

①9



Octrooiraad
Nederland

①1 Publikatienummer: **9201941**

①2 **A TERINZAGELEGGING**

②1 Aanvraagnummer: **9201941**

⑤1 Int.Cl.⁵:
G10L 5/00

②2 Indieningsdatum: **05.11.92**

③0 Voorrang:
06.11.91 KR 9119617

④3 Ter inzage gelegd:
01.06.93 I.E. 93/11

⑦1 Aanvrager(s):
**Korea Telecommunication Authority te Seoul,
Zuid-Korea**

⑦2 Uitvinder(s):
**Chong Rak Lee te Seoul, Zuid-Korea.
Yong Kyu Park te Puchon, Zuid-Korea**

⑦4 Gemachtigde:
**Ir. L.C. de Bruijn c.s.
Nederlandsch Octrooibureau
Scheveningseweg 82
2517 KZ 's-Gravenhage**

⑤4 **Spraaksegmentcoderings- en toonhoogteregelingswerkwijzen voor spraaksynthesestelsels**

⑤7 De onderhavige uitvinding betreft een technologie die betrekking heeft op een spraaksynthesestelsel en maakt gebruik van een periodieke golfvorm-ontledings- en herplaatsingswerkwijze die een coderingswerkwijze is waarin signalen van een stemhebbend geluidsinterval in oorspronkelijke spraak ontleed worden in golven die elk overeenkomen met een spraakgolfvorm voor één periode die gemaakt is door elke glottispuls, waarbij de golven respectievelijk worden gecodeerd en opgeslagen. Het is een technologie waarbij de golven die zich het dichtst bij de posities bevinden waar de golven geplaatst moeten worden uit de opgeslagen golven geselecteerd en gedecodeerd worden en zodanig gesuperponeerd worden dat de oorspronkelijke geluidskwaliteit gehandhaafd kan worden terwijl de duur en toonhoogtefrequentie van het spraaksegment willekeurig kunnen worden ingesteld.

NL A 9201941

De aan dit blad gehechte afdruk van de beschrijving met conclusie(s) en eventuele tekening(en) bevat afwijkingen ten opzichte van de oorspronkelijk ingediende stukken; deze laatste kunnen bij de Octrooiraad op verzoek worden ingezien.

Spraaksegmentcoderings- en toonhoogteregelingswerkwijzen voor spraak-synthesestelsels.

De uitvinding heeft betrekking op een spraaksynthesestelsel en op
 5 een werkwijze voor het synthetiseren van spraak, en meer in het bijzonder op een spraaksegmentcoderings- en een toonhoogteregelingswerkwijze die de kwaliteit van gesynthetiseerde spraak significant verbetert.

Het principe van de onderhavige uitvinding kan niet alleen direct worden toegepast op spraaksynthese maar ook op de synthese van andere
 10 geluiden (klanken), zoals de geluiden van muziekinstrumenten of zang, die elk eigenschappen hebben die vergelijkbaar zijn met die van spraak, of op spraakcodering met een zeer lage snelheid of op spraaksnelheid-conversie. De onderhavige uitvinding zal in het onderstaande met de nadruk op spraaksynthese worden beschreven.

15 Er zijn spraaksynthese-werkwijzen voor het ten uitvoer leggen van een stelsel voor tekst-naar-spraak-synthese, die talloze vocabulaires kunnen synthetiseren door het in spraak omzetten van tekst, d.w.z. reeksen van tekens. De werkwijze die eenvoudig ten uitvoer kan worden gelegd en het meest algemeen wordt toegepast is echter de werkwijze van segmentele spraaksynthese, ook synthese-door-aaneenschakeling-werkwijze
 20 genoemd, waarin menselijke spraak wordt bemonsterd en in fonetische eenheden, zoals half-lettergrepen of tweeklanken, wordt ontleed om korte spraaksegmenten te verkrijgen die vervolgens gecodeerd en in een geheugen opgeslagen worden, en wanneer de tekst wordt ingevoerd, wordt deze in
 25 fonetische transcripties omgezet. Spraaksegmenten die overeenkomen met de fonetische transcripties worden dan sequentieel uit het geheugen teruggehaald en gedecodeerd om de met de invoertekst overeenkomende spraak te synthetiseren.

In dit type segmentele spraaksynthese-werkwijzen is één van de
 30 belangrijkste elementen voor het bepalen van de kwaliteit van de gesynthetiseerde spraak de coderingswerkwijze van de spraaksegmenten. In de bekende segmentele spraaksynthesewerkwijze van een spraaksynthesestelsel wordt hoofdzakelijk een "vocoderings"-werkwijze met een lage spraakkwaliteit gebruikt als de spraakcoderingswerkwijze voor het opslaan van
 35 spraaksegmenten. Dit is echter één van de meest belangrijke oorzaken van de kwaliteitsvermindering van gesynthetiseerde spraak. Een korte beschrijving met betrekking tot de spraaksegment-coderingswerkwijze volgens de stand van de techniek wordt in het onderstaande gegeven.

Een spraakcoderingswerkwijze kan in het algemeen worden ingedeeld

in een golfvorm-coderingswerkwijze met een goede spraakkwaliteit en een vocoderingswerkwijze met een geringe spraakkwaliteit. Aangezien de golfvorm-coderingswerkwijze een werkwijze is die de spraakgolfvorm als zodanig beoogt over te dragen, is het zeer moeilijk om de toonhoogte-frequentie en -duur te veranderen, zodat het onmogelijk is om tijdens het uitvoeren van de spraaksynthese de intonatie en de snelheid van de spraak in te stellen. Het is ook onmogelijk om de spraaksegmenten onderling vloeiend samen te voegen zodat de golfvorm-coderingswerkwijze principieel niet geschikt is voor het coderen van de spraaksegmenten.

Integendeel, wanneer de vocoderingswerkwijze (ook een analyse-synthese-werkwijze genoemd) gebruikt wordt, kunnen het toonhoogtepatroon en de duur van het spraaksegment willekeurig veranderd worden. Aangezien de spraaksegmenten ook vloeiend kunnen worden samengevoegd door de spectrale omhullende-schattingsparameters te interpoleren zodat de vocoderingswerkwijze geschikt is voor de coderingsmiddelen voor tekst-naar-spraak-synthese, worden in de meeste huidige spraaksynthesestelsels vocoderingswerkwijzen toegepast, zoals lineaire predictieve (voorspellende) codering ("LPC") of formant-vocodering. Aangezien de kwaliteit van gecodeerde spraak echter laag is wanneer de spraak gecodeerd wordt met gebruikmaking van de vocoderingswerkwijze kan de gesynthetiseerde spraak, die verkregen is door het decoderen van de opgeslagen spraaksegmenten en het aaneenvoegen daarvan, geen betere spraakkwaliteit hebben dan die welke door de vocoderingswerkwijze wordt geboden.

Tot dusver gedane pogingen om de door de vocoderingswerkwijze geboden spraakkwaliteit te verbeteren vervangen de gebruikte pulstrein door een activeringssignaal dat een minder kunstmatige golfvorm bezit. Een zo'n dergelijke poging omvatte het gebruiken van een golfvorm met een gepiektheid die kleiner was dan die van de puls, bijvoorbeeld een driehoekige golfvorm of een golf in de vorm van een halve cirkel of een golfvorm die vergelijkbaar is met een glottispuls. Een andere poging omvatte het selecteren van een als voorbeeld dienende toonhoogtepuls van één of enkele overblijvende signaaltoonhoogteperioden die verkregen zijn door invers filteren en het in plaats van de puls gebruiken van één voorbeeldpuls voor de gehele tijdsperiode of voor een wezenlijke lange tijdsperiode. Dergelijke pogingen om de puls te vervangen door een activeringspuls met andere golfvormen hebben de spraakkwaliteit niet verbeterd of hebben deze slechts in geringe mate verbeterd, indien dit al het geval is, en hebben nooit gesynthetiseerde spraak opgeleverd met een kwaliteit die in de buurt kwam van natuurlijke spraak.

Het is het oogmerk van de onderhavige uitvinding om spraak te synthetiseren met een hoge kwaliteit en met een natuurlijkheid en een begrijpelijkheid die overeenkomt met menselijke spraak, door het toepassen van een nieuwe spraaksegment-coderingswerkwijze die goede spraak-
 5 kwaliteit- en toonhoogteregeling mogelijk maakt. De werkwijze van de onderhavige uitvinding combineert de voordelen van de golfvormcoderingswerkwijze, die goede spraakkwaliteit levert maar niet de mogelijkheid heeft om de toonhoogte te regelen, en de vocoderingswerkwijze, die toonhoogteregeling biedt maar een lage spraakkwaliteit bezit.

10 De onderhavige uitvinding maakt gebruik van een periodieke- golfvorm-ontledingswerkwijze, hetgeen een coderingswerkwijze is die een signaal in een stemhebbende geluidssector in de originele spraak ontleedt in golven ("wavelets") die equivalent zijn aan de spraakgolfvormen van één periode die door glottispulsen gemaakt worden, om het ontlede signaal te
 15 coderen en op te slaan, en een op tijdvervorming gebaseerde golfverplaatsingswerkwijze, die een golfvorm-synthese-werkwijze is die in staat is de duur en de toonhoogtefrequentie van het spraaksegment willekeurig in te stellen, terwijl de kwaliteit van de oorspronkelijke spraak behouden blijft, door het selecteren van golven die zich het dichtst bij posities
 20 bevinden waar golven tussen opgeslagen golven geplaatst moeten worden, en door het vervolgens decoderen van de geselecteerde golven en het superponeren daarvan. Ten behoeve van deze uitvinding worden muziekgeluiden behandeld als stemhebbende geluiden.

De voorgaande doeleinden moeten zodanig worden opgevat, dat zij
 25 slechts enkele van de meer relevante eigenschappen en toepassingen van de uitvinding vormen. Vele andere gunstige resultaten kunnen verkregen worden door de geopenbaarde uitvinding op een verschillende wijze toe te passen of de uitvinding binnen het kader van de beschrijving te wijzigen. Andere doeleinden en een beter begrip van de uitvinding kan derhalve
 30 worden verkregen door te verwijzen naar zowel de onderstaande samenvatting van de uitvinding als de gedetailleerde beschrijving, die de uitvoeringsvorm van de uitvinding beschrijven, in aanvulling op de reikwijdte van de uitvinding die bepaald wordt door de conclusies, samen- genomen met de bijgevoegde tekeningen.

35 Spraaksegmentcoderings- en toonhoogteregelingswerkwijzen voor spraaksynthesestelsels van de onderhavige uitvinding worden gedefinieerd door de conclusies, waarbij specifieke uitvoeringsvormen in de bijgevoegde tekeningen zijn getoond. Teneinde de uitvinding samen te vatten heeft de uitvinding betrekking op een werkwijze die het mogelijk maakt

spraak te synthetiseren die de kwaliteit van natuurlijke spraak benadert door de duur en de toonhoogtefrequentie daarvan in te stellen door golven van elke periode qua golfvorm te coderen, deze in een geheugen op te slaan, en deze op het moment van de synthese te decoderen en op geschikte
 5 tijdstippen te plaatsen zodat zij het gewenste toonhoogtepatroon bezitten, en hen vervolgens te superponeren om natuurlijke spraak, zang, muziek en dergelijke op te wekken.

De onderhavige uitvinding omvat een spraaksegment-coderingswijze voor gebruik met een spraaksynthesestelsel, waarbij de werkwijze
 10 het vormen van golven omvat door het verkrijgen van parameters die in elk analyse-tijdsinterval een spectrale omhullende weergeven. Dit wordt uitgevoerd door een periodiek of quasi-periodiek digitaal signaal, zoals stemhebbende spraak, te analyseren met de spectrumschattingstechniek. Een oorspronkelijk signaal wordt eerst gedeconvolveerd tot een impulsresponsie
 15 die wordt weergegeven door de spectrale-omhullende-parameters en een periodiek of quasi-periodiek toonhoogte-pulstreinsignaal met een vrijwel vlakke spectrale omhullende. Een activeringssignaal, dat verkregen wordt door het toevoegen van bemonsteringen met de waarde nul achter een toonhoogte-pulssignaal van één periode, dat verkregen is door het toonhoogte-pulstreinsignaal periode voor periode te segmenteren zodat in elke perio-
 20 de een toonhoogtepuls is bevat, en een impulsresponsie die overeenkomt met een verzameling van spectrale omhullende parameters in hetzelfde tijdsinterval als het activeringssignaal, worden geconvolveerd om voor die periode een golf te vormen.

25 De golven kunnen, in plaats van van te voren gevormd te worden door golfvormcodering en in een geheugen te worden opgeslagen, gevormd worden door het samenbrengen van informatie, die verkregen is door het golfvorm-coderen van een toonhoogte-pulssignaal van elk periode-interval, verkregen door segmentatie, met informatie die verkregen is door coderen van
 30 een verzameling van spectrale-omhullende-schattingsparameters met hetzelfde tijdsinterval als de bovengenoemde informatie, of met een impulsresponsie die overeenkomt met de parameters en het in een geheugen opslaan van de golfinformatie. Er zijn twee werkwijzen voor het produceren van gesynthetiseerde spraak door gebruik te maken van de in het
 35 geheugen opgeslagen golfinformatie. De eerste werkwijze is het vormen van elke golf door het convolveren van een activeringssignaal, dat verkregen is door het toevoegen van bemonsteringen met de waarde nul aan een toonhoogte-pulssignaal van één periode, verkregen door het decoderen van de informatie, en een impulsresponsie die overeenkomt met de gedecodeerde

spectrale-omhullende-parameters in hetzelfde tijdsinterval als het activeringssignaal, en het vervolgens toewijzen van de golven aan geschikte tijdstippen zodat zij een gewenst toonhoogtepatroon en duurpatroon bezitten, het op de tijdstippen plaatsen van hen, en het superponeren van hen.

5 De tweede werkwijze is het vormen van een synthetisch activerings-signaal door het op zodanige wijze toewijzen van de toonhoogte-pulssignalen, die verkregen zijn door het decoderen van de golfvorminformatie, aan geschikte tijdstippen, dat zij een gewenst toonhoogtepatroon en duurpatroon bezitten en het plaatsen van hen op de tijdstippen, en het vormen
10 van een verzameling van synthetische spectrale omhullende-parameters ofwel door het in de tijd comprimeren of expanderen van de verzameling van tijdfuncties van de parameters op basis van deelsegment-voor-deelsegment, afhankelijk van de vraag of de duur van een deelsegment in een te synthetiseren spraaksegment respectievelijk kleiner of groter is dan die
15 van een overeenkomstig deelsegment in het oorspronkelijke spraaksegment, of door de verzameling van tijdfuncties van de parameters van één periode synchroon met het daarmee samengebrachte toonhoogte-pulssignaal van één periode, dat geplaatst is voor het vormen van het gesynthetiseerde activeringssignaal, en om het synthetische activeringssignaal te convolueren
20 met een impulsresponsie die overeenkomt met de verzameling van synthetische spectrale omhullende-parameters door gebruikmaking van een in de tijd variërend filter of door gebruikmaking van een op de FFT (snelle Fourier-transformatie) gebaseerde snelle convolutietechniek. In de laatste methode treedt een leeg interval op wanneer een gewenste toonhoogteperiode langer is dan de oorspronkelijke toonhoogteperiode en vindt een
25 overlappingsinterval plaats wanneer de gewenste toonhoogteperiode korter is dan de oorspronkelijke toonhoogteperiode.

In het overlappingsinterval wordt het synthetische activeringssignaal verkregen door de overlappende toonhoogte-pulssignalen bij elkaar op
30 te tellen of door één van hen te selecteren, en de spectrale omhullende-parameter wordt verkregen door één van de overlappende spectrale omhullende-parameters te selecteren of door een gemiddelde waarde van de twee overlappende parameters te gebruiken.

In het lege interval wordt het synthetische activeringssignaal
35 verkregen door dit te vullen met bemonsteringen met de waarde nul, en de synthetische spectrale-omhullende-parameter wordt verkregen door de waarden van de spectrale-omhullende-parameters bij de begin- en eindpunten van de voorgaande en volgende perioden voor en na het midden van het lege interval te herhalen, of door één van de twee waarden of een gemiddelde

waarde van de twee waarden te herhalen, of door dit te vullen met waarden en de twee waarden op vloeiende wijze te verbinden.

De onderhavige uitvinding omvat verder een toonhoogteregelingswerk-
wijze voor een spraaksynthesestelsel, die het mogelijk maakt de duur en
5 toonhoogte van een spraaksegment te regelen door een op tijdsvervorming
gebaseerde golfverplaatsingswerkwijze die het mogelijk maakt om spraak te
synthetiseren met vrijwel dezelfde kwaliteit als natuurlijke spraak, door
belangrijke grenstijdstippen zoals het startpunt, het eindpunt en de
vaste punten in een spraaksegment en toonhoogte-pulsposities van elke
10 golf of van elk toonhoogte-pulssignaal te coderen en tegelijkertijd in
een geheugen op te slaan op het moment dat elk spraaksegment wordt opge-
slagen, en op het moment van de synthese een tijdvervormingsfunctie te
verkrijgen door gewenste grenstijdstippen te vergelijken met oorspronke-
lijke opgeslagen grenstijdstippen die overeenkomen met de gewenste grens-
15 tijdstippen, het bepalen van de oorspronkelijke tijdstippen die overeen-
komen met elke gewenste toonhoogte-pulspositie door gebruikmaking van de
tijdvervormingsfunctie, het selecteren van golven met toonhoogte-pulspo-
sities die zich het dichtst bij de oorspronkelijke tijdposities bevinden
en het plaatsen van hen op geschikte toonhoogte-pulsposities, en het
20 superponeren van de golven.

De toonhoogteregelingswerkwijze kan verder het produceren van syn-
thetische spraak omvatten door het selecteren van toonhoogte-pulssignalen
van één periode en spectrale omhullende-parameters die overeenkomen met
de toonhoogte-pulssignalen, in plaats van de golven, en het plaatsen
25 daarvan en het convolueren van de geplaatste toonhoogte-pulssignalen en
impulsresponsie die overeenkomt met de spectrale omhullende-parameters om
golven te produceren, en het superponeren van de geproduceerde golven, of
het convolueren van een synthetisch activeringssignaal, verkregen door
het superponeren van de geplaatste toonhoogte-pulssignalen, en een in de
30 tijd variërende impulsresponsie die overeenkomt met synthetische spectra-
le omhullende-parameters die gemaakt zijn door het aaneenvoegen van de
geplaatste spectrale omhullende-parameters.

