsen. Man kan benytte en hvilken som helst av de kjente metodene for dette, f.eks. den som beskrives i B.S. Atal & J.R. Remde, "A New Model of LPC Excitation for producing Natural Sounding Speech at Low Bit rates", Proc. IEEE Int. Conf. ASSP, Paris, 1982, side 614.

En søke-enhet 12 er vist i fig. 2 for dette formål: enhetens utgang mater forskyvningsenheten 9 for å fastlegge

den roterende forskyvning som skal anvendes på eksiteringen som genereres av generatoren 8. Effektivt sett utvelger dette, ut fra 1024 eksiteringer som er tillatt i kodeboken, en særskilt klasse av eksiteringer, nemlig de med den største pulsen i den særskilte posisjonen som bestemmes av søke-enheten 12.

Utgangssignalet fra subtrahereren 10 går til en styrings-enhet som også leverer adresser til lageret 7 og forskyvningsverdier til forskyvningsenheten 9. Formålet med styringsenheten er å fastlegge hvilke av de 32 mulige eksiteringene slik de representeres av den utvalgte klassen, som gir det minste utgangssignal fra subtrahereren (vanligvis den midlere, kvadrerte verdi av forskjellene over en gruppe). Endelig fastlagt oppføring og forskyvning ledes ut i form av et kodeord C og en forskyvningsverdi S til utgangsmultiplekseren 5.

Bestemmelsen i styringsenheten av oppføringen for en gitt talegruppe som er tilgjengelig på utgangen av bufferen 11, er som følger:

- (i) før suksessive kodeord (kodebok-adresser) til lageret 7.
- (ii) tilfør til hver kodebok-oppføring en slik forskyvning at den største pulsen beveges til posisjon som indikeres ut fra "flerpuls"-søket.
- (iii) Overvåk utgangssignalet fra subtrahereren 10 for alle 32 oppføringer for å finne ut hvilken som gir opphav til lavest midlere kvadrerte forskjell.
- (iv) Legg ut kodeordet (kodebok-adressen) og forskyvningsverdien til multiplekseren.

Sammenlignet med en konvensjonell CELP-koder som benytter en kodebok med 1024 oppføringer, er det en liten reduksjon i signal/støy-forholdet som oppnås, på grunn av de begrensninger som er lagt på eksiteringene (dvs. at de faller i 32 gjensidig forskyvbare klasser). Det er imidlertid en reduksjon av kodebokens størrelse, og således av lager-behovet for lageret

7. Dessuten reduseres mengden av beregninger som må utføres av styringsenheten 13 betydelig, siden bare 32 tester må utføres i stedet for 1024.

For å ta hensyn til den ikke helt optimale utvelgelsen som ligger innebygget i "flerpuls-søket", kan den ovennevnte prosessen også innbefatte eksiteringer som forskyves noen posisjoner før og etter den posisjon som finnes i søket.

Dette kan oppnås ved at styringsenheten legger til/ trekker fra passende verdier til den forskyvningsverdi som leveres til forskyvningsenheten 9, slik det indikeres ved hjelp av den stiplede forbindelseslinjen. Siden det filtrerte utgangssignalet med en tidsforskjøvet versjon av en gitt eksitering er en tidsforskjøvet versjon av filtrets respons på den gitte eksiteringen, kan imidlertid disse forskyvningene i stedet utføres ved hjelp av en andre forskyvningsenhet 14 plassert etter syntese-filteret 6. Imidlertid er ikke lenger resultatet korrekt når det opptrer syklisk ombytting: Dette problemet kan løses ved (a) ikke å utføre forskyvninger som forårsaker syklisk ombytting, (b) å utføre forskyvningen, men å tillate pulser å gå tapt heller enn å bli syklisk ombyttet (idet dekoderen informeres), eller (c) å tillate syklisk ombytting, men utføre en korreksjon for å ta hensyn til feilen.