Een stemhebbende spraak-syntheseinrichting van een spraaksynthese-
stelsel wordt beschreven en omvat een decoderings-deelblok 9 dat golf-
35 informatie produceert door het decoderen van golfcodes van het spraakseg-
ment-opslagblok 5. Een duurregelings-deelblok 10 produceert tijdvervor-
mingsgegevens uit invoer van duurgegevens uit een prosodie-opwekkings-
deelsysteem 2 en grenstijdstippen die zijn opgenomen in kopinformatie van
het spraaksegmenten-opslagblok 5. Een toonhoogteregelings-deelblok 11

produceert op zodanige wijze toonhoogte-pulspositie-informatie dat dit een intonatiepatroon heeft zoals is aangegeven door intonatiepatroon-gegevens van invoer van de kopinformatie van het spraaksegment-opslagblok 5, de intonatiepatroon-gegevens van het prosodie-opwekkings-deelstelsel 5 en de tijdvervormingsinformatie van het duurregelings-deelblok 10. Een energieregelingsdeelblok 12 produceert versterkingsinformatie zodat gesynthetiseerde spraak het klemtoonpatroon heeft zoals aangegeven wordt door klemtoonpatroongegegevens uit invoer van de klemtoonpatroongegegevens van het prosodie-opwekkingsdeelstelsel 2, de tijdvervormingsinformatie 10 van het duurregelings-deelblok 10 en toonhoogte-pulspositie-informatie van het toonhoogte-regelings-deelblok 11. Een golfvorm-samenstellings-deelblok 13 produceert een stemhebbend spraaksignaal uit invoer van de golf informatie van het decoderings-deelblok 9, de tijdvervormingsinformatie van het duurregelings-deelblok 10, de toonhoogte-pulspositie-informatie 15 van het toonhoogteregelings-deelblok 11 en de versterkingsinformatie van het energieregelingsdeelblok 12.

Volgens de onderhavige uitvinding wordt tekst dus ingevoerd in het fonetische voorverwerkings-deelstelsel 1, waar deze wordt omgezet in fonetische transcriptiesymbolen en syntactische analysegegevens. De syntactische analysegegevens worden uitgevoerd naar een prosodie-opwekkings-deelstelsel 2. Het prosodie-opwekkings-deelstelsel 2 voert prosodie-informatie uit naar het spraaksegment-aaneenschakelings-deelstelsel 3. De fonetische transcriptiesymbolen die zijn uitgevoerd uit het voorverwerkings-deelstelsel, worden ook ingevoerd in het spraaksegment-aaneenschakeling-deelstelsel 3. De fonetische transcriptiesymbolen worden dan ingevoerd in het spraaksegment-selectieblok 4 en de overeenkomstige prosodie-gegevens worden ingevoerd in het stemhebbende stemgeluid-syntheseblok 6 en in het niet-stemhebbend-geluid-syntheseblok 7. In het spraaksegment-selectieblok 4 wordt elk ingevoerd transcriptiesymbool vergeleken met een 25 overeenkomstige spraaksegment-synthese-eenheid en een geheugenadres van de passende synthese-eenheid, dat overeenkomt met elk ingevoerd fonetisch transcriptiesymbool, wordt in het spraaksegment-opslagblok 5 bepaald uit een spraaksegmenttabel. Het adres van de passende synthese-eenheid wordt dan uitgevoerd naar het spraaksegment-opslagblok 5, waar het overeenkomstige spraaksegment in gecodeerde golfvorm geselecteerd wordt voor elk 30 van de adressen van de bij elkaar passende synthese-eenheden. Het geselecteerde spraaksegment wordt in de vorm van een gecodeerde golf uitgevoerd naar het stemhebbend-geluid-syntheseblok 6 voor stemhebbend-geluid en naar het niet-stemhebbend-geluid-syntheseblok 7 voor niet-stemhebbend

geluid. Het stemhebbend-geluid-syntheseblok 6, dat de tijdvervormings-
 golfverplaatsingswerkwijze gebruikt om spraakgeluid te synthetiseren, en
 het niet-stemhebbend-geluid-syntheseblok 7 geven digitale synthetische
 spraaksignalen af aan de digitaal-analoog-omzetter voor het omzetten van
 5 de ingevoerde digitale signalen in analoge signalen die de gesyntheti-
 seerde spraakgeluiden zijn.

Voor het toepassen van de onderhavige uitvinding wordt spraak en/of
 muziek eerst op magneetband opgenomen. Het resulterende geluid wordt dan
 van analoge signalen omgezet in digitale signalen door de analoge signa-
 10 len met een laagdoorlaatfilter te filteren en dan de gefilterde signalen
 toe te voeren aan een analoog-digitaal-omzetter. De resulterende gedigi-
 taliseerde spraaksignalen worden dan gesegmenteerd tot een aantal spraak-
 segmenten met geluiden die overeenkomen met synthese-eenheden, zoals
 fonemen, tweeklanken, halflettergrepen en dergelijke, door gebruik te
 15 maken van bekende spraakbewerkingshulpmiddelen. Elk resulterende spraak-
 segment wordt dan gescheiden in stemhebbende en niet-stemhebbende segmen-
 ten met gebruikmaking van bekende stemhebbende-/niet-stemhebbende-
 detectie- en spraakbewerkingshulpmiddelen. De niet-stemhebbende spraak-
 segmenten worden gecodeerd door bekende vocoderingswerkwijzen die witte
 20 toevalsruis toepassen als een bron van niet-stemhebbende spraak. De voco-
 deringswerkwijzen omvatten LPC, homomorfe formant-vocoderingswerkwijzen,
 en dergelijke.

De stemhebbende spraaksegmenten worden gebruikt om golven $s_j(n)$ te
 vormen overeenkomstig de werkwijze die in het onderstaande in figuur 4 is
 25 geopenbaard. De golven $s_j(n)$ worden dan gecodeerd door gebruikmaking van
 een geschikte golfvorm-coderingswerkwijze. Bekende golfvorm-coderings-
 werkwijzen omvatten pulscodemodulatie (PCM), adaptieve verschil-puls-
 codemodulatie (ADPCM), adaptieve predictieve (voorspellende) codering
 (APC) en dergelijke. De resulterende gecodeerde stemhebbende spraakseg-
 30 menten worden opgeslagen in het spraaksegment-opslagblok 5, zoals in de
 figuren 6A en 6B is getoond. De gecodeerde niet-stemhebbende spraakseg-
 menten worden ook opgeslagen in het spraaksegment-opslagblok 5.

De meer relevante en belangrijke eigenschappen van de onderhavige
 uitvinding zijn in het bovenstaande in grote trekken weergegeven zodat de
 35 hiernavolgende gedetailleerde beschrijving van de uitvinding beter begre-
 pen zal worden en zodat de onderhavige bijdrage aan de techniek ten volle
 geapprecieerd kan worden. Aanvullende eigenschappen van de hierna
 beschreven uitvinding vormen het onderwerp van de conclusies van de uit-
 vinding. Deskundigen zullen inzien dat het idee en de specifieke uitvoe-

ringsvorm, zoals hier beschreven, eenvoudig gebruikt kan worden als basis voor het wijzigen of het ontwerpen van andere structuren voor het uitvoeren van dezelfde doeleinden als de onderhavige uitvinding. Verder kunnen deskundigen zich realiseren dat dergelijke equivalente constructies niet
 5 buiten het kader en de omvang van de uitvinding vallen, zoals deze in de conclusies is uiteengezet.

Voor een beter begrip van de aard en de oogmerken van de uitvinding wordt verwezen naar de hiernavolgende gedetailleerde beschrijving, samen- genomen met de bijgevoegde tekeningen, waarin:

10 Figuur 1 het tekst-naar-spraak-synthesestelsel van de spraak- segment-synthesewerkwijze weergeeft;

 Figuur 2 het spraaksegment-aaneenschakelingsdeelstelsel weergeeft;

 Figuren 3A tot en met 3T golfvormen weergeven ten behoeve van de uitleg van het principe van de periodieke golfvorm-ontledingswerkwijze en
 15 de golfverplaatsingswerkwijze volgens de onderhavige uitvinding;

 Figuur 4 een blokdiagram weergeeft ten behoeve van de uitleg van de periodieke golfvorm-ontledingswerkwijze;

 Figuren 5A tot en met 5E blokdiagrammen weergeven ten behoeve van de uitleg van de procedure van de blinde deconvolutiewerkwijze;

20 Figuren 6A en 6B code-opmaken weergeven voor de in het spraak- segment-opslagblok opgeslagen stemhebbende spraaksegment-informatie;

 Figuur 7 het stemhebbende spraaksyntheseblok volgens de onderhavige uitvinding weergeeft; en

 Figuren 8A en 8B grafieken weergeven ten behoeve van de uitleg van
 25 de duur en toonhoogte-regelingswerkwijze volgens de onderhavige uitvin- ding.

In de verschillende aanzichten van de tekeningen verwijzen overeen- komstige verwijzingstekens naar overeenkomstige delen.

De structuur van het tekst-naar-spraak-synthesestelsel van de
 30 spraaksegment-synthesewerkwijze volgens de stand van de techniek bestaat uit drie deelstelsels:

- A. een fonetisch voorbewerkingsdeelstelsel (1);
- B. een prosodie-opwekkingsdeelstelsel (2); en
- C. een spraaksegment-aaneenschakelingdeelstelsel (3) zoals
 35 getoond in figuur 1.

Wanneer de tekst van een toetsenbord, een computer of enig ander stelsel is ingevoerd in het tekst-naar-spraak-synthesestelsel analyseert het fonetische voorbewerkings-deelstelsel (1) de syntaxis van de tekst en verandert de tekst dan in een reeks fonetische transcriptiesymbolen door

daarop fonetische hercoderingsregels toe te passen. Het prosodie-opwekkingsdeelstelsel (2) wekt intonatiepatroongegevens en klemtoonpatroongegevens op met gebruikmaking van syntactische analysegegevens, zodat geschikte intonatie en klemtoon kunnen worden toegepast op de reeks van
 5 fonetische transcriptiesymbolen, en voert dan de gegevens toe aan het spraaksegment-aaneenschakeling-deelstelsel (3). Het prosodie-opwekkingsdeelstelsel (2) verschaft ook de gegevens met betrekking tot de duur van elk foneem aan het spraaksegment-aaneenschakelingdeelstelsel (3).

De bovenstaande drie soorten prosodische gegevens, d.w.z. de intonatiepatroongegevens, de klemtoonpatroongegevens en de gegevens met
 10 betrekking tot de duur van elk foneem, worden in het algemeen samen met de door het fonetische voorbewerkingsdeelstelsel (1) gegenereerde reeks fonetische transcriptiesymbolen naar het spraaksegment-aaneenschakelingsdeelstelsel (3) gezonden, hoewel zij onafhankelijk van de reeks fonetische transcriptiesymbolen naar het spraaksegment-aaneenschakelingsdeelstelsel (3) kunnen worden overgedragen.
 15

Het spraaksegment-aaneenschakelingsdeelstelsel (3) wekt continue spraak op door het sequentieel ophalen van geschikte spraaksegmenten, die gecodeerd zijn en opgeslagen zijn in het geheugen daarvan, in overeenstemming met de reeks fonetische transcriptiesymbolen (niet getoond), en
 20 door het decoderen daarvan. Op dit moment kan het spraaksegment-aaneenschakelingsdeelstelsel (3) synthetische spraak genereren met de intonatie, klemtoon en spraaksnelheid zoals bedoeld is door het prosodische opwekkingsdeelstelsel (2) door het regelen van de energie (intensiteit), de duur en de toonhoogteperiode van elk spraaksegment, in overeenstemming met de prosodische informatie.
 25

De onderhavige uitvinding verbetert de spraakkwaliteit in vergelijking met gesynthetiseerde spraak volgens de stand van de techniek op opmerkelijke wijze door het verbeteren van de coderingswerkwijze voor het
 30 opslaan van de spraaksegmenten in het spraaksegment-aaneenschakelingsdeelstelsel (3). Een beschrijving met betrekking tot de werking van het spraaksegment-aaneenschakelingsdeelstelsel (3) onder verwijzing naar figuur 3 volgt in het onderstaande.

Wanneer de reeks van fonetische transcriptiesymbolen, gevormd door
 35 het fonetische voorbewerkingsdeelstelsel (1), wordt ingevoerd in het spraaksegment-selectieblok (4), selecteert het spraaksegment-selectieblok (4) sequentieel de synthese-eenheden, zoals tweeklanken en halflettergrepen, door continu de reeks van binnenkomende fonetische transcriptiesymbolen te inspecteren, en bepaalt de adressen van de spraaksegmenten

die overeenkomen met de geselecteerde synthese-eenheden uit het geheugen daarvan, zoals in tabel 1 is weergegeven. Tabel 1 toont een voorbeeld van de spraaksegmenttabel die ligt opgeslagen in het spraaksegment-selectieblok (4) dat op tweeklanken gebaseerde spraaksegmenten selecteert. Dit
 5 resulteert in het vormen van een adres van het geselecteerde spraaksegment dat wordt uitgevoerd naar het spraaksegment-opslagblok (5).

De met de adressen van het spraaksegment overeenkomende spraaksegmenten worden gecodeerd volgens de werkwijze van de onderhavige uitvinding, die later wordt beschreven, en worden opgeslagen op de adressen van
 10 het geheugen van het spraaksegment-opslagblok (5).

Tabel 1

	fonetisch transcriptie- symbool van spraaksegment	geheugenadres (hexadecimaal)
15		
	/ai/	0000
	/au/	0021
	/ab/	00A3
20	/ad/	00FF
	.	.
	.	.
	.	.
25		

Wanneer het adres van het geselecteerde spraaksegment uit het spraaksegment-selectieblok (4) wordt ingevoerd in het spraaksegment-opslagblok (5), haalt het spraaksegment-opslagblok (5) de overeenkomstige spraaksegmentgegevens uit het geheugen in het spraaksegment-opslagblok
 30 (5) en zendt deze naar een stemhebbend-klank-syntheseblok (6) indien het een stemhebbende klank of een stemhebbende fricatieve klank is, of naar een niet-stemhebbende-klank-syntheseblok (7) indien het een niet-stemhebbende klank is. Dat wil zeggen, het stemhebbende-klank-syntheseblok (6) synthetiseert een digitaal spraaksignaal dat overeenkomt met de stemhebbende-klank-spraaksegmenten; en het niet-stemhebbende-klank-syntheseblok
 35 (7) synthetiseert een digitaal spraaksignaal dat overeenkomt met het niet-stemhebbende-klank-spraaksegment. Elk digitaal gesynthetiseerd spraaksignaal van het stemhebbende-klank-syntheseblok (6) en het niet-stemhebbende-klank-syntheseblok (7) wordt dan omgezet in een analoog
 40 signaal.

Het resulterende digitale gesynthetiseerde spraaksignaal, dat wordt uitgevoerd vanuit het stemhebbende-klank-syntheseblok (6) of het niet-stemhebbende-klank-syntheseblok (7), wordt dan gezonden naar een D/A-

omzettingsblok (8), dat bestaat uit een digitaal-analoog-omzetter, een analoog laagdoorlaatfilter en een analoge versterker, en wordt in een analoog signaal omgezet om gesynthetiseerd spraakgeluid te verschaffen.

Wanneer het stemhebbende-klank-syntheseblok (6) en het niet-stem-
 5 hebbende-klank-syntheseblok (7) de spraaksegmenten aaneenschakelen verschaffen zij de prosodie, zoals bedoeld is door het prosodie-opwekkende deelstelsel (2), aan gesynthetiseerde spraak door op juiste wijze de duur, de intensiteit en de toonhoogtefrequentie van het spraaksegment in te stellen op basis van de prosodie-informatie, d.w.z. intonatiepatroon-
 10 gegevens, klemtoonpatroongegevens, en duurgegevens.

De voorbereiding van het spraaksegment voor opslag in het spraaksegment-opslagblok (5) is als volgt. Eerst wordt een synthese-eenheid geselecteerd. Dergelijke synthese-eenheden omvatten foneem-, allofoon-, difoon-, lettergreep-, halflettergreep-, CVC-, VCV-, CV-, VC-eenheden
 15 (hier geeft "C" een medeklinker en "V" respectievelijk een klinkerfoneem weer) of combinaties daarvan. De synthese-eenheden die in de onderhavige spraaksynthese-werkwijze het meest algemeen gebruikt worden zijn de twee-klanken en de halflettergrepen.

Het spraaksegment dat overeenkomt met elk element van een verzameling van de synthese-eenheden wordt gescheiden van de spraakmonsters die werkelijk door een mens zijn uitgesproken. Het aantal elementen van de verzameling van synthese-eenheden is derhalve gelijk aan het aantal spraaksegmenten. In het geval bijvoorbeeld waarin halflettergrepen als de synthese-eenheden in het Engels worden gebruikt, is het aantal half-
 25 lettergrepen ongeveer 1000 en dienovereenkomstig is het aantal spraaksegmenten ook ongeveer 1000. Dergelijke spraaksegmenten bestaan in het algemeen uit een niet-stemhebbende klankinterval en een stemhebbende klankinterval.

Bij de onderhavige uitvinding worden het niet-stemhebbende spraak-
 30 segment en het stemhebbende spraaksegment, die verkregen zijn door het scheiden van het spraaksegment volgens de stand van de techniek in het niet-stemhebbende klankinterval en het stemhebbende klankinterval, gebruikt als de fundamentele synthese-eenheid. Het synthesegedeelte voor spraak met een niet-stemhebbende klank wordt bereikt overeenkomstig de
 35 stand van de techniek, zoals in het onderstaande wordt besproken. De synthese voor spraak met een stemhebbende klank wordt bereikt volgens de onderhavige uitvinding.

De niet-stemhebbende spraakelementen worden dus gedecodeerd in het niet-stemhebbende-klanksyntheseblok (7), dat in figuur 2 is getoond. Bij

het decoderen van de niet-stemhebbende klank is in de stand van de techniek opgemerkt dat het gebruik van een kunstmatig wit willekeurig ruis-sig-
 naal als een activeringssignaal voor een synthesefilter de kwaliteit
 van de gedecodeerde spraak niet aantast. Bij het coderen en decoderen van
 5 de niet-stemhebbende spraaksegmenten kan derhalve de vocoderingswerkwijze
 volgens de stand van de techniek als zodanig worden toegepast, in welke
 werkwijze de witte ruis als het activeringssignaal wordt gebruikt. In de
 synthese van niet-stemhebbende klanken volgens de stand van de techniek
 kan het witte ruissignaal bijvoorbeeld worden opgewekt door een toevals-
 10 getallen-opwekkingsalgoritme en kan vervolgens worden toegepast, of het
 vooraf opgewekte en in het geheugen opgeslagen witte ruissignaal kan
 tijdens het synthetiseren uit het geheugen worden opgeroepen, of een
 restsignaal dat verkregen is door het filteren van het niet-stemhebbende-
 klankinterval van de werkelijke spraak met gebruikmaking van een invers
 15 spectraal omhullende filter en opgeslagen is in een geheugen kan uit het
 geheugen worden opgeroepen bij het synthetiseren. Indien het niet noodza-
 kelijk is om de duur van het niet-stemhebbende spraaksegment te verande-
 ren kan een zeer eenvoudige coderingswerkwijze kan worden toegepast waar-
 in het niet-stemhebbende klankgedeelte gecodeerd wordt volgens een golf-
 20 vorm-coderingswerkwijze zoals pulscodemodulatie (PCM) of adaptieve ver-
 schil-pulscodemodulatie (ADPCM) en wordt opgeslagen. Deze wordt dan bij
 het synthetiseren gedecodeerd om te worden gebruikt.

De onderhavige uitvinding heeft betrekking op een coderings- en
 synthesesewerkwijze van de stemhebbende spraaksegmenten die de kwaliteit
 25 van de gesynthetiseerde spraak bepaalt. Een beschrijving met betrekking
 tot een dergelijke werkwijze, met de nadruk op het spraaksegment-opslag-
 blok en het stemhebbende-klank-syntheseblok, is (6) in figuur 2 getoond.

De stemhebbende spraaksegmenten onder de in het geheugen van het
 spraaksegment-opslagblok (5) opgeslagen spraaksegmenten worden vooraf
 30 overeenkomstig de periodieke-golfvorm-ontledingswerkwijze van de onder-
 havige uitvinding ontleed in golven met een toonhoogte-periodieke compo-
 nent, en daarin opgeslagen. Het stemhebbende-klank-syntheseblok (6) syn-
 thetiseert spraak met de gewenste toonhoogte en duurpatronen door over-
 eenkomstig de op tijdvervorming gebaseerde golfverplaatsingswerkwijze de
 35 golven op de juiste wijze te selecteren en te rangschikken. Het principe
 van deze werkwijzen wordt onder verwijzing naar de tekeningen in het
 onderstaande beschreven.

Stemhebbende spraak $s(n)$ is een periodiek signaal dat verkregen
 wordt wanneer een periodieke glottisgolf, opgewekt bij de stembanden, het

akoestische stemkanaalfilter $V(f)$, bestaande uit de mondholte, keelholte en neusholte, doorloopt. Hier wordt aangenomen dat het stemkanaalfilter $V(f)$ een frequentiekaracteristiek vanwege een lipstralings-effect omvat. Een spectrum $S(f)$ van stemhebbende spraak wordt gekenmerkt door:

- 5 1. Een fijne structuur die snel varieert ten opzichte van de frequentie " f "; en
2. Een spectrale omhullende die langzaam ten opzichte daarvan varieert,

 waarbij de eerste veroorzaakt wordt door periodiciteit van het
10 stemhebbende spraaksignaal en de laatste het spectrum van een glottispuls en de frequentiekaracteristiek van het stemkanaalfilter weergeeft.

Het spectrum $S(f)$ van de stemhebbende spraak neemt dezelfde vorm aan als de vorm die verkregen wordt wanneer de fijne structuur van een pulstrein vanwege de harmonische componenten die aanwezig zijn bij hele
15 veelvouden van de toonhoogtefrequentie F_0 vermenigvuldigd wordt met een spectrale omhullende functie $H(f)$. Stemhebbende spraak $s(n)$ kan derhalve beschouwd worden als een uitgangssignaal voor het geval dat een periodiek toonhoogte-pulstreinsignaal $e(n)$ met een vlakke spectrale omhullende en dezelfde periode als de stemhebbende spraak $S(n)$ wordt ingevoerd in een
20 in de tijd variërend filter met dezelfde frequentieresponsie-karakteristiek als voor de spectrale omhullende-functie $H(f)$ van de stemhebbende spraak $s(n)$. In het tijdsdomein gezien is de stemhebbende spraak $s(n)$ een convolutie van de impulsresponsie $h(n)$ van het filter $H(f)$ en het periodieke toonhoogte-pulstreinsignaal $e(n)$. Aangezien $H(f)$ overeenkomt met de
25 spectrale omhullende-functie van de stemhebbende spraak $s(n)$ wordt naar het in de tijd variërende filter met $H(f)$ als frequentieresponsie-karakteristiek verwezen als een spectraal omhullende- filter of een synthese-filter.

In figuur 3A is een signaal weergegeven voor vier perioden van een
30 glottisgolfvorm. Gewoonlijk zijn die golfvormen van de glottispulsen die de glottisgolfvorm vormen onderling vergelijkbaar maar niet compleet identiek, en ook de tijd van het interval tussen de aangrenzende glottispulsen is onderling vergelijkbaar maar niet compleet gelijk. Zoals boven is beschreven wordt de stemhebbende spraakgolfvorm $s(n)$ van figuur 3C
35 opgewekt wanneer de in figuur 3A getoonde glottisgolfvorm $g(n)$ gefilterd wordt door het spraakkanaalfilter $V(f)$. De glottisgolfvorm $g(n)$ bestaat uit de glottispulsen $g_1(n)$, $g_2(n)$, $g_3(n)$ en $g_4(n)$, die qua tijd van elkaar verschillen, en wanneer deze door het stemkanaalfilter $V(f)$ worden gefilterd, worden de in figuur 3B getoonde golven $s_1(n)$, $s_2(n)$, $s_3(n)$ en

$s_4(n)$ opgewekt. De stemhebbende spraakgolfvorm $s(n)$, die in figuur 3C is getoond, wordt gegenereerd door het superponeren van dergelijke golven.

Een fundamenteel idee van de onderhavige uitvinding is dat als de golven, die een stemhebbend spraaksignaal vormen, verkregen kunnen worden door het stemhebbende spraaksignaal te ontleden, spraak met een willekeurig accent en intonatiepatroon gesynthetiseerd kan worden door de intensiteit van de golven en de tijdsintervallen daartussen te wijzigen.

Aangezien de in figuur 3C getoonde stemhebbende-spraak-golfvorm $s(n)$ werd opgewekt door het superponeren van de golven die elkaar in de tijd overlappen, is het moeilijk om de golven uit de spraakgolfvorm $s(n)$ terug te winnen.

Om de golfvorm van elke periode elkaar in het tijddomein niet te laten overlappen moet de golfvorm een gepiekte golfvorm zijn waarin de energie geconcentreerd wordt rond één tijdstip, zoals in figuur 3F kan worden gezien.

Een gepiekte golfvorm is een golfvorm die in het frequentiedomein een vrijwel vlakke spectrale omhullende bezit. Wanneer een stemhebbende spraakgolfvorm $s(n)$ is gegeven kan een periodiek toonhoogte-pulstreinsignaal $e(n)$ met een vlakke spectrale omhullende, zoals getoond in figuur 3F, als uitgangssignaal verkregen worden door de omhullende van het spectrum $S(f)$ van de golfvorm $s(n)$ te schatten en dit in te voeren in een invers spectraal omhullendefilter $1/H(f)$ met de inverse van de omhullendefunctie $H(f)$ als frequentiekaracteristiek. De figuren 4, 5A en 5B hebben betrekking op deze stap.

Omdat de toonhoogtepuls-golfvormen van elke periode, die zoals getoond in figuur 3F het periodieke toonhoogte-pulstreinsignaal $e(n)$ vormen, elkaar in het tijddomein niet overlappen, kunnen zij gescheiden worden. Het principe van de periodieke-golfvorm-ontledingswerkwijze is dat aangezien de gescheiden "toonhoogte-pulssignalen voor één periode" $e_1(n)$, $e_2(n)$, ... een in wezen vlak spectrum bezitten, als zij terug in het spectrale omhullende filter $H(f)$ worden ingevoerd zodat de signalen het oorspronkelijke spectrum hebben, dat dan de golven $s_1(n)$, $s_2(n)$, enz., zoals getoond in figuur 3B, verkregen kunnen worden.

Figuur 4 is een blokdiagram van de periodieke-golfvorm-ontledingswerkwijze van de onderhavige uitvinding, waarin het stemhebbende spraaksegment in golven wordt ontleed. De stemhebbende spraakgolfvorm $s(n)$, die een digitaal signaal is, wordt verkregen door het analoge stemhebbende spraaksignaal of het geluidssignaal van een muziekinstrument met een laagdoorlaatfilter in bandbreedte te beperken en door de resulterende

signalen om te zetten in digitale signalen en op een magneetschijf op te slaan in de vorm van pulscodemodulatie (PCM) door verscheidene bits tegelijkertijd te groeperen, en worden dan, wanneer zij nodig zijn, voor bewerking opgeroepen.

5 De eerste stap van het golf-voorbewerkingsproces volgens de periodieke-golfvorm-ontledingswerkwijze is een blinde deconvolutie waarin de stemhebbende spraakgolfvorm $s(n)$ (periodiek signaal $s(n)$) gedeconvolveerd wordt tot een impulsresponsie $h(n)$, die een tijddomeinfunctie is van de spectrale omhullendefunctie $H(f)$ van het signaal $s(n)$, en een periodiek
10 toonhoogte-pulstreinsignaal $e(n)$ met een vlakke spectrale omhullende en dezelfde periode als het signaal $s(n)$. Zie de figuren 5A en 5B en de daarop betrekking hebbende bespreking.

Zoals beschreven is voor de blinde deconvolutie de spectrumschattingstechniek, waarmee de spectrale omhullendefunctie $H(f)$ uit het signaal $s(n)$ geschat wordt, essentieel.
15

Spectrumschattings technieken volgens de stand van de techniek kunnen in drie werkwijzen worden onderverdeeld:

1. een blok-analyse-werkwijze;
2. een toonhoogte-synchrone-analyse-werkwijze; en
- 20 3. een sequentiële analyse-werkwijze die afhangt van de lengte van een analyse-interval.

De blokanalyse-werkwijze is een werkwijze waarin het spraaksignaal wordt onderverdeeld in blokken met een constante duur in de orde van 10-20 ms (milliseconden), en dan wordt de analyse uitgevoerd met betrek-
25 king tot het constante aantal spraakbemonsteringen dat in elk blok aanwezig is, waardoor één verzameling (meestal 10-16 parameters) van spectrale omhullende-parameters voor elk blok worden verkregen, voor welke werkwijze een homomorfe analysemethode en een blokanalysemethode met lineaire predictie typerend zijn.

30 De toonhoogte-synchrone-analyse-werkwijze verkrijgt één verzameling van spectrale-omhullende-parameters voor elke periode door een analyse uit te voeren op elk periodieke-spraaksignaal dat verkregen werd door het spraaksignaal onder te verdelen met de toonhoogte-periode als eenheid (zoals getoond in figuur 3C), voor welke werkwijze de analyse-door-syn-
35 these-werkwijze en de toonhoogte-synchroon-analyse-werkwijze met lineaire predictie typerend zijn.

In de sequentiële analyse-werkwijze wordt één verzameling spectrale-omhullende parameters verkregen voor elk spraakmonster (zoals getoond in figuur 3D) door het schatten van het spectrum voor elk spraakmonster,

voor welke werkwijze de kleinste-kwadraten-werkwijze en de recursieve kleinste-kwadraten-werkwijze, die een soort van adaptieve filterwerkwijze vormen, typerend zijn.

5 Figuur 3D toont de variatie in de tijd van de eerste vier reflectiecoëfficiënten van 14 reflectiecoëfficiënten $k_1, k_2, \dots, k_{14}$, die een verzameling van spectrale-omhullende-parameters vormen welke verkregen is door de sequentiële analysewerkwijze (zie figuur 5A). Zoals uit de tekening kan worden opgemaakt veranderen de waarden van de spectrale omhullende parameters continu vanwege de continue beweging van de articulatorische organen, hetgeen betekent dat de impulsresponsie $h(n)$ van het
10 spectrale omhullende-filter continu verandert. Aangenomen dat $h(n)$ in een interval van één periode niet verandert wordt hier, ten behoeve van de uitleg, $h(n)$ tijdens de eerste, tweede en derde periode, respectievelijk aangegeven als $h(n)_1, h(n)_2, h(n)_3$, zoals getoond in figuur 3E.

15 Een verzameling van omhullende-parameters die verkregen is door verschillende spectrumschattingstechnieken, zoals een cepstrum $CL(i)$, dat een verzameling parameters is die verkregen is door de homomorfe analysewerkwijze, en een verzameling predictiecoëfficiënten $\{a_i\}$ of een verzameling reflectiecoëfficiënten $\{k_i\}$, of een verzameling van lijnspectrum-
20 paren, enz., die verkregen is door het toepassen van de recursieve kleinste-kwadraten-werkwijze of de lineaire predictie-werkwijze, wordt op dezelfde wijze behandeld als de $H(f)$ of $h(n)$, omdat deze de frequentie-karakteristiek $H(f)$ of de impulsresponsie $h(n)$ van het spectrale-omhullende-filter kan vormen. In het navolgende wordt de impulsresponsie der-
25 halve ook de verzameling van spectrale-omhullende-parameters genoemd.

De figuren 5A en 5B tonen werkwijzen voor de blinde deconvolutie.

Figuur 5A toont een blinde deconvolutiewerkwijze die wordt uitgevoerd door gebruikmaking van de lineaire predictie-analysewerkwijze of door gebruikmaking van de recursieve kleinste-kwadraten-werkwijze, welke
30 beide werkwijzen zijn volgens de stand van de techniek. Indien een stemhebbende spraakgolfvorm $s(n)$ gegeven is, zoals getoond in figuur 3C, worden de predictiecoëfficiënten $\{a_1, a_2, \dots, a_N\}$ of de reflectiecoëfficiënten $\{k_1, k_2, \dots, k_N\}$, die de spectrale omhullende parameters zijn die de frequentie-karakteristiek $H(f)$ of de impulsresponsie $h(n)$ van het
35 spectrale omhullende filter representeren, verkregen met gebruikmaking van de analysewerkwijze met lineaire predictie of de recursieve kleinste-kwadraten-werkwijze. Normaal zijn 10-16 predictiecoëfficiënten voldoende voor de orde van de predictie "N". Met gebruikmaking van de predictiecoëfficiënten $\{a_1, a_2, \dots, a_N\}$ en de reflectiecoëfficiënten $\{k_1, k_2, \dots$

kN als de spectrale-omhullende- parameter kan een invers spectrale-om-
 hullende-filter (of eenvoudigweg een invers filter genoemd) met de fre-
 quentiekarakteristiek van $1/H(f)$, die de inverse is van de frequentie-
 karakteristiek $H(f)$ van het spectrale- omhullende-filter, door een des-
 5 kundige op eenvoudige wijze worden samengesteld. Indien de stemhebbende
 spraakgolfvorm het ingangssignaal is voor het inverse spectrale-omhullen-
 de-filter, dat in de lineaire predictie-analysewerk wijze of in de recur-
 sieve kleinste-kwadraten-werkwijze ook lineaire-predictie-foutfilter
 wordt genoemd, kan het periodieke toonhoogte-pulstreinsignaal van het
 10 type van figuur 3F met de vlakke spectrale omhullende, dat een predictie-
 fout-sig naal of een restsig naal wordt genoemd, als uitgangssig naal van
 het filter worden verkregen.

De figuren 5B en 5C tonen de blinde deconvolutiewerkwijze met
 gebruikmaking van de homomorfe analysewerk wijze, die een blokanalysewerk-
 15 wijze is, terwijl figuur 5B de werk wijze toont die wordt uitgevoerd door
 een frequentiedeling en figuur 5C respectievelijk de werk wijze toont die
 wordt uitgevoerd door invers filteren.

Een beschrijving van figuur 5B volgt. Spraakmonsters voor analyse
 van één blok worden verkregen door het stemhebbende spraaksig naal $s(n)$ te
 20 vermenigvuldigen met een tapse vensterfunctie, zoals een Hamming-venster,
 met een duur van ongeveer 10-20 ms. Een cepstrale reeks $c(i)$ wordt dan
 verkregen door de spraakmonsters te verwerken met gebruikmaking van een
 serie homomorfe verwerkingsprocedures die bestaan uit een discrete Fou-
 rier-transformatie, een complex logaritme en een inverse discrete Fou-
 25 rier-transformatie, zoals getoond in figuur 5D. Het cepstrum is een func-
 tie van de quefreny, wat een eenheid is die vergelijkbaar is met de
 tijd.

Een laagquefrent cepstrum $CL(i)$, dat zich rond de oorsprong bevindt
 en de spectrale omhullende van de stemhebbende spraak $s(n)$ weergeeft, en
 30 een hoogquefrent cepstrum $CH(i)$, dat een periodiek toonhoogte-pulstrein-
 sig naal $e(n)$ weergeeft, kunnen in het quefrentiedomein van elkaar worden
 gescheiden. Dat wil zeggen, het vermenigvuldigen van het cepstrum $c(i)$
 met respectievelijk een laagquefrente vensterfunctie en een hoogquefrente
 vensterfunctie levert respectievelijk $CL(i)$ en $CH(i)$ op. Het respectieve-
 35 lijk onderwerpen hiervan aan een inverse homomorfe verwerkingsprocedure,
 zoals getoond in figuur 5E, geeft de impulsresponsie $h(n)$ en het toon-
 hoogte-pulstreinsig naal $e(n)$. Aangezien het onderwerpen van $CH(i)$ aan de
 inverse homomorfe verwerkingsprocedure niet direct het toonhoogte-puls-
 treinsig naal $e(n)$ oplevert maar het toonhoogte-pulstreinsig naal van één

blok, vermenigvuldigd met een tijdvensterfunctie $w(n)$, oplevert, kan $e(n)$ in dit geval verkregen worden door het opnieuw vermenigvuldigen van het toonhoogte-pulstreinsignaal met een inverse tijdvensterfunctie $1/w(n)$ die overeenkomt met de inverse van $w(n)$.

5 De werkwijze van figuur 5C is gelijk aan die van figuur 5B, met uitzondering van het feit dat in figuur 5C $CL(i)$ in plaats van $CH(i)$ wordt gebruikt bij het verkrijgen van het periodieke toonhoogte-pulstreinsignaal $e(n)$. Dat wil zeggen, door gebruik te maken van de eigenschap dat een impulsresponsie $h^{-1}(n)$ die overeenkomt met $1/H(f)$, die de
10 inverse is van de frequentiekaracteristiek $H(f)$, verkregen kan worden door het verwerken van $-CL(i)$, die verkregen wordt door de tegengestelde van $CL(i)$ te nemen, kan in deze werkwijze door de inverse homomorfe verwerkingsprocedure het periodieke toonhoogte-pulstreinsignaal $e(n)$ als uitgangssignaal verkregen worden door een filter met een eindige impuls-
15 responsie (FIR) te vervaardigen dat $h^{-1}(n)$ als impulsresponsie heeft, en door in het filter een oorspronkelijk spraaksignaal $s(n)$, dat niet vermenigvuldigd is met een vensterfunctie, in te voeren. Deze werkwijze is een inverse filterwerkwijze die in wezen gelijk is aan die van figuur 5A, met uitzondering van het feit dat terwijl in de homomorfe analyse van figuur
20 5C het inverse spectrale-omhullende-filter $1/H(f)$ vervaardigd wordt door een impulsresponsie $h^{-1}(n)$ van het inverse spectrale-omhullende-filter te verkrijgen, in figuur 5A het inverse spectrale omhullende filter $1/H(f)$ direct kan worden samengesteld aan de hand van de predictiecoëfficiënten $\{a_i\}$ of de reflectiecoëfficiënten $\{k_i\}$ die door de lineaire-predictie-
25 analysewerkwijze zijn verkregen.