Det gjenstår å beskrive hvordan kodeboken genereres. Denne genereringen kan foregå ved hjelp av gaussiske støyteknikker, på den måte som allerede er foreslått i "Stochastic Coding of Speech Signals at very low Bit Rates", B.S. Atal & M.R. Schroeder, Proc. IEEE Int. Conf. on Communications, 1984, side 1610-1613. En ytterligere fordel kan imidlertid oppnås ved å generere kodeboken ved hjelp av statistisk analyse av de resultater som frembringes av en flerpuls-koder. Dette kan fjerne den tilnærmelse som ligger i antagelsen om at den første pulsen som utledes ved hjelp av "flerpuls-søket" er den største, siden oppføringene i kodeboken da kan lagres med den først oppnådde pulsen i en standard-posisjon, og forskyves slik at denne pulsen bringes til den posisjon som enheten utleder.

Skjønt de forskjellige funksjons-elementene som vises i fig. 2 er vist separat, kan i realiteten noen eller alle av dem utføres ved hjelp av de samme innretningene. En av de kommersielt tilgjengelige integrerte kretser for digital signalbehandling (DSP) kan eksempelvis benyttes dersom den programmeres på passende måte.

Selv om muligheten for "flerpuls-søk" er beskrevet i sammenhengen med forskjøvnede kodebok-oppføringer, kan den også anvendes i andre situasjoner hvor de tillatte eksiteringer kan inndeles i klasser innen hvilke alle eksiteringer har den største, eller mest signifikante, puls i en særkilt posisjon innen gruppen. Posisjonen for den utledede pulsen benyttes så for å velge ut passende klasse, og bare kodebok-oppføringene i den klassen behøver å testes.

Fig. 3 viser en dekodeer for å reprodusere signaler som er kodet ved hjelp av apparatet vist i fig. 2.

En demultiplekser 31 mottar signaler på inngangen 30 og (a) leverer filter-koeffisienter til et syntese-filter 32, (b) leverer kodeord (kodebok-adresser) til adresse-inngangen på et kodebok-lager 33, (c) leverer forskyvningsverdier til en forskyvningsenhet 34 som befordreer utgangssignalet fra en eksiterings-generator 35 forbundet med lageret 33 til inngangen på syntese-filteret 32. Tale-utgangssignalet fra filteret 32 leveres via en digital/analog-omvandler 36 til en utgang 37.

P A T E N T K R A V

1. Tale-kodeinnretning, omfattende:

midler anordnet for under drift å generere, ut fra suksessive tidsgruppe-perioder i inngangs-talesignaler, filterinformasjon som definerer suksessive representasjoner av et syntese-filters respons, og for å utmate filterinformasjonen;

midler anordnet for under drift å motta, for hver enkelt av suksessive tidsgruppe-perioder av talen, inngangstalesignalene og den respektive filterinformasjon, og for å generere eksiteringsinformasjon, omfattende:

(a) et datalager som definerer flere eksiteringsgrupper som hver består av flere pulser;

(b) midler for å utvelge en eksiteringsgruppe fra de nevnte flere grupper og rotasjonsmessig forskjøvne versjoner av gruppene;

k a r a k t e r i s e r t v e d at utvelgelsesmidlene er anordnet for å

- (i) bestemme, i en flerhet av enkeltpuls-grupper som hver består av en enkelt puls i en forskjellige posisjon, hvilken gruppe som tilfredsstiller det kriterium at den når den påtrykkes inngangen på et filter med respons definert av filterinformasjonen, vil frembringe en gruppe med syntetisk tale som ligner mest på gruppen med inngangs-tale;
- (ii) bestemme hvilken av de nevnte flere lagrede grupper som, når den forskyves rotasjonsmessig i en grad som utledes fra den bestemte pulsens posisjon, tilfredsstiller kriteriet; og
- (iii) generere data som identifiserer lager-oppføringen og graden eller størrelsen av dens eventuelle rotasjonsmessige forskyvning.

2. Tale-kodeinnretning ifølge krav 1,

k a r a k t e r i s e r t v e d at de nevnte rotasjonsmessig forskjøvne versjoner består av de lagrede grupper som

hver er forskjøvet i en grad som tilsvarer den bestemte puls-
posisjonen.

3. Tale-kodeinnretning ifølge krav 1,
k a r a k t e r i s e r t v e d at de rotasjonsmessig
forskjøvne versjonene omfatter de lagrede grupper som hver er
forskjøvet i en grad som tilsvarer den bestemte puls-
posisjonen, og de grupper som er utsatt for tilleggs-
forskyvninger som er små i forhold til gruppens størrelse.