In de op de homomorfe analyse gebaseerde blinde deconvolutie kan de impulsresponsie $h(n)$ of het laag-quefrente cepstrum $CL(i)$, getoond door stippellijnen in de figuren 5B en 5C, gebruikt worden als de verzameling van spectrale-omhullende-parameters. Wanneer de impulsresponsie $\{h(0),$
30 $h(1), \dots, h(N-1)\}$ wordt gebruikt, omvat een verzameling van spectrale-omhullende-parameters normalerwijs een behoorlijk aantal parameters waarbij de orde van N 90 - 120 bedraagt, terwijl het aantal parameters kan worden verminderd tot 50-60, waarbij N 25 - 30 bedraagt, wanneer het cepstrum $\{CL(-N)_m, CL(-N+1), \dots, 0, \dots, CL(N)\}$ wordt gebruikt.

35 Zoals boven is beschreven wordt de stemhebbende spraakgolfvorm $s(n)$ overeenkomstig de procedure van figuur 5 gedeconvolueerd tot in de impulsresponsie $h(n)$ van het spectrale omhullende filter en het periodieke toonhoogte-pulstreinsignaal $e(n)$.

Zodra het toonhoogte-pulstreinsignaal en de spectrale-omhullende-

parameters volgens de blinde deconvolutieprocedure verkregen zijn worden de toonhoogte-pulsposities P_1 , P_2 , enz. verkregen uit het periodieke toonhoogte-pulstreinsignaal $e(n)$ of het spraaksignaal $s(n)$ door toepassing van een toonhoogte-pulspositie-detectiealgoritme in het tijddomein, zoals het tijdvak-detectie-algoritme. Vervolgens worden de in de figuren 3H, 3K, 3N getoonde toonhoogte-pulssignalen $e_1(n)$, $e_2(n)$ en $e_3(n)$ respectievelijk verkregen door het periodiek segmenteren van het toonhoogte-pulstreinsignaal $e(n)$, zodat één toonhoogte-puls bevat is in één periode-interval, zoals in figuur 3F is getoond. De posities van de segmentatie kunnen bepaald worden als middelpunten tussen de toonhoogtepulsen of punten die een constante tijd voor elke toonhoogte-puls liggen. Aangezien de positie van elke toonhoogte-puls in de tijd samenvalt met het eindgedeelte van elke glottispuls, zoals uit een vergelijking van de figuren 3A en 3F kan worden opgemaakt, verdient het de voorkeur om als positie van de segmentatie een punt op een constante tijd achter elke toonhoogte-puls te kiezen, zoals in figuur 3F door de stippellijn is aangegeven. Aangezien de toonhoogte-puls echter het grootste effect op de hoorbaarheid vormt zijn er tussen deze gevallen geen significante verschillen in de gesynthetiseerde spraak.

Als de door deze werkwijze verkregen toonhoogte-pulssignalen $e_1(n)$, $e_2(n)$, $e_3(n)$, enz. respectievelijk opnieuw worden geconvolveerd met de $h_1(n)$, $h_2(n)$, $h_3(n)$ van figuur 3E, die impulsresponsies zijn tijdens het periode-interval van de toonhoogte-pulssignalen $e_1(n)$, $e_2(n)$, $e_3(n)$, enz., worden de bedoelde golven verkregen zoals getoond in de figuren 3I, 3L, 3(0). Een dergelijke convolutie kan op geschikte wijze worden uitgevoerd door het invoeren van elk toonhoogte-pulstreinsignaal in het spectrale omhullende filter $H(f)$, dat de spectrale-omhullende-parameters gebruikt als de filtercoëfficiënten, zoals in figuur 4 is getoond. In gevallen waarin de lineaire predictiecoëfficiënten of de reflectiecoëfficiënten of de lijnspectrum-paren bijvoorbeeld gebruikt worden als de spectrale-omhullende-parameters zoals in de lineaire-predictie-analyse-werkwijze wordt een IIR-filter (filter met oneindige impulsresponsie) samengesteld met de lineaire predictiecoëfficiënten of de reflectiecoëfficiënten of de lijnspectrum-paren als filtercoëfficiënten. In gevallen waarin de impulsresponsie gebruikt wordt voor de spectrale omhullende parameters, zoals in de homomorfe-analyse-werkwijze, wordt een FIR-filter samengesteld met de impulsresponsie als de aftakingscoëfficiënten. Aangezien het synthesesfilter niet direct kan worden samengesteld als de spectrale-omhullende-parameter een logaritmische oppervlakteverhouding of

het cepstrum is, moeten de spectrale-omhullende-parameters terug worden getransformeerd naar de reflectiecoëfficiënten of de impulsresponsie om gebruikt te worden als de coëfficiënten van het IIR- of FIR-filter. Als het toonhoogte-pulssignaal voor één periode hetingangssignaal is voor
 5 het spectrale-omhullende- filter dat is samengesteld zoals boven is beschreven, waarbij de filtercoëfficiënten in de tijd veranderen overeenkomstig de spectrale omhullende parameters die overeenkomen met hetzelfde moment als elk monster van het toonhoogte-pulssignaal, wordt dan de golf voor die periode uitgevoerd.

10 Om die reden werden de "tijdfunctie-golfvormen" van de spectrale-omhullende-parameters verwijderd op hetzelfde punt waarop $e(n)$ verwijderd werd om het toonhoogte-pulssignaal voor elke periode te verkrijgen. In het geval van sequentiële analyse bijvoorbeeld worden de spectrale omhullende-parameters $k_1(n)_1$, $k_2(n)_1$, enz. van de eerste periode, zoals
 15 getoond in figuur 3G, verkregen door het verwijderen van de spectrale-omhullende-parameters die overeenkomen met dezelfde tijdperiode als het toonhoogte-pulssignaal $e_1(n)$ van de eerste periode, dat getoond is in figuur 3H, uit de tijdfuncties $k_1(n)$, $k_2(n)$, enz. van de spectrale-omhullende-parameters, zoals in figuur 3D is getoond. De spectrale-
 20 omhullende-parameters van de tweede en derde periode, die in de figuren 3J en 3M zijn aangegeven als een getrokken lijn, kunnen ook worden verkregen op een overeenkomstige manier zoals boven is genoemd. In figuur 4 zijn de reflectiecoëfficiënten k_1 , k_2 , ..., k_N en de impulsresponsie $h(0)$, $h(1)$, ..., $h(N-1)$ getoond als een typische verzameling van spectrale-
 25 le-omhullende-parameters, waar zij worden aangegeven als $k_1(n)$, $k_2(n)$, ..., $k_n(n)$ en $h(0,n)$, $h(1,n)$, ..., $h(N-1, n)$ om te onderstrepen dat zij functies van de tijd zijn. Op overeenkomstige wijze zal in gevallen waar in het cepstrum $CL(i)$ gebruikt wordt als de verzameling van spectrale-omhullende-parameters dit worden aangegeven als $CL(i,n)$.

30 Aangezien de tijdfuncties van de spectrale omhullende parameters, in tegenstelling tot de sequentiële analysewerkwijze, niet verkregen worden in het geval van de toonhoogte-synchrone analysewerkwijze of de blok-analysewerkwijze, maar de waarden van de spectrale omhullende parameters worden verkregen die ontstaan zijn over het analyse-interval, zou
 35 het nodig zijn om de tijdfuncties van de spectrale-omhullende- parameters uit de waarden van de spectrale-omhullende-parameters te maken en vervolgens de tijdfuncties periode voor periode in segmenten te verdelen om de spectrale omhullende parameters voor één periode te verkrijgen. In werkelijkheid is het echter handig om, in plaats van het samenstellen van de

tijdfuncties, als volgt te werk te gaan. Dat wil zeggen, aangezien een verzameling van spectrale omhullende parameters met constante waarden overeenkomt met elk toonhoogte-periode-interval, zoals in figuur 8B door een onderbroken lijn is getoond, vertonen de spectrale-omhullende-parame-
 5 ters in het geval van de toonhoogte-synchrone analysewerkwijze geen verandering, zelfs wanneer hun tijdfuncties periode voor periode in segmenten worden onderverdeeld. De spectrale-omhullende-parameters die voor één periode in een buffer moeten worden opgeslagen, zijn derhalve geen tijdfuncties maar constanten, die onafhankelijk van de tijd zijn.

10 In het geval van de blokanalyse-werkwijze zijn de waarden van een spectrale omhullende parameter voor één periode die bij één blok behoort, bijvoorbeeld $k_1(n)1, k_1(n)2, \dots, k_1(n)M$ niet alleen constant onafhankelijk van de tijd maar ook identiek, aangezien een verzameling van constante spectrale omhullende parameters per blok wordt verkregen (hier staat
 15 $k_1(n)j$ voor de tijdfunctie van k_1 voor het j -de periode-interval, en geeft M het aantal toonhoogte-periode-intervallen weer dat tot een blok behoort).

In het geval van de blokanalysewerkwijze moet opgemerkt worden dat wanneer het toonhoogte-pulssignaal over de grens van twee aangrenzende
 20 blokken ligt, de waarden van de spectrale omhullende parameters van het voorgaande blok en het volgende blok respectievelijk gebruikt zullen worden voor de voorgaande en volgende signaaldelen die met betrekking tot de blokgrens in delen zijn gesplitst.

Zoals uit figuur 3I kan worden opgemaakt is de duur van de golf
 25 niet noodzakelijkerwijs gelijk aan één periode. Voor het aan het spectrale omhullende filter toevoeren van het toonhoogte-pulssignaal en de spectrale omhullende parameters van één periodelengte, die verkregen zijn door de periodieke segmentatie, zijn de processen van het toevoegen van nullen en het verlengen van parameters, wat in figuur 4 is getoond, nood-
 30 zakelijk om de duur van het toonhoogte-pulssignaal en de spectrale-omhullende-parameters ten minste gelijk aan de effectieve duur van de golf te doen zijn. Het proces van het toevoegen van nullen heeft tot doel de totale duur van het toonhoogte-pulssignaal gelijk aan de vereiste lengte te maken door het toevoegen van bemonsteringen met de waarde nul
 35 na het toonhoogte-pulssignaal van één periode. Het proces van het verlen-
 gen van parameters heeft tot doel de totale duur van de spectrale-omhul-
 lende-parameter gelijk te maken aan de vereiste lengte door het toevoegen van de spectrale omhullende parameter voor de volgende perioden na de spectrale omhullende parameter van één periodelengte. Zelfs indien een

eenvoudige werkwijze van het herhaald toevoegen van de laatste waarde van de spectrale omhullende parameters van een periode of de eerste waarde van de spectrale omhullende parameters van de volgende periode wordt gebruikt, wordt de kwaliteit van de gesynthetiseerde spraak echter niet
 5 significant verminderd.

Het feit dat de effectieve duur van de door het spectrale-omhullende-filter op te wekken golf afhangt van de waarden van de spectrale omhullende parameters maakt het moeilijk om deze van te voren te schatten. Als in overweging wordt genomen dat de effectieve duur van de golf
 10 in het geval van mannelijke spraak twee perioden vanaf de toonhoogtepulspositie en in het geval van vrouwelijke spraak of de spraak van kinderen drie perioden vanaf de toonhoogtepulspositie is, is het echter in de meeste gevallen gunstig om te beslissen dat de duur van het "verlengde" toonhoogtepulssignaal" wordt gevormd door het toevoegen van nullen
 15 en dat de "toegevoegde spectrale omhullende parameters" die gevormd moeten worden door het toevoegen van parameters respectievelijk drie en vier periodelengten aannamen voor mannelijke en vrouwelijke spraak in het geval dat periodieke segmentatie direct achter de toonhoogtepulsen wordt uitgevoerd, aangezien er in de praktijk geen significante fouten optreden. In figuur 3G worden verlengde spectrale omhullende parameters voor
 20 de eerste periode van het interval "ad" van drie perioden, gemaakt door het toevoegen van de spectrale-omhullende-parameters voor het interval "bd" van twee perioden dat is aangegeven door een onderbroken lijn naast de spectrale-omhullende-parameter van het interval "ab" van de eerste
 25 periode, dat verkregen is door de periodieke segmentatie, als voorbeeld getoond. In figuur 3H is een verlengd toonhoogtepulssignaal voor de eerste periode van het interval "ad" van drie perioden, dat gemaakt is door het toevoegen van bemonsteringen met de waarde nul aan het interval "bd" van twee perioden, naast het toonhoogtepulssignaal van het interval
 30 "ab" met één periode, dat verkregen is door de periodieke segmentatie, als voorbeeld getoond.

Aangezien de duur na het toevoegen van nullen en de parameterverlenging is toegenomen tot drie of vier perioden, terwijl de duur van het toonhoogtepulssignaal en de spectrale-omhullende-parameters voor het
 35 toevoegen van nullen en de parameterverlenging één periode bedraagt, zijn in het geval zoals boven is beschreven buffers verschaft tussen de periodieke segmentatie en de parameterverlenging, zoals in figuur 4 is getoond, en worden het toonhoogtepulssignaal en de spectrale-omhullende-parameters die verkregen zijn door de periodieke segmentatie dan opgesla-

gen in de buffers en worden opgevraagd wanneer zij nodig zijn, zodat een tijdelijk bufferen tot stand is gebracht.

Indien het verlengde toonhoogte-pulssignaal en de verlengde spectrale omhullende parameters verkregen worden door het toevoegen van nul-
 5 len en het parameter verlengen in figuur 4, kan het "golfsignaal" $s_1(n)$ voor de eerste periode van de lengte van het interval van drie perioden, zoals het interval "ad" zoals in figuur 3I is getoond, uiteindelijk worden verkregen door het invoeren van het verlengde toonhoogte-pulssignaal van de eerste periode, zoals het interval "ad" van figuur 3H, in het
 10 spectrale omhullende filter $H(f)$ en het synchroon variëren van de coëfficiënten op dezelfde wijze als de verlengde spectrale-omhullende-parameters van de eerste periode, zoals het interval "ad" van figuur 3G. De golfsignalen $s_2(n)$ en $s_3(n)$ voor respectievelijk de tweede en derde periode kunnen op overeenkomstige wijze worden verkregen.

15 Zoals boven is beschreven, wordt de stemhebbende spraakgolfvorm $s(n)$ uiteindelijk door de procedure van figuur 4 ontleed in de golven die de golfvorm $s(n)$ vormen. Het zal duidelijk zijn dat het terug naar de oorspronkelijke punten herrangschikken van de golven van de figuren 3I, 3L en 3(0), die verkregen zijn door ontleding, figuur 3B oplevert en
 20 indien de golven gesuperponeerd zijn wordt weer de oorspronkelijke spraakgolfvorm $s(n)$ verkregen, zoals in figuur 3C is getoond. Als de golven van de figuren 3I, 3L en 3(0) opnieuw worden gerangschikt door het variëren van de onderlinge afstanden en vervolgens worden gesuperponeerd zoals in figuur 3P is getoond, wordt de spraakgolf verkregen met een
 25 verschillend toonhoogte-patroon, zoals in figuur 3Q is getoond. Als zodanig maakt het op de juiste wijze variëren van het tijdsinterval tussen de golven die door ontleding zijn verkregen de synthese van spraak mogelijk met een willekeurig gewenst toonhoogte-patroon, d.w.z. de intonatie. Op overeenkomstige wijze maakt het op de juiste manier variëren van de ener-
 30 gie van de golven de synthese van spraak mogelijk met het willekeurig gewenste klemtoonpatroon.

In het in figuur 2 getoonde spraaksegment-opslagblok wordt elk stemhebbend spraaksegment, dat volgens de in figuur 4 getoonde werkwijze is ontleed in evenveel golven als het aantal toonhoogte-pulsen, opgesla-
 35 gen in de opmaak zoals deze in figuur 6A is getoond, welke de spraaksegment-informatie wordt genoemd. In een kopveld, dat het voorste deel is van de spraaksegment-informatie, worden grenstijdstippen $B_1, B_2, \dots, B_L$, die belangrijke tijdstippen in het spraaksegment zijn, en toonhoogte-pulsposities $P_1, P_2, \dots, P_M$ van elk in de synthese van elke golf gebruikt

toonhoogte-pulssignaal opgeslagen, waarin het aantal bemonsteringen dat overeenkomt met elk tijdstip wordt vastgelegd, waarbij de eerste bemonsteringspositie van het eerste toonhoogte-pulssignaal $e_1(n)$ als 0 wordt genomen. Het grenstijdstip is de positie in de tijd van de grenspunten
 5 tussen de deelsegmenten die resulteren wanneer het spraaksegment in verschillende deelsegmenten wordt onderverdeeld. Een klinker met daarvoor en daarna medeklinkers kan bijvoorbeeld worden beschouwd als bestaande uit drie deelsegmenten voor spraak met lage snelheid, aangezien de klinker kan worden onderverdeeld in een blijvende-toestand-interval van het mid-
 10 delste deel en twee overgangsintervallen die aanwezig zijn voor en na het blijvende-toestand-interval, en drie eindpunten van de deelsegmenten worden opgeslagen als de grenstijdstippen in het kopveld van het spraaksegment. In het geval waarin het bemonsteren bij een hogere spraaksnelheid wordt uitgevoerd, worden echter twee grenstijdstippen in de kop-
 15 informatie opgeslagen omdat het overgangsinterval één punt wordt, zodat het spraaksegment van de klinker beschouwd kan worden als bestaande uit twee deelsegmenten.

In het golf-codeveld, hetgeen het laatste deel van de spraaksegment-informatie is, worden golfcodes opgeslagen die codes zijn die ver-
 20 kregen zijn door golfvorm-coderen van de golf die overeenkomt met elke periode. De golven kunnen gecodeerd worden door een eenvoudige golfvorm-coderingswerkwijze, zoals PCM, maar omdat de golven een significante correlatie op korte termijn en lange termijn bezitten kan de hoeveelheid voor opslag noodzakelijk geheugen aanmerkelijk worden verminderd als de
 25 golven op efficiënte wijze qua golf gecodeerd worden door gebruik te maken van ADPCM met een toonhoogte-voorspellende lus, een adaptieve voorspellende codering of een digitale adaptieve delta-modulatie-werkwijze. De werkwijze waarin de door ontleding verkregen golven qua golf worden gecodeerd, waarbij de resulterende codes worden opgeslagen en, op het
 30 moment van de synthese, de codes worden gedecodeerd, opnieuw gerangschikt en gesuperponeerd om de gesynthetiseerde spraak te produceren, wordt de "golfvorm-code-opslagwerkwijze" genoemd.

Het toonhoogte-pulssignaal en de corresponderende spectrale- omhullende-parameters kunnen beschouwd worden als zijnde identiek aan de golf,
 35 omdat zij de ingrediënten zijn waarmee de golf kan worden gevormd. Een werkwijze is derhalve ook mogelijk waarin de "broncodes", die verkregen zijn door het coderen van de toonhoogte-pulssignalen en de spectrale-omhullende-parameters, worden opgeslagen en de golven gemaakt worden met de toonhoogte-pulssignalen en de spectrale-omhullende- parameters die

verkregen zijn door het decoderen van de broncodes, en de golven worden dan opnieuw gerangschikt en gesuperponeerd om de gesynthetiseerde spraak te produceren. Deze werkwijze wordt de "broncode-opslagwerkwijze" genoemd. Deze werkwijze komt overeen met die waarin het toonhoogte-pulssignaal en de spectrale-omhullende-parameters die in de buffers zijn opgeslagen, in plaats van de golven die in figuur 4 als het uitgangssignaal zijn verkregen, in hetzelfde periode-interval worden samengevoegd en vervolgens worden opgeslagen in het spraaksegment-opslagblok. In de broncode-opslagwerkwijze worden derhalve de procedures na de buffer in figuur 4, d.w.z. de parameter-verlengingsprocedure, de nul-toevoegingsprocedure en de filterprocedure door het synthese-filter $H(f)$, uitgevoerd in het golfvorm-samenstellings-deelblok in figuur 7.