4. Tale-kodeinnretning ifølge krav 2 eller 3,
k a r a k t e r i s e r t v e d at den grad av
forskyvning som tilsvarer den bestemte puls-posisjonen, er den
forskyvning som bringer den største puls i eksiteringsgruppen
til samme posisjon innen gruppen som den bestemte enkeltpuls.

5. Tale-kodeinnretning ifølge krav 3 eller 4,
k a r a k t e r i s e r t v e d at hver av de flere
lagrede eksiteringsgrupper er generert av en innstillings-
sekvens som omfatter identifisering av posisjonen inne i en
enkelt puls-gruppe som tilfredsstillter det nevnte kriterium,
for en enkelt første puls, fulgt av bestemmelse av ytterligere
pulser som skal inkluderes i eksiteringsgruppen, og ved at den
nevnte grad av forskyvning som tilsvarer den bestemte
posisjonen, er den forskyvning som bringer den første pulsen i
eksiteringsgruppen til samme posisjon i gruppen som posisjonen
for enkeltpulsen i den enkeltpuls-gruppe som er bestemt av
utvelgelsesmidlene.

Fig.1

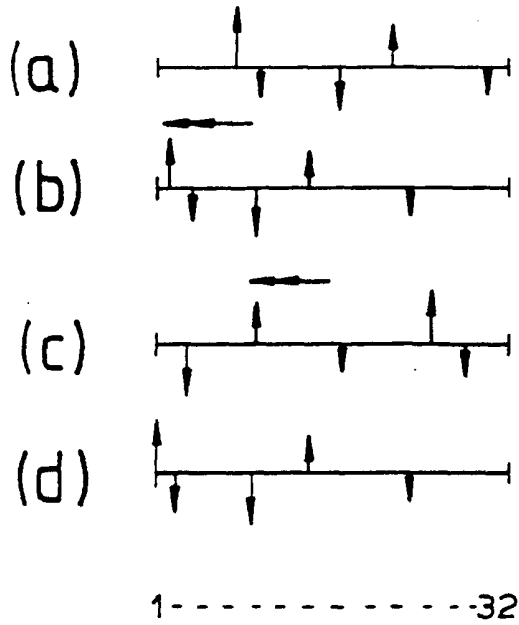


Fig.3

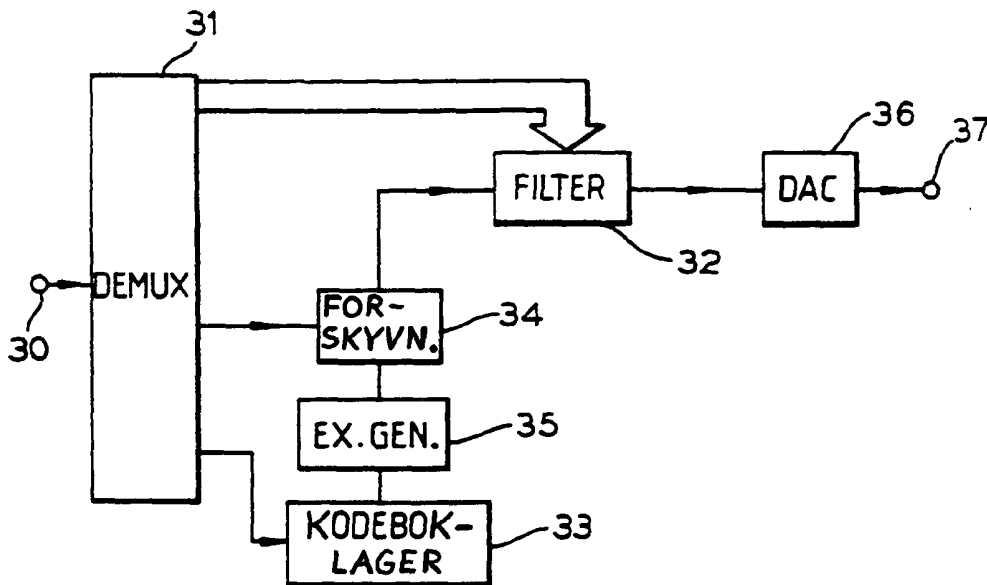


Fig. 2