In de broncode-opslagwerkwijze is de opmaak van de spraaksegment-informatie zoals getoond in figuur 6B, die gelijk is aan figuur 6A afgezien van de inhoud van het golfcodeveld. Dat wil zeggen, de toonhoogte-pulssignalen en de spectrale omhullende parameters die nodig zijn voor de synthese van de golven worden in plaats van de golven gecodeerd en opgeslagen op de posities waar de golf voor elke periode in figuur 6A moet worden opgeslagen.

De spectrale-omhullende-parameters worden gecodeerd volgens de quantiseringswerkwijze voor de spectrale omhullende parameters volgens de stand van de techniek en worden in het golfcodeveld opgeslagen. Op dat moment kan de codering op efficiënte wijze worden uitgevoerd als de spectrale omhullende parameters voor de quantisatie op geschikte wijze zijn getransformeerd. Het verdient bijvoorbeeld de voorkeur om de predictiecoëfficiënten te transformeren in de parameters van het lijnspectrumpaar en de reflectiecoëfficiënten in de log-oppervlakte-verhoudingen en deze te quantiseren. Aangezien de impulsresponsie tussen aangrenzende bemonsteringen en tussen aangrenzende impulsresponsies sterk gecorreleerd is, kan de hoeveelheid gegevens die voor opslag nodig is aanzienlijk worden gereduceerd als deze volgens een verschilcoderingswerkwijze qua golfvorm zijn gecodeerd. In het geval van de cepstrumparameters is een coderingswerkwijze bekend waarin de cepstrumparameter zodanig getransformeerd wordt dat de hoeveelheid gegevens significant gereduceerd kan worden.

Anderzijds wordt het toonhoogte-pulssignaal volgens een geschikte golfvorm-coderingswerkwijze gecodeerd en wordt de resulterende code opgeslagen in het golfcodeveld. De toonhoogte-pulssignalen hebben een geringe correlatie op korte termijn maar hebben onderling een significante correlatie op lange termijn. Als de golfvorm-coderingswerkwijze, zoals de

toonhoogte-voorspellende adaptieve PCM-codering met een toonhoogte-voorspellende lus, wordt gebruikt, kan derhalve gesynthetiseerde spraak met een hoge kwaliteit worden verkregen, zelfs wanneer de hoeveelheid voor opslag benodigd geheugen wordt gereduceerd tot drie bits per bemonstering. De predictiecoëfficiënt van een toonhoogte-predictor kan een waarde hebben die voor elke toonhoogte-periode verkregen wordt volgens een autocorrelatiewerkwijze, of kan een constante waarde bezitten. In het eerste stadium van de codering kan het toonhoogte-predictie-effect versterkt worden door een normalisatie door het delen van het te coderen toonhoogte-pulssignaal door de vierkantswortel van een gemiddelde energie "G" per bemonstering. Het decoderen wordt uitgevoerd in het stemhebbende-spraak-syntheseblok, en het toonhoogte-pulssignaal wordt tot diens oorspronkelijke grootte teruggebracht door in de eindtrap van de decodering weer met "G" te vermenigvuldigen.

In figuur 6B wordt de spraaksegment-informatie getoond voor het geval waarin een lineaire predictieve analysewerkwijze wordt gebruikt die 14 reflectiecoëfficiënten als spectrale omhullende parameters gebruikt. Als het analyse-interval voor de lineaire predictieve analyse de toonhoogte-periode is komen 14 reflectiecoëfficiënten overeen met elk toonhoogte-pulssignaal en worden opgeslagen. Als het analyse-interval een blok met een zekere lengte is hebben de reflectiecoëfficiënten voor verschillende toonhoogte-pulsen in één blok dezelfde waarden, zodat de hoeveelheid voor de opslag van de golven benodigd geheugen gereduceerd wordt. In dit geval moet, zoals boven is beschreven, de positie van het grenspunt tussen blokken aanvullend in het kopveld worden opgeslagen, aangezien de reflectie-coëfficiënten van het voorste blok of het laatste blok op het moment van de synthese gebruikt worden voor het toonhoogte-pulssignaal dat over de grens van twee blokken ligt, afhankelijk van het feit of de bemonsteringen van het signaal voor of na het grenspunt liggen. Als de sequentiële analysewerkwijze, zoals de recursieve kleinste-kwadraten-werkwijze, wordt gebruikt, worden de reflectiecoëfficiënten k_1 , k_2 , ..., k_{14} continue functies van de tijdsindex "n", zoals getoond in figuur 3D, en is een grote hoeveelheid geheugen benodigd om de tijdfunctie $k_1(n)$, $k_2(n)$, ..., $k_{14}(n)$ op te slaan. Als het geval van figuur 3 als voorbeeld wordt genomen, worden de golfvormen voor het interval "ab" van figuur 3G en figuur 3H als de eerste periode en voor het interval "bc" van figuur 3J en figuur 3K als de tweede periode en voor het interval "cd" van figuur 3M en figuur 3N als de derde periode van het golf-codeveld opgeslagen in het golf-codeveld.

De golfvormcode-opslagwerkwijze en de broncode-opslagwerkwijze zijn in wezen dezelfde werkwijze, en in feite is de golfvormcode die verkregen wordt wanneer de golven volgens een efficiënte golfvorm-coderingswerkwijze zoals APC (adaptief predictief coderen) in de golfvormcode-opslag-
 5 werkwijze wordt gecodeerd qua inhoud vrijwel gelijk aan de broncode die verkregen wordt in de broncode-opslagwerkwijze. De golfvormcode in de golfvormcode-opslagwerkwijze en de broncode in de broncode-opslagwerkwijze worden samen de golfcode genoemd.

Figuur 7 geeft de interne configuratie weer van het stemhebbende-
 10 spraak-syntheseblok van de onderhavige uitvinding. De golfcodes die zijn opgeslagen in het golfcodeveld van de spraaksegment-informatie, die ontvangen is uit het spraaksegment-opslagblok, worden door een decoderings-deelblok 9 gedecodeerd in een procedure die het omgekeerde is van de procedure waarin zij werden gecodeerd. De golfsignalen die verkregen zijn
 15 wanneer de golfvormcodes gedecodeerd worden in de golfvormcode-opslagwerkwijze, of de toonhoogte-pulssignalen die verkregen zijn wanneer de broncodes worden gedecodeerd in de broncode-opslagwerkwijze en de spectrale omhullende parameters die gekoppeld worden met de toonhoogte-pulssignalen, worden de golfinformatie genoemd, en worden aan het golfvorm-
 20 samenstellings-deelblok beschikbaar gesteld. Anderzijds is de in het kopveld van de spraaksegment-informatie opgeslagen kopinformatie de invoer voor een duurregelings-deelblok 10 en een toonhoogterege-
 25 lings-deelblok 11.

Het duurregelings-deelblok van figuur 7 ontvangt als invoer de
 25 duurgegevens in de prosodische informatie en de grenstijdstippen die zijn begrepen in de spraaksegment-kopinformatie, en produceert de tijdvervormingsinformatie door de duurgegevens en de grenstijdstippen te gebruiken en verschaft de geproduceerde tijdvervormingsinformatie aan het golfvorm-
 30 samenstellingsdeelblok 13, het toonhoogterege-
 35 lings-deelblok en het energieregelings-deelblok. Als de totale duur van het spraaksegment langer of korter wordt, wordt de duur van deelsegmenten die het spraaksegment vormen overeenkomstig langer of korter, waarbij de verhouding van de expansie of compressie afhangt van de eigenschap van elk deelsegment. In het geval van een klinker met daarvoor en daarachter medeklinkers bijvoor-
 beeld heeft de duur van het constante-toestand-interval dat zich in het midden bevindt een wezenlijk grotere veranderingssnelheid dan de overgangsintervallen aan beide zijden van de klinker. Het duurregelings-deelblok vergelijkt de duur BL van het oorspronkelijke spraaksegment dat is opgeslagen en de duur van het te synthetiseren spraaksegment, aangeduid

door de duurgegevens, en verkrijgt de duur van elk te synthetiseren deel-
segment die overeenkomt met de duur van elk oorspronkelijk deelsegment
door hun veranderingssnelheid of de duurregel te gebruiken, waardoor de
grenstijdstippen van de gesynthetiseerde spraak wordt verkregen. De oor-
5 spronkelijke grenstijdstippen B1, B2, enz. en de grenstijdstippen B'1,
B'2, enz. van de gesynthetiseerde spraak die qua overeenkomst gekoppeld
zijn met de oorspronkelijke grenstijdstippen worden samen de tijdvervorme-
ningsinformatie genoemd, waarbij in het geval van figuur 8 bijvoorbeeld
de tijdvervormingsinformatie kan worden weergegeven door {(B1, B'1), (B1,
10 B'2), (B2, B'3), (B3, B'3), (B4, B'4)}.

De functie van het toonhoogte-regelings-deelblok van figuur 7 is
het op zodanige wijze produceren van de toonhoogte-pulspositie-informa-
tie, dat de gesynthetiseerde spraak het intonatiepatroon heeft dat wordt
aangegeven door de intonatiepatroongegevens, en het te verschaffen aan
15 het golfvorm-samenstellings-deelblok en aan het energieregelings-deel-
blok. Het toonhoogte-regelings-deelblok ontvangt als invoer de intonatie-
patroongegevens die de doeltoonhoogte-frequentiewaarden voor elk foneem
omvatten, en produceert een toonhoogte-contour dat de continue variatie
van de toonhoogte-frequentie met betrekking tot de tijd weergeeft door de
20 doel-toonhoogte-frequentiewaarden vloeiend te verbinden. Het toonhoogte-
regelings-deelblok kan een micro-intonatie-fenomeen vanwege een obstruent
aan de toonhoogte-contour weergeven. In dit geval wordt de toonhoogte-
contour echter een discontinue functie waarin de toonhoogte-frequentie-
waarde abrupt in de tijd varieert op het grenspunt tussen het obstruent-
25 foneem en de aangrenzende andere foneem. De toonhoogte-frequentie wordt
verkregen door de toonhoogte-contour op de eerste toonhoogte-pulspositie
van het spraaksegment te bemonsteren, en de toonhoogte-periode wordt
verkregen door een inverse van de toonhoogte-frequentie te nemen, en
vervolgens wordt het punt voorafgaande aan de toonhoogte-periode bepaald
30 als de tweede toonhoogte-pulspositie. De volgende toonhoogte-periode
wordt dan uit de toonhoogte-frequentie op dat punt verkregen en de vol-
gende toonhoogte-pulsperiode wordt op diens beurt verkregen, en de herha-
ling van een dergelijke procedure zou alle toonhoogte-pulsposities van de
gesynthetiseerde spraak kunnen opleveren. De eerste toonhoogte-pulsposi-
35 tie van het spraaksegment of, in het geval van het eerste spraaksegment
van een serie continu stemhebbende spraaksegmenten van de gesyntheti-
seerde spraak, diens aangrenzende bemonsteringen, kunnen als eerste
bemonstering worden aangemerkt, en de eerste toonhoogte-pulspositie voor
het volgende spraaksegment kan worden aangemerkt als het punt dat over-

eenkomt met de positie van de toonhoogte-puls naast de laatste toonhoogte-puls van het voorafgaande spraaksegment, enz. Het toonhoogteregelings-deelblok zendt de toonhoogte-pulsposities P'1, P'2, enz. van de als zodanig verkregen gesynthetiseerde spraak en de oorspronkelijke toonhoogte-pulsposities P1, P2, enz., die bevat zijn in de spraaksegment-kopinformatie, samen in een bindsel naar het golfvorm-samenstellings-deelblok en het energieregelings-deelblok, waar zij de zogenaamde toonhoogte-pulspositie-informatie vormen. In het geval van figuur 8 bijvoorbeeld kan de toonhoogte-pulspositie-informatie worden weergegeven als {(P1, P2, ... P9), (P'1, P'2, ..., P'8)}.

Het energieregelingsdeelblok van figuur 7 produceert versterkingsinformatie waardoor de gesynthetiseerde spraak het klemtoonpatroon heeft zoals is aangegeven door de klemtoonpatroon-gegevens, en zendt dit naar het golfvorm-samenstellings-deelblok. Het energieregelingsdeelblok ontvangt als invoer de klemtoonpatroongegegevens die de doel-amplitudewaarden voor elk foneem zijn, en produceert een energiecontour die de continue variatie van de amplitude in de tijd weergeeft door deze vloeiend te verbinden. Aangenomen wordt dat de spraaksegmenten vooraf, op het moment van opslag, genormaliseerd worden zodat zij een relatieve energie hebben overeenkomstig de klasse van het spraaksegment om het relatieve verschil in energie voor elk foneem weer te geven. In het geval van klinkers bijvoorbeeld heeft een lage klinker een grotere energie per tijdseenheid dan een hoge klinker, en een nasaal geluid heeft ongeveer de helft van de energie per tijdseenheid vergeleken met een klinker. Verder is de energie tijdens het sluitinterval van een plofgeluid zeer gering. Wanneer de spraaksegmenten worden opgeslagen zullen zij derhalve gecodeerd worden na vooraf te zijn bijgesteld, zodat zij een dergelijke relatieve energie bezitten. In dit geval wordt de in het energieregelings-deelblok geproduceerde energiecontour een versterking die vermenigvuldigd moet worden met de te synthetiseren golfvorm. Het energieregelings-deelblok verkrijgt de versterkingswaarden G1, G2, enz. op elke toonhoogte-pulspositie P'1, P'2, enz. van de synthetische spraak door gebruik te maken van de energiecontour en de toonhoogte-pulspositie-informatie, en verschaft deze aan het golfvorm-samenstellings-deelblok, waarbij deze de versterkingsinformatie wordt genoemd. In het geval van figuur 8 bijvoorbeeld kan de versterkingsinformatie worden weergegeven als {(P'1, G1), (P'2, G2), ..., (P'8, G8)}.

Het golfvorm-samenstellings-deelblok van figuur 7 ontvangt als invoer de in het bovenstaande beschreven golfinformatie, tijdvervormings-

informatie, toonhoogte-pulspositie-informatie en versterkingsinformatie, en produceert uiteindelijk het stemhebbende spraaksignaal. Het golfvorm-samenstellings-deelblok produceert de spraak met het intonatiepatroon, klemtoonpatroon en de duur zoals aangegeven door de prosodische informatie door gebruik te maken van de golfinformatie die van het decoderings-deelblok is ontvangen. Op dit moment worden sommige van de golven herhaald en worden sommige weggelaten. De duurgegevens, intonatiepatroon-gegevens en klemtoonpatroon-gegevens die in de prosodische informatie zijn bevat zijn onderling afhankelijke indicatieve informatie, terwijl zij in onderlinge relatie moeten worden behandeld omdat, wanneer de golfvorm met de golf-informatie wordt gesynthetiseerd, er een onderlinge relatie tussen deze drie informatievormen bestaat. Eén van de belangrijkste problemen bij het samenstellen van een golfvorm is welke golf geselecteerd moet worden als de golf die op elke toonhoogte-pulspositie van de gesynthetiseerde spraak moet worden aangebracht. Als niet de juiste golven geselecteerd en aangebracht worden, kan geen synthetische spraak met een goede kwaliteit worden verkregen. In het onderstaande wordt een beschrijving gegeven van de werking van het golfvorm-samenstellings-deelblok met gebruikmaking van de op tijdvervorming gebaseerde golfverplaatsingswerkwijze van de onderhavige uitvinding, die een golfverplaatsingswerkwijze is welke in staat is om een hoge kwaliteit bij het synthetiseren van de spraak te verkrijgen door gebruik te maken van de vanuit het spraaksegment-opslagblok verkregen spraaksegmentinformatie.

De golfvorm-synthese-procedure van stemhebbende spraak van het golfvorm-samenstellings-deelblok bestaat uit twee trappen, dat wil zeggen de golfvorm-verplaatsings-trap met gebruikmaking van de tijdvervormingsfunctie en de superpositie-trap voor het superponeren van de verplaatste golven.

Dat wil zeggen, in het geval van de golfvorm-code-opslagwerkwijze worden uit de als golfinformatie ontvangen golfsignalen de meest geschikte geselecteerd voor de toonhoogte-pulsposities van de synthetische spraak en worden op hun toonhoogte-pulsposities geplaatst, en hun versterkingen worden ingesteld, en daarna wordt de gesynthetiseerde spraak geproduceerd door hen te superponeren.

In de broncode-opslag-werkwijze worden het toonhoogte-pulssignaal en de spectrale-omhullende-parameters voor elke periode die overeenkomt met het toonhoogte-puls-signaal als golf-informatie ontvangen. In dit geval zijn twee synthetische spraaksamenstellingswerkwijzen mogelijk. De eerste werkwijze is het verkrijgen van elke golf door het aan het synthe-

sefilter verschaffen van de spectrale-omhullende-parameters en het toonhoogte-pulssignaal voor een intervallengte van twee tot vier perioden, verkregen door het uitvoeren van de procedures die overeenkomen met de rechterzijde van het buffer van figuur 4, dat wil zeggen het in het

5 bovenstaande beschreven parameter-verlengen en het nullen-toevoegen aan de golfinformatie, en het vervolgens met de golven samenstellen van de synthetische spraak volgens een procedure die identiek is aan die in de golfvorm-codeopslag-werkwijze. Deze werkwijze is in principe gelijk aan het samenstellen van de synthetische spraak in de golfvorm-codeopslag-

10 werkwijze, en derhalve zal worden afgezien van een afzonderlijke beschrijving. De tweede werkwijze omvat het verkrijgen van een synthetisch toonhoogte-pulstreinsignaal of synthetisch activeringssignaal met een vlakke spectraleomhullende maar met een toonhoogte-patroon dat verschilt van dat van het oorspronkelijke periodieke toonhoogte-pulstreinsignaal door het selecteren van de meest geschikte voor de toonhoogte-

15 pulsposities van de synthetische spraak onder de toonhoogte-pulssignalen en het plaatsen van hen en het instellen van hun versterkingen, en het daarna superponeren van hen, en het verkrijgen van synthetische spectrale-omhullende-parameters die vervaardigd worden door het in verband brengen van de spectrale-omhullende-parameter met elk toonhoogte-pulssignaal

20 dat het synthetische toonhoogte-pulstreinsignaal of synthetische activeringssignaal vormt, en het vervolgens produceren van de gesynthetiseerde spraak door het verschaffen van het synthetische activeringssignaal en de synthetische spectrale-omhullende-parameters aan het synthesesfilter. Deze

25 twee werkwijzen zijn in wezen identiek, afgezien van het feit dat de volgorde tussen het synthesesfilter en de superpositieprocedure in de samenstelling van de synthetische spraak is omgekeerd.

De in het bovenstaande beschreven synthetische spraak-samenstelling-werkwijze wordt in het onderstaande onder verwijzing naar figuur 8

30 beschreven. De golf-verplaatsings-werkwijze kan in principe zowel in de golfvorm-codeopslag-werkwijze als de broncode-opslag-werkwijze worden toegepast. Derhalve zullen de synthetische spraak-golfvorm-samenstellingsprocedures in de twee werkwijzen tegelijkertijd onder verwijzing naar figuur 8 worden beschreven.

35 In figuur 8A is de correlatie tussen het oorspronkelijke spraaksegment en het te synthetiseren spraaksegment weergegeven. De oorspronkelijke grenstijdstippen B1, B2, enz., aangegeven door stippellijnen, de grenstijdstippen B'1, B'2, enz. van de gesynthetiseerde klank en de correlatie tussen deze, aangegeven door arceringen, zijn bevat in de tijd-

vervormingsinformatie die uit het duurregelings-deelblok wordt ontvangen. Bovendien zijn de oorspronkelijke toonhoogte-pulsposities P1, P2, enz., aangegeven door de getrokken lijnen, en zijn de toonhoogte-pulsposities P'1, P'2, enz. van de gesynthetiseerde klank bevat in de vanuit het toon-
 5 hoogte-besturings-deelblok ontvangen toonhoogte-pulspositie-informatie. Ten behoeve van de uitleg van figuur 8 wordt aangenomen dat de toonhoogte-periode van de oorspronkelijke spraak en de toonhoogte-periode van de gesynthetiseerde klank (geluid) respectievelijk constant zijn en dat de laatste 1,5 maal de eerste bedraagt.

10 Het golfvorm-samenstellings-deelblok vormt eerst de tijdvervormingsfunctie, zoals getoond in figuur 8B, door gebruik te maken van de oorspronkelijke grenstijdstippen, de grenstijdstippen van het gesynthetiseerde geluid en de correlatie tussen deze. De abscis van de tijdvervormingsfunctie geeft de tijd "t" van het oorspronkelijke spraaksegment
 15 weer, en de ordinaat geeft de tijd "t" van het te synthetiseren spraaksegment weer. Omdat bijvoorbeeld in figuur 8A het eerste deelsegment en het laatste deelsegment van het oorspronkelijke spraaksegment respectievelijk gecomprimeerd zouden moeten worden tot $\frac{2}{3}$ van de tijd en geëxpandeerd worden tot twee maal de tijd, verschijnt de correlatie daarvan in
 20 de tijdvervormingsfunctie van figuur 8B respectievelijk als de lijnen met de helling van $\frac{2}{3}$ en 2. Het tweede deelsegment varieert niet in duur en verschijnt derhalve in de tijdvervormingsfunctie als een lijn met de helling 1. Het tweede deelsegment van het te synthetiseren spraaksegment resulteert uit de herhaling van het grenstijdstip "B1" van het oorspronkelijke spraaksegment, en in tegenstelling daarmee wordt het derde deel-
 25 segment van het oorspronkelijke spraaksegment gevarieerd tot één grenstijdstip "B'3" in het te synthetiseren spraaksegment. De correlaties in dergelijke gevallen verschijnen respectievelijk als een verticale lijn en als een horizontale lijn. De tijdvervormingsfunctie wordt dus verkregen
 30 door het grenstijdstip van het oorspronkelijke spraaksegment en het grenstijdstip van het te synthetiseren spraaksegment in overeenstemming met het grenstijdstip van het oorspronkelijke spraaksegment als twee punten te presenteren en door deze met een lijn te verbinden. Het kan in sommige gevallen mogelijk zijn om de correlatie tussen de deelsegmenten meer in
 35 overeenstemming met de werkelijkheid te presenteren door de punten met een vloeiende lijn te verbinden.

In de golfvorm-codeopslag-werkwijze bepaalt het golfvorm-samenstellings-deelblok het oorspronkelijke tijdstip dat overeenkomt met de toonhoogte-pulspositie van het synthetische geluid door gebruik te maken van

de tijdvervormingsfunctie, en bepaalt de golf waarvan de toonhoogte-pulspositie het dichtst bij het oorspronkelijke tijdstip ligt en plaatst de golf vervolgens op de toonhoogte-pulspositie van het synthetische geluid.

In de volgende trap vermenigvuldigt het golfvorm-samenstellings-
 5 deelblok elk geplaatst golfsignaal met de versterking die overeenkomt met de toonhoogte-pulspositie van het golfsignaal dat uit de versterkingsinformatie is bepaald, en verkrijgt uiteindelijk het gewenste synthetische geluid door de qua versterking aangepaste golfsignalen door eenvoudig optellen daarvan te superponeren. In figuur 3Q is het door een derge-
 10 lijke superpositieprocedure geproduceerde synthetische geluid weergegeven voor het geval waarin de golven van fig. 3I, fig. 3L, fig. 3(0) als in fig. 3P worden verplaatst oftewel "herplaatst".

Op overeenkomstige wijze bepaalt in de broncode-opslag-werkwijze het golfvorm-samenstellings-deelblok het oorspronkelijke tijdstip dat
 15 overeenkomt met de toonhoogte-pulspositie van het synthetische geluid door gebruik te maken van de tijdvervormingsfunctie, en bepaalt het toonhoogte-pulssignaal waarvan de toonhoogte-pulspositie zich het dichtst bij het oorspronkelijke tijdstip bevindt, en plaatst vervolgens het toonhoogte-pulssignaal op de toonhoogte-pulspositie van het synthetische geluid.
 20 De nummers voor de toonhoogte-pulssignalen of de golven die op deze wijze op elke toonhoogte-pulspositie van het te synthetiseren spraaksegment zijn geplaatst, zijn in de figuren 8A en 8B getoond. Zoals uit de tekeningen kan worden opgemaakt, zijn enkele van de golven die het oorspronkelijke spraaksegment vormen weggelaten vanwege de compressie van de
 25 deelsegmenten, en enkele zijn vanwege de expansie van de deelsegmenten herhaald gebruikt. In figuur 8 werd aangenomen dat het toonhoogte-pulssignaal voor elke periode verkregen werd door direct achter elke toonhoogte-puls te segmenteren.

De superpositie van de golven in de golfvorm-codeopslag-werkwijze
 30 is equivalent aan de superpositie van de toonhoogte-pulssignalen in de broncode-opslag-werkwijze. In het geval van de broncode-opslag-werkwijze vermenigvuldigt het golfvorm-samenstellings-deelblok derhalve elk verplaatst toonhoogte-pulssignaal met de, met de toonhoogte-pulspositie van het verplaatste toonhoogte-pulssignaal overeenkomende, versterking die
 35 bepaald is uit de versterkingsinformatie, en verkrijgt uiteindelijk het gewenste synthetische activeringssignaal door het superponeren van de qua versterking aangepaste toonhoogte-pulssignalen. Aangezien de meeste energie geconcentreerd is in de toonhoogte-puls kan het echter in dit geval mogelijk zijn om het synthetische activeringssignaal te maken door eerst

een synthetische activeringssignaal zonder versterkingsaanpassing te verkrijgen door de geplaatste toonhoogte-pulssignalen te superponeren en vervolgens het synthetische activeringssignaal zonder versterkingsaanpassing te vermenigvuldigen met het energie-contour dat in het energie-regelings-deelblok is gegenereerd, in plaats van het superponeren van de
 5 qua constante versterking aangepaste toonhoogte-pulssignalen. Figuur 3R toont het synthetische activeringssignaal dat verkregen wordt wanneer de toonhoogte-pulssignalen van de figuren 3H, 3K en 3N volgens een dergelijke procedure worden verplaatst, zodat het toonhoogte-patroon hetzelfde
 10 wordt als voor het geval van figuur 3P.

In de broncode-opslag-werkwijze moet het golfvorm-samenstellings-deelblok de synthetische spectrale-omhullende-parameters maken, en twee manieren zijn mogelijk, d.w.z. de in figuur 8A getoonde werkwijze met compressie en expansie in de tijd en de in figuur 8B getoonde synchrone
 15 correspondentiewerkwijze. Als de spectrale-omhullende-parameters continue functies van de tijd zijn en de omhullende van het spraakspectrum volledig weergeven, kunnen de synthetische spectrale-omhullende-parameters op eenvoudige wijze worden verkregen door de oorspronkelijke spectrale-omhullende-parameters op een deelsegment-voor-deelsegment-basis in de tijd
 20 te comprimeren of te expanderen. In figuur 8A is de door de sequentiële analyse-werkwijze verkregen spectrale-omhullende-parameter weergegeven als een onderbroken curve en is de spectrale-omhullende-parameter, die gecodeerd is door de curve te benaderen door verschillende punten zoals A, B, C enz. met lijnsegmenten te verbinden, met een getrokken lijn weer-
 25 gegeven. Aangezien slechts de tijdposities van elk punt variëren om de punten A', B', C', enz. als resultaat van de compressie en expansie in de tijd op te leveren, is een dergelijke lijnsegment-coderings-werkwijze in het bijzonder geschikt voor het geval van een compressie en expansie in de tijd. In het geval van het gebruik van de blok-analyse-werkwijze of de
 30 synchrone toonhoogte-analyse-werkwijze kan de werkwijze met compressie en expansie in de tijd niet de gewenste kwaliteit van het synthetische geluid geven aangezien de spectrale aanpassing niet nauwkeurig is en de variatie in de tijd van de spectrale-omhullende-parameter discontinu is, in welk geval het de voorkeur verdient om de synchrone-correspondentie-
 35 werkwijze te gebruiken, waarin de synthetische spectrale-omhullende-parameters worden samengesteld door de spectrale-omhullende-parameters voor elk toonhoogte-periode-interval te correleren met elk overeenkomstige toonhoogte-pulssignaal, zoals in figuur 8B is getoond. Dat wil zeggen, aangezien de golf in de golfvorm-codeopslag-werkwijze equivalent is aan

het toonhoogte-pulssignaal en de corresponderende spectrale-omhullende-parameters voor hetzelfde toonhoogte-periode-interval, kunnen de synthetische spectrale-omhullende-parameters worden gemaakt door de spectrale-omhullende-parameters voor één periode-interval synchroon te plaatsen in
 5 hetzelfde periode-interval van elk geplaatst toonhoogte-pulssignaal. In figuur 8B zijn k_1 , dat één van de spectrale-omhullende-parameters is, en k'_1 , dat de met k_1 overeenkomende synthetische spectrale-omhullende-parameter is welke samengesteld is door zulke werkwijzen voor de blokanalyse-werkwijze en de synchrone toonhoogte-analyse-werkwijze, respectievelijk
 10 met een getrokken lijn en een onderbroken lijn getoond. Zoals in het bovenstaande is vermeld kan uiteraard met de door de sequentiële analyse-werkwijze verkregen spectrale-omhullende-parameter de synthetische spectrale-omhullende-parameter volgens de werkwijze van figuur 8A worden samengesteld. Als bijvoorbeeld het toonhoogte-pulssignaal voor elke
 15 periode is verplaatst zoals in figuur 3R is getoond, zijn de spectrale-omhullende-parameters voor elke periode overeenkomstig de toonhoogte-pulssignalen geplaatst zoals in figuur 3S is getoond.

Op het moment van het samenstellen van het synthetische activeringingssignaal en de synthetische spectrale-omhullende-parameters in de
 20 broncode-opslag-werkwijze resulteert een leeg interval tussen twee aangrenzende toonhoogte-periode-intervallen, zoals met arceringslijnen in figuur 8 is getoond, indien de toonhoogte-periode van het gesynthetiseerde geluid langer is dan de oorspronkelijke toonhoogte-periode. Indien de toonhoogte-periode van het gesynthetiseerde geluid korter is dan de
 25 oorspronkelijke toonhoogte-periode treden overlappingsintervallen op waarin twee aangrenzende toonhoogte-periode-intervallen elkaar overlappen. Het overlappingsinterval "fb" en het lege interval "gh" zijn bij wijze van voorbeeld in de figuren 3R en 3S getoond. Zoals in het voorgaande is beschreven zullen de verplaatste toonhoogte-pulssignalen op het
 30 tijdstip van overlappen worden gesuperponeerd. Het is echter redelijk dat de overeenkomstig de toonhoogte-pulssignalen verplaatste spectrale-omhullende-parameters op het tijdstip van overlappen worden gemiddeld in plaats van gesuperponeerd. De samenstellingswerkwijze van het synthetische activeringingssignaal en de synthetische spectrale-omhullende-parameters, waarin de lege intervallen en de overlappingsintervallen worden
 35 meegenomen, is als volgt.

De bemonsteringen met de waarde nul worden op het moment van het samenstellen van het synthetische activeringingssignaal in het lege interval ingevoegd. In het geval van stemhebbend fricatief geluid kan een natuur-

lijker geluid gesynthetiseerd worden indien in plaats van de bemonsteringen met de waarde nul het hoogdoorlaat-gefilterde ruissignaal in het lege interval wordt ingevoegd. De verplaatste toonhoogte-pulssignalen moeten in het overlappingsinterval worden opgeteld. Aangezien een dergelijke
 5 toevoegingswerkwijze hinderlijk is, is het geschikt om een afknottingswerkwijze te gebruiken waarin slechts één signaal uit twee elkaar in het overlappingsinterval overlappende toonhoogte-pulssignalen wordt geselecteerd. De kwaliteit van het gesynthetiseerde geluid met gebruikmaking van de afknottingsmethode wordt niet significant verminderd. In figuur 3R was
 10 het lege interval "gh" gevuld met bemonsteringen met de waarde nul, en werd het toonhoogte-pulssignaal van het voorste interval in het overlappingsinterval fb geselecteerd. Dat wil zeggen, in het geval van het plaatsvinden van een overlapping werd de voorste van de overlappingsintervallen van elk toonhoogte-pulssignaal afgeknot, en deze werkwijze
 15 bezit fysisch meer betekenis dat de werkwijze waarin de toonhoogte-pulssignalen worden gemaakt door direct vóór de toonhoogte-puls te segmenteren en waarin tijdens de synthese de laatste van de overlappingsintervallen van het toonhoogte-pulssignaal wordt afgeknot indien zij elkaar overlappen, zoals eerder is beschreven. In werkelijkheid geven de werk-
 20 wijzen geen significant verschil in de geluidskwaliteit van het gesynthetiseerde geluid.

Op het moment van het samenstellen van de synthetische spectrale-omhullende-parameter is het ideaal dat het lege interval gevuld is met de waarden die lineair variëren van een waarde van de spectrale-omhullende-
 25 parameter aan het eind van het voorgaande periode-interval tot een waarde van de spectrale-omhullende-parameter aan het beginpunt van de volgende periode, en dat in het overlappingsinterval de spectrale-omhullende-parameter geleidelijk varieert van de spectrale-omhullende-parameter van de voorgaande periode tot die van de volgende periode door gebruikmaking van
 30 de interpolatiemethode waarin het gemiddelde van twee overlappende spectrale-omhullende-parameters wordt verkregen met weegwaarden die lineair in de tijd variëren. Aangezien deze werkwijzen echter hinderlijk zijn kan de volgende werkwijze worden gebruikt die handiger is en die de geluidskwaliteit niet significant vermindert. Dat wil zeggen, voor de
 35 spectrale-omhullende-parameter in het lege interval kan, zoals in figuur 8B, de waarde van de spectrale-omhullende-parameter aan het eind van het voorgaande periode-interval herhaald worden gebruikt, of kan de waarde van de spectrale-omhullende-parameter aan het begin van het volgende periode-interval herhaald worden gebruikt, het rekenkundige gemiddelde

van de twee spectrale-omhullende-parameters kan worden gebruikt, of de waarden van de spectrale-omhullende-parameter aan de eind- en aan de beginpunten van de voorgaande en de volgende periode-intervallen kunnen respectievelijk voor en na het midden van het lege interval, dat een
 5 grens vormt, worden gebruikt. Voor de spectrale-omhullende-parameter in het overlappingsinterval kan eenvoudigweg een willekeurig deel worden gekozen dat overeenkomt met de geselecteerde toonhoogte-puls. Aangezien in figuur 3S bijvoorbeeld het toonhoogte-pulssignaal voor het voorgaande periode-interval geselecteerd was als het synthetische activeringssignaal
 10 in het overlappingsinterval "fb" waren de parameterwaarden voor het voorgaande periode-interval op overeenkomstige wijze gekozen als de synthetische spectrale-omhullende-parameters. In het lege interval "gh" van figuur 8B en 3S werden de parameterwaarden van de spectrale-omhullende-parameter aan het eind van het voorgaande periode-interval herhaald
 15 gebruikt. In het geval van figuur 3S, waarin de spectrale-omhullende-parameter een continue functie van de tijd is, leveren de werkwijze waarin de laatste waarde van het voorgaande periode-interval of de eerste waarde van het volgende periode-interval gedurende het lege interval herhaald wordt gebruikt, en de werkwijze waarin de twee waarden gedurende
 20 het lege interval lineair worden gevarieerd, hetzelfde resultaat op.

Als eenmaal alle synthetische activeringssignalen en synthetische spectrale-omhullende-parameters voor een segment zijn samengesteld, vlakt het golfvorm-samenstellingsdeelblok normaal beide uiteinden van de samengestelde synthetische spectrale-omhullende-parameters af met gebruik-
 25 making van de interpolatiewerkwijze, zodat de variatie van de spectrale-omhullende-parameters tussen aangrenzende spraaksegmenten vloeiend is. Als het synthetische activeringssignaal en de als boven samengestelde spectrale-omhullende-parameters respectievelijk als activeringssignaal en filtercoëfficiënten in het golfvorm-samenstellingsdeelblok in het synthe-
 30 sefilter worden ingevoerd, wordt het gewenste synthetische geluid uiteindelijk door het synthesefilter afgegeven. Het synthetische activerings-signaal dat verkregen wordt wanneer de toonhoogte-pulssignalen van de figuren 3H, 3K en 3N zodanig worden verplaatst dat het toonhoogte-patroon gelijk is aan dat van figuur 3P, is in figuur 3R getoond, en de syntheti-
 35 sche spectrale-omhullende-parameters die verkregen zijn door het doen corresponderen van spectrale-omhullende-parameters voor één periode van figuur 3G, 3J en 3M met de toonhoogte-pulssignalen in het synthetische activeringssignaal van figuur 3R, zijn getoond in figuur 3S. Het vormen van een in de tijd variërend synthesefilter met als filtercoëfficiënten

de variërende reflectiecoëfficiënten zoals in figuur 3S zijn getoond, en het invoeren van het synthetische activeringssignaal zoals in figuur 3R is getoond in het in de tijd variërende synthesefilter, levert het gesynthetiseerde geluid van de figuur 3T op, dat vrijwel gelijk is aan het
 5 gesynthetiseerde geluid van figuur 3P.

Als nu de golfvorm-code-opslagwerkwijze en de broncode-opslagwerkwijze worden vergeleken kunnen de twee werkwijzen qua principe als identiek worden beschouwd. Wanneer de spraaksegmenten met een slechte mogelijkheid tot verbinden aan elkaar worden geschakeld bestaat er het verschil dat het in het geval van het broncode-opslagwerkwijze door gebruik-
 10 making van de interpolatiewerkwijze mogelijk is om vloeiend verbonden geluid te synthetiseren door de spectrale-omhullende-parameters af te vlakken, maar dat dit in het geval van de golfvorm-codeopslag-werkwijze onmogelijk is. Verder vereist de broncode-opslagwerkwijze een kleiner
 15 geheugen dan de golfvorm-code-opslagwerkwijze, aangezien de golfvorm van slechts één periodelengte per golf in de broncode-opslagwerkwijze opgeslagen moet worden, en deze heeft het voordeel dat het makkelijk is om de functie van het spraakhebbend-geluid-syntheseblok en de functie van het bovenbeschreven niet-spraakhebbend-geluid-syntheseblok te integreren. In
 20 het geval waarin de homomorfe analysewerkwijze wordt gebruikt kan het cepstrum of de impulsresponsie in de golfvorm-codeopslag-werkwijze gebruikt worden als de verzameling van spectrale-omhullende-parameters, terwijl het in de broncode-opslagwerkwijze praktisch onmogelijk is om het cepstrum te gebruiken dat de op blokken gebaseerde berekening vereist,
 25 aangezien de duur van het syntheseblok met de waarden van de constante synthetische spectrale-omhullende-parameters van blok tot blok varieert, zoals kan worden opgemaakt uit de synthetische spectrale-omhullende-parameter van figuur 8B die door een getrokken lijn is weergegeven. De broncode-opslagwerkwijze volgens de onderhavige uitvinding maakt gebruik van
 30 de toonhoogte-puls van één periode als de activeringspuls. Dit verschilt echter in zoverre van de gebruikelijke toonhoogte-puls-activeringswerkwijze volgens de stand van de techniek die beoogt de puls te substitueren door een monster-toonhoogte-puls, dat bij de onderhavige uitvinding de toonhoogte-puls van elke periode en de spectrale-omhullende-parameters
 35 van elke periode in overeenstemming met de toonhoogte-puls worden samengevoegd om de golf van elke periode te produceren.

Zoals uit de bovenstaande beschrijving kan worden opgemaakt is de onderhavige uitvinding geschikt voor het coderen en decoderen van een spraaksegment van een tekst-naar-spraak-synthesestelsel van de segmentele

spraak-synthese-werkwijze. Aangezien de onderhavige uitvinding een werkwijze omvat waarin de totale en gedeeltelijke duur en toonhoogte-patroon van willekeurige fonetische eenheden zoals foneem, halflettergreep, difoon en deelsegment, enz., die spraak vormen, vrij en onafhankelijk
 5 kunnen worden gewijzigd, kan de onderhavige uitvinding gebruikt worden in een spraaksnelheid-omzettingssysteem of een tijdschaal-modificatiesysteem dat de stemsnelheid met een constante verhouding zodanig wijzigt dat deze sneller of langzamer is dan de oorspronkelijke snelheid zonder het intonatiepatroon van de spraak te veranderen, en de uitvinding kan ook worden
 10 gebruikt bij een zang-stemsynthesestelsel of een spraakcoderingssysteem met een zeer lage snelheid, zoals een fonetische vocoder of een segment-vocoder die de spraak overdraagt door de duur en toonhoogte van vooraf opgeslagen spraaksegmenten te veranderen.

Een ander toepassingsgebied van de onderhavige uitvinding is een
 15 muzikaal geluid-synthese-systeem, zoals een op de bemonsteringswerkwijze gebaseerd elektronisch muziekinstrument. Aangezien vrijwel alle het geluid binnen het gamma van de elektronische muziekinstrumenten qua golfvorm digitaal zijn gecodeerd, opgeslagen en gereproduceerd wanneer zij in de stand van de techniek wat betreft bemonsteringswerkwijzen voor elek-
 20 tronische muziekinstrumenten vanuit het toetsenbord en dergelijke werden opgevraagd, bestaat er het nadeel dat een grote hoeveelheid geheugen benodigd is voor de opslag van het muziekgeluid. Indien echter de periodieke golfvormontledings- en golfverplaatsingswerkwijze van de onderhavige uitvinding wordt gebruikt, kan de vereiste hoeveelheid geheugen aan-
 25 zienlijk worden gereduceerd aangezien de klanken van verschillende toonhoogten gesynthetiseerd kunnen worden door de tonen van slechts enkele soorten toonhoogten te bemonsteren. Het muziekgeluid bestaat kenmerkend uit drie gedeelten, dat wil zeggen een inzet, een aangehouden deel en een afvallend deel. Aangezien de spectraleomhullende niet alleen tussen de
 30 drie delen maar ook binnen het aangehouden deel geleidelijk varieert, verandert het timbre ook overeenkomstig. Als de muzikale geluidssegmenten derhalve volgens de bovenbeschreven periodieke golfvorm-ontledingswerkwijze worden gecodeerd en opgeslagen, waarbij de geschikte punten waarop het spectrum wezenlijk varieert als de grenstijdstippen worden genomen,
 35 en als het geluid volgens de bovenbeschreven, op tijdvervorming gebaseerde golf-verplaatsingswerkwijze wordt gesynthetiseerd wanneer er verzoeken van het toetsenbord en dergelijke zijn, dan kan een muziekgeluid met een willekeurige gewenste toonhoogte worden gesynthetiseerd. In gevallen waarin het muziek-geluidssignaal echter volgens de lineaire predictieve ana-

lysewerkwijze wordt gedeconvolveerd omdat er een tendens is dat de preciese spectrale omhullende niet is verkregen en de toonhoogte-puls niet scherp is, wordt aanbevolen om het aantal voor analyse gebruikte spectrale-omhullende-parameters te reduceren en het signaal voor de analyse te
5 differentiëren.

Hoewel de uitvinding in diens voorkeursvorm met een zekere mate van specificiteit is beschreven zullen deskundigen inzien dat de onderhavige beschrijving van de voorkeursvorm slechts bij wijze van voorbeeld is beschreven en dat een groot aantal wijzigingen kunnen worden aangebracht
10 zonder buiten het wezen en het kader van de uitvinding te treden.

Conclusies

1. Spraaksegment-coderingswerkwijze voor gebruik in een spraaksynthesestelsel, omvattende:

- 5 het vormen van golven door het in elk analyse-tijdsinterval verkrijgen van parameters die een spectraleomhullende weergeven door een periodiek of quasi-periodiek digitaal signaal te analyseren door gebruikmaking van een spectrumschattingstechniek, door het oorspronkelijke signaal te deconvolueren tot een impulsresponsie die door de spectrale-om-
- 10 hullende-parameters wordt weergegeven en een periodiek of quasi-periodiek toonhoogte-pulstreinsignaal met een vrijwel vlakke spectraleomhullende, en door het convolueren van een activeringssignaal dat verkregen is door het toevoegen van bemonsteringen met de waarde nul na een toonhoogte-pulssignaal van één periode dat verkregen is door het toonhoogte-puls-
- 15 treinsignaal voor elke periode te segmenteren zodat in elke periode één toonhoogte-puls is bevat en een impulsresponsie die overeenkomt met de spectrale-omhullende-parameters in hetzelfde tijdsinterval als het activeringssignaal; en

- het qua golfvorm coderen van golven van elke periode, het in geheugen opslaan daarvan, en het tijdens de spraaksynthese decoderen van de
- 20 golven, het instellen van de duur en de toonhoogtefrequentie door het zodanig toewijzen van de golven aan geschikte tijdstippen dat zij het gewenste toonhoogte-patroon bezitten, het op de genoemde tijdstippen plaatsen van hen en het superponeren van hen, waardoor spraak wordt
- 25 gesynthetiseerd.

2. Spraaksegment-coderingswerkwijze volgens conclusie 1, waarin de golven worden gevormd door het samenbrengen van informatie die verkregen is door het qua golfvorm coderen van een toonhoogte-pulssignaal van elke periode dat verkregen is door segmentatie met informatie welke verkregen
- 30 is door het coderen van een verzameling van spectrale-omhullende-parameters van één periode tijdens hetzelfde tijdsinterval en het opslaan van de informatie in de geheugens, en waarbij tijdens de synthese golven worden gemaakt door het convolueren van een activeringssignaal dat verkregen is door het toevoegen van bemonsteringen met de waarde nul aan een
- 35 toonhoogte-pulssignaal van één periode dat verkregen is door het decoderen van de genoemde informatie en een impulsresponsie die overeenkomt met de gedecodeerde spectrale-omhullende-parameters in hetzelfde tijdsinterval als het activeringssignaal.

3. Spraaksegment-coderingswerkwijze volgens conclusie 2, waarin het

gesynthetiseerde spraakgeluid tijdens de synthese wordt geproduceerd door het convolueren van een synthetisch activeringssignaal dat gevormd is door het zodanig aan geschikte tijdstippen toewijzen van door decoderen van de informatie verkregen toonhoogte-pulssignalen dat zij een gewenst
 5 toonhoogte-patroon bezitten, en het op de tijdstippen plaatsen daarvan, waarbij een leeg interval dat plaatsvindt wanneer een gewenste toonhoogte-periode langer is dan een oorspronkelijke toonhoogte-periode, gevuld wordt met bemonsteringen met de waarde nul en waarbij in een overlappingsinterval, dat optreedt wanneer de gewenste toonhoogte-periode korter
 10 is dan oorspronkelijke toonhoogte-periode, de overlappende toonhoogte-pulssignalen bij elkaar worden opgeteld of één van de signalen wordt geselecteerd, en een in de tijd variërende impulsresponsie die overeenkomt met een verzameling van synthetische spectrale-omhullende-parameters die ofwel gevormd wordt door het in de tijd comprimeren of expanderen van
 15 de verzameling van tijdfuncties van de parameters op een segment-voor-segment-basis, afhankelijk of de duur van een deelsegment in een te synthetiseren spraaksegment respectievelijk korter of langer is dan die van een overeenkomstig deelsegment in het oorspronkelijke spraaksegment, of door het lokaliseren van de verzameling van tijdfuncties van de parameters van één periode, synchroon met het samengebrachte toonhoogte-pulssignaal van één periode dat geplaatst is om het synthetische activerings-
 20 signaal te vormen, waarbij in het laatste geval een synthetische spectrale-omhullende-parameter in het lege interval wordt verkregen door de waarde van de spectrale-omhullende-parameter aan het eind van de voorgaande periode of de waarde van de spectrale-omhullende-parameter aan het begin van de volgende periode te herhalen of een gemiddelde waarde van de twee waarden te bepalen of door te vullen met waarden die de twee waarden vloeiend verbinden, of door de waarden van de spectrale-omhullende-parameters aan het eind en het begin van de voorgaande en volgende periodes
 25 voor en na het midden van het lege interval te herhalen en waarbij de synthetische spectrale-omhullende-parameter in het overlappingsinterval verkregen wordt door één van de overlappende spectrale-omhullende-parameters te selecteren of door een gemiddelde waarde van de twee overlappende parameters te gebruiken.

35 4. Toonhoogteregelingswerkwijze voor een spraaksynthesestelsel, omvattende:

het tegelijkertijd regelen van de duur en toonhoogte van een spraaksegment door een op tijdvervorming gebaseerde golf-verplaatsingswerkwijze door het coderen van grenstijdstippen met inbegrip van beginpunt,

eindpunt en stationair punt in spraaksegment- en toonhoogte-pulsposities van elke golf of elk toonhoogte-pulssignaal van één periode, en het tijdens het opslaan van elk spraaksegment tegelijkertijd opslaan daarvan in geheugens, en tijdens de synthese verkrijgen van een tijdvervormings-
 5 functie door het vergelijken van gewenste grenstijdstippen en oorspronkelijke grenstijdstippen die in overeenstemming met de gewenste grenstijdstippen zijn opgeslagen, het bepalen van de oorspronkelijke tijdstippen die overeenkomen met elke gewenste toonhoogte-pulspositie door gebruik te maken van de tijdvervormingsfunctie, het selecteren van golven met toon-
 10 hoogte-pulsposities die zich het dichtst bij de oorspronkelijke tijdstippen bevinden en het plaatsen daarvan op gewenste toonhoogte-pulsposities, en het superponeren van de golven.

5. Toonhoogte-regelingswerkwijze voor een spraaksynthesestelsel volgens conclusie 4, verder omvattende het produceren van synthetische
 15 spraak door het selecteren van toonhoogte-pulssignalen van één periode en spectrale-omhullende-parameters die overeenkomen met de toonhoogte-pulssignalen en het plaatsen daarvan, en het convolueren van de geplaatste toonhoogte-pulssignalen en impulsresponsie in overeenstemming met de genoemde spectrale-omhullende-parameters om golven te produceren, en het
 20 superponeren van de geproduceerde golven.

6. Toonhoogteregeleingswerkwijze voor een spraaksynthesestelsel volgens conclusie 4, verder omvattende het produceren van gesynthetiseerde spraak door het selecteren van toonhoogte-pulssignalen van één
 periode en spectrale-omhullende-parameters die overeenkomen met de toon-
 25 hoogte-pulssignalen en het plaatsen daarvan, en het convolueren van een synthetische activeringssignaal dat verkregen is door het superponeren van de gelokaliseerde toonhoogte-pulssignalen volgens de werkwijze van conclusie 3, en een in de tijd variërende impulsresponsie die overeenkomt met synthetische spectrale-omhullende-parameters welke gemaakt zijn door
 30 het aaneenvoegen van de geplaatste spectrale-omhullende-parameters volgens de werkwijze van conclusie 3.

7. Stemhebbende-spraak-synthese-inrichting voor gebruik in een spraaksynthesestelsel, omvattende:

een decoderings-deelblok (9) dat golfinformatie produceert door het
 35 decoderen van golfcodes van het spraaksegment-opslagblok (5);

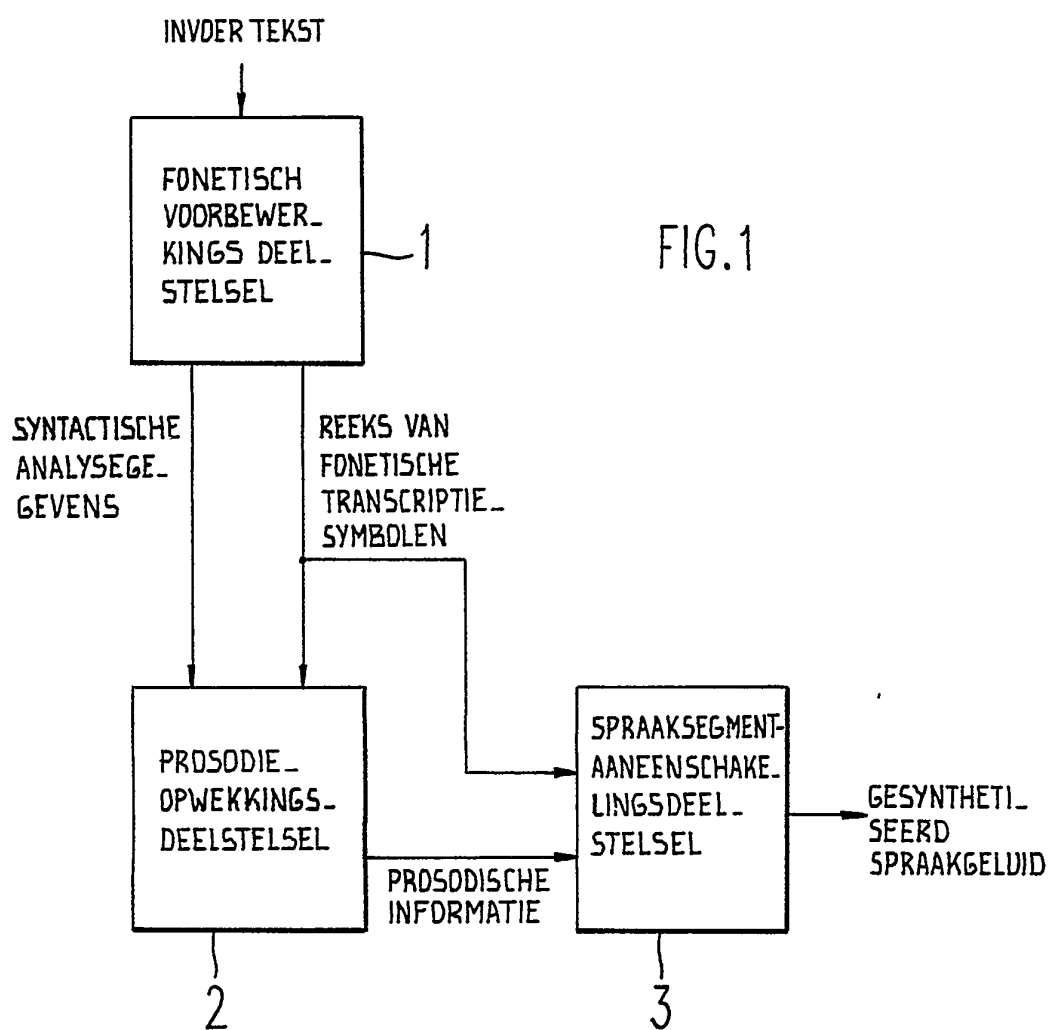
een duurregelings-deelblok (10) dat tijdvervormingsinformatie produceert uit een invoer van duurgegevens van een prosodie-opwekkings-deelstelsel (2) en grenstijdstippen die bevat zijn in kopinformatie van het spraaksegment-opslagblok (5);

een toonhoogte-regelingsdeelblok (11) dat zodanig toonhoogte-puls-
positie-informatie produceert dat het een intonatiepatroon heeft zoals
aangegeven wordt door intonatiepatroon-gegevens uit invoer van de kopin-
formatie van het spraaksegment-opslagblok (5), de intonatiepatroon-gege-
5 vens van het prosodie-opwekkings-deelstelsel en de tijdvervormingsi-
nformatie van het duurregelings-deelblok (10);

een energieregelings-deelblok (12) dat zodanig versterkingsinform-
atie produceert dat de gesynthetiseerde spraak een klemtoonpatroon heeft
zoals aangegeven wordt door de klemtoonpatroon-gegevens uit invoer van de
10 klemtoonpatroon-gegevens van het prosodie- opwekkings-deelstelsel (2), de
tijdvervormingsinformatie van het duurregelings-deelblok (10) en toon-
hoogte-pulspositie-informatie van het toonhoogte-regelingsdeelblok (11);
en

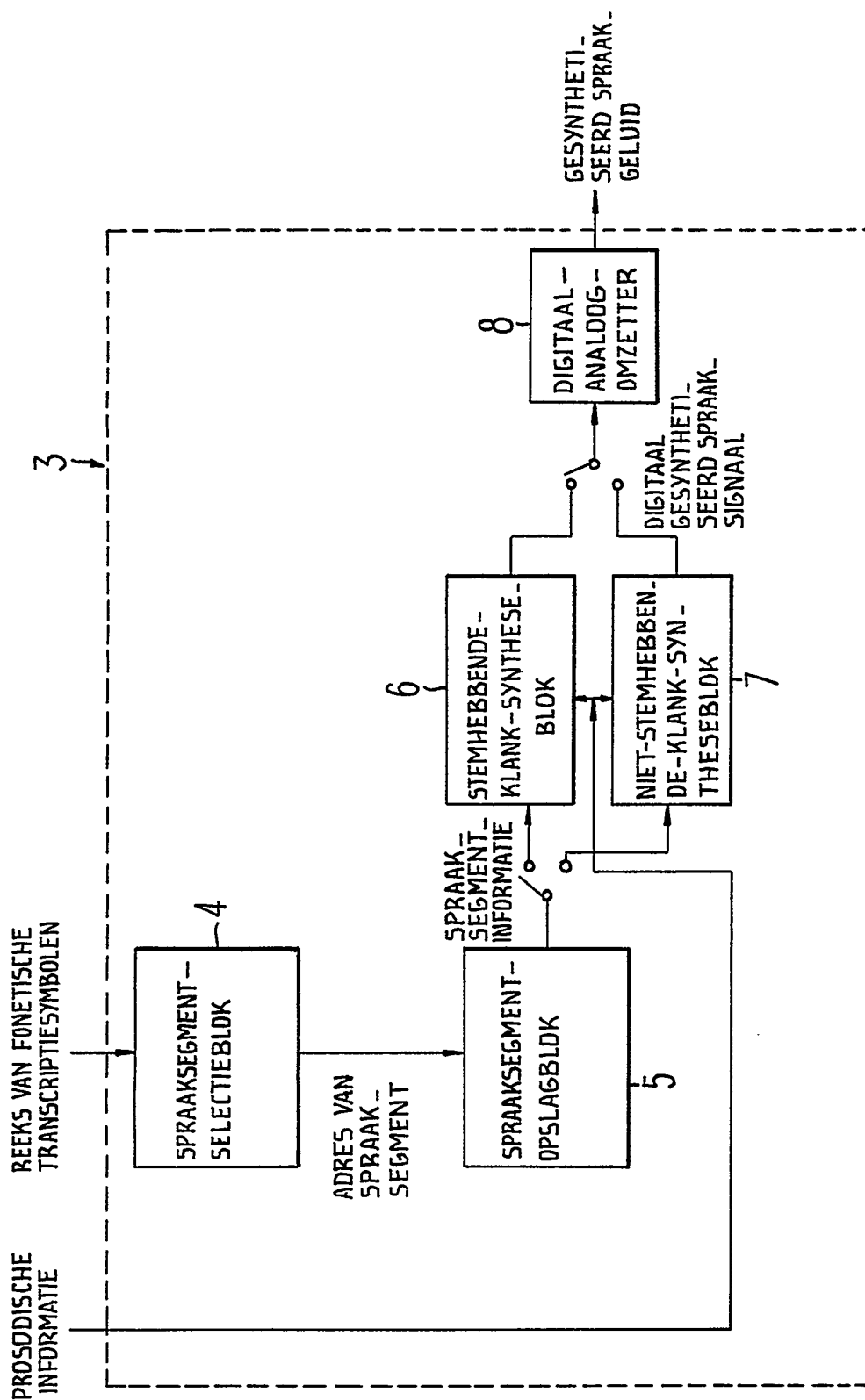
een golfvorm-samenstellings-deelblok (13) dat een stemhebbend spra-
15 aksignaal produceert uit invoer van de golfinformatie van het decode-
ringsdeelblok (9), de tijdvervormingsinformatie uit het duurregelings-
deelblok (10), de toonhoogte-pulspositie-informatie van het toonhoogte-
regelings-deelblok (11) en de versterkingsinformatie van het energie-
regelings-deelblok (12).

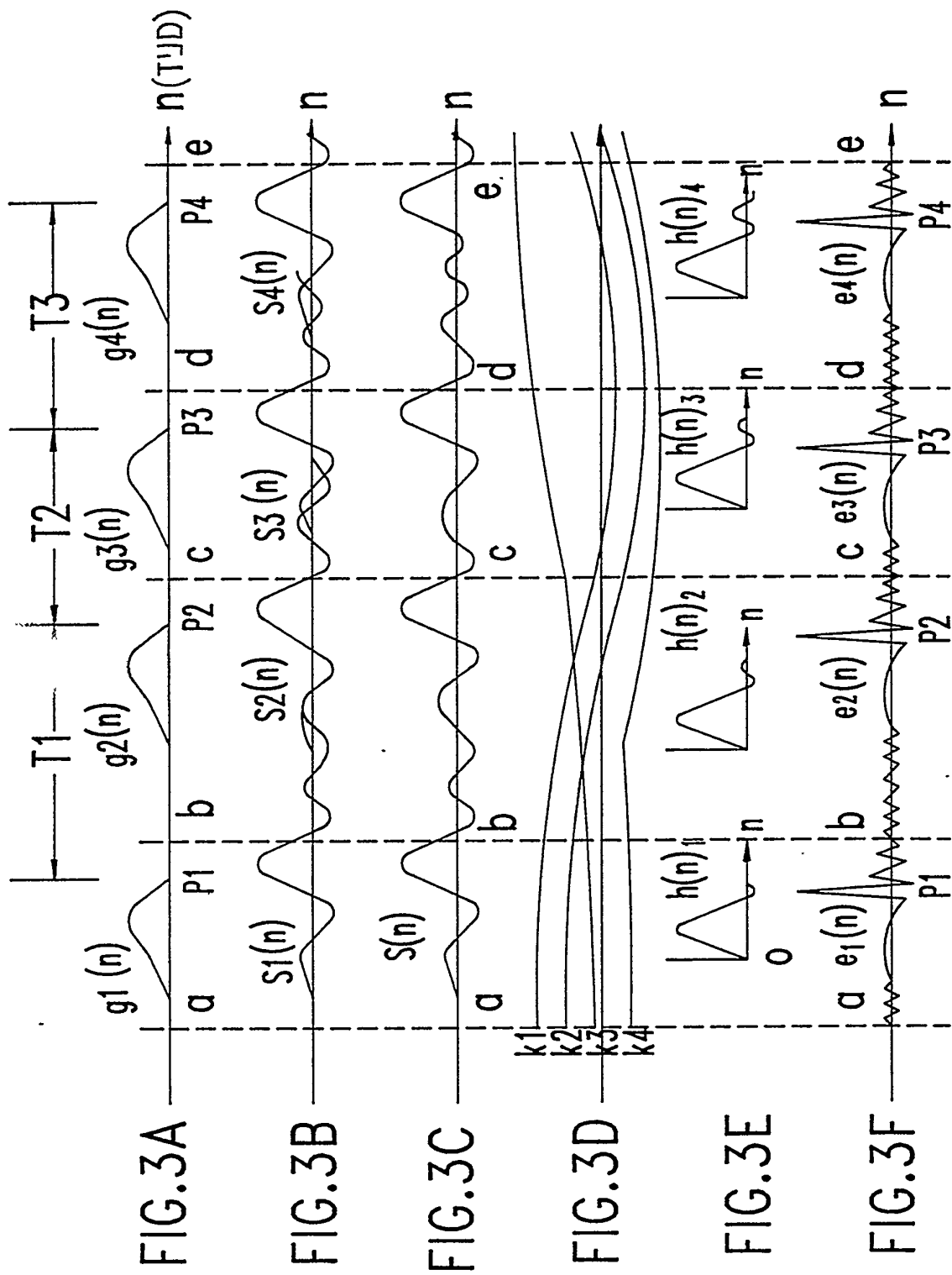
20

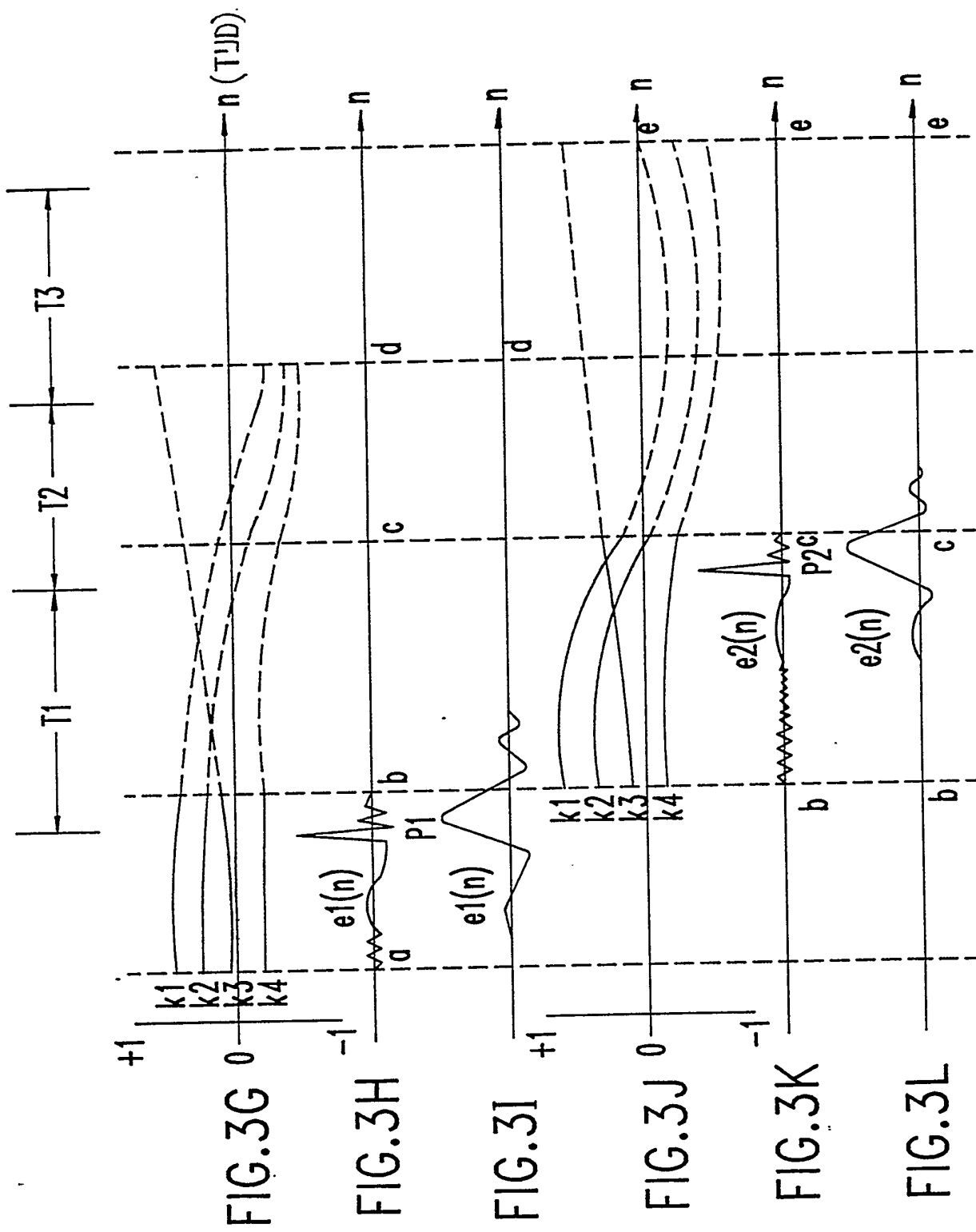


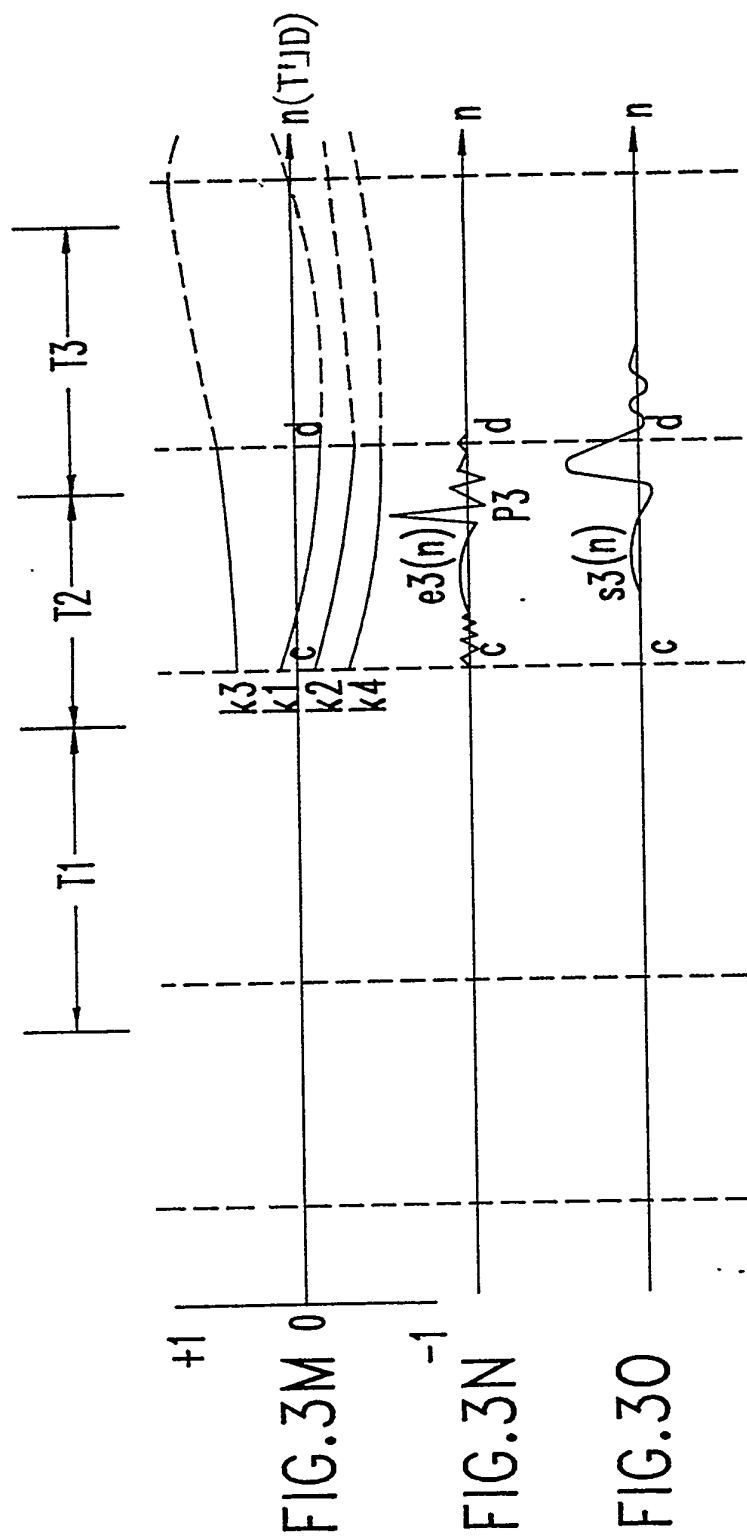
9201941

FIG.2









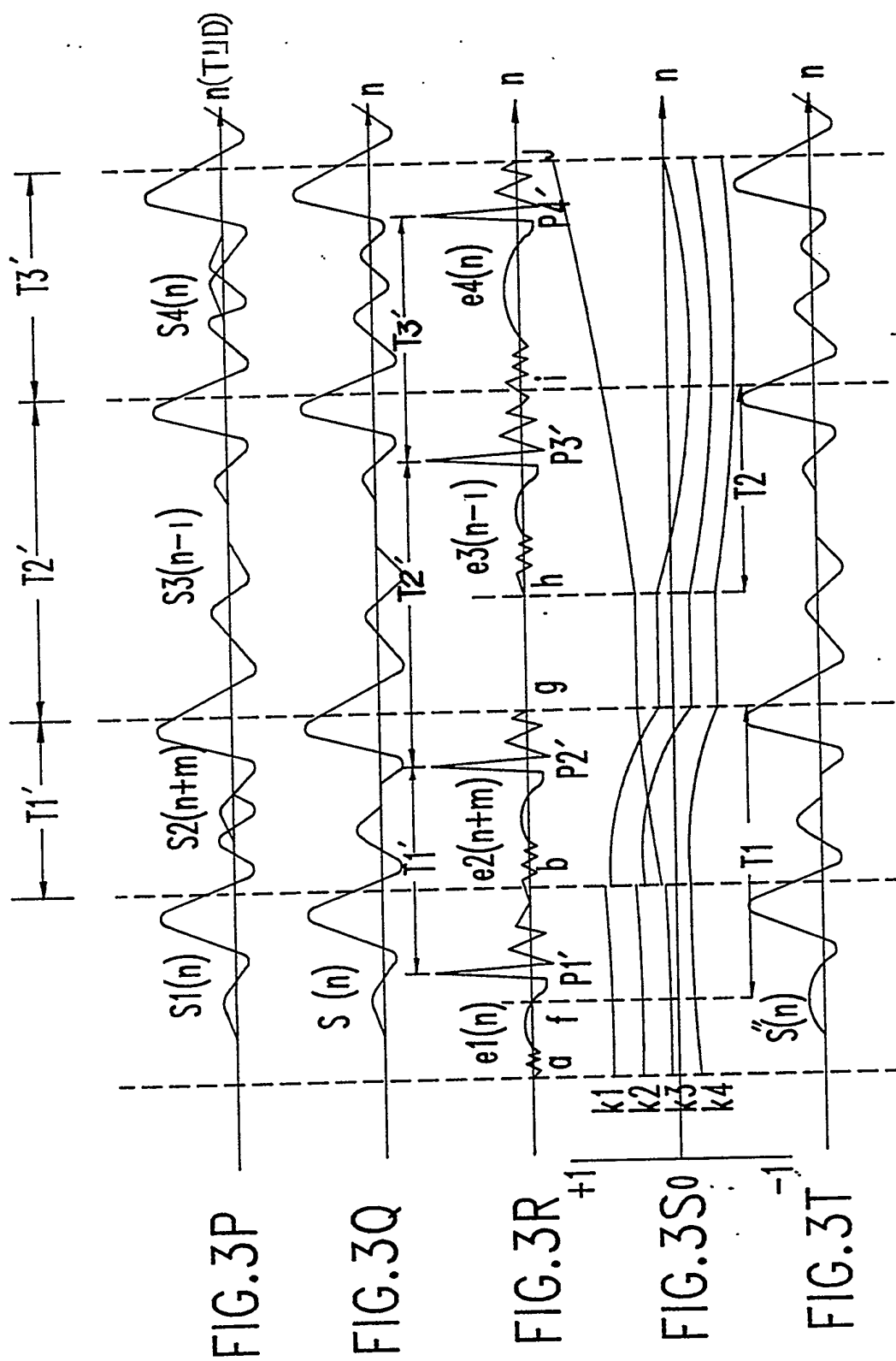


FIG.4

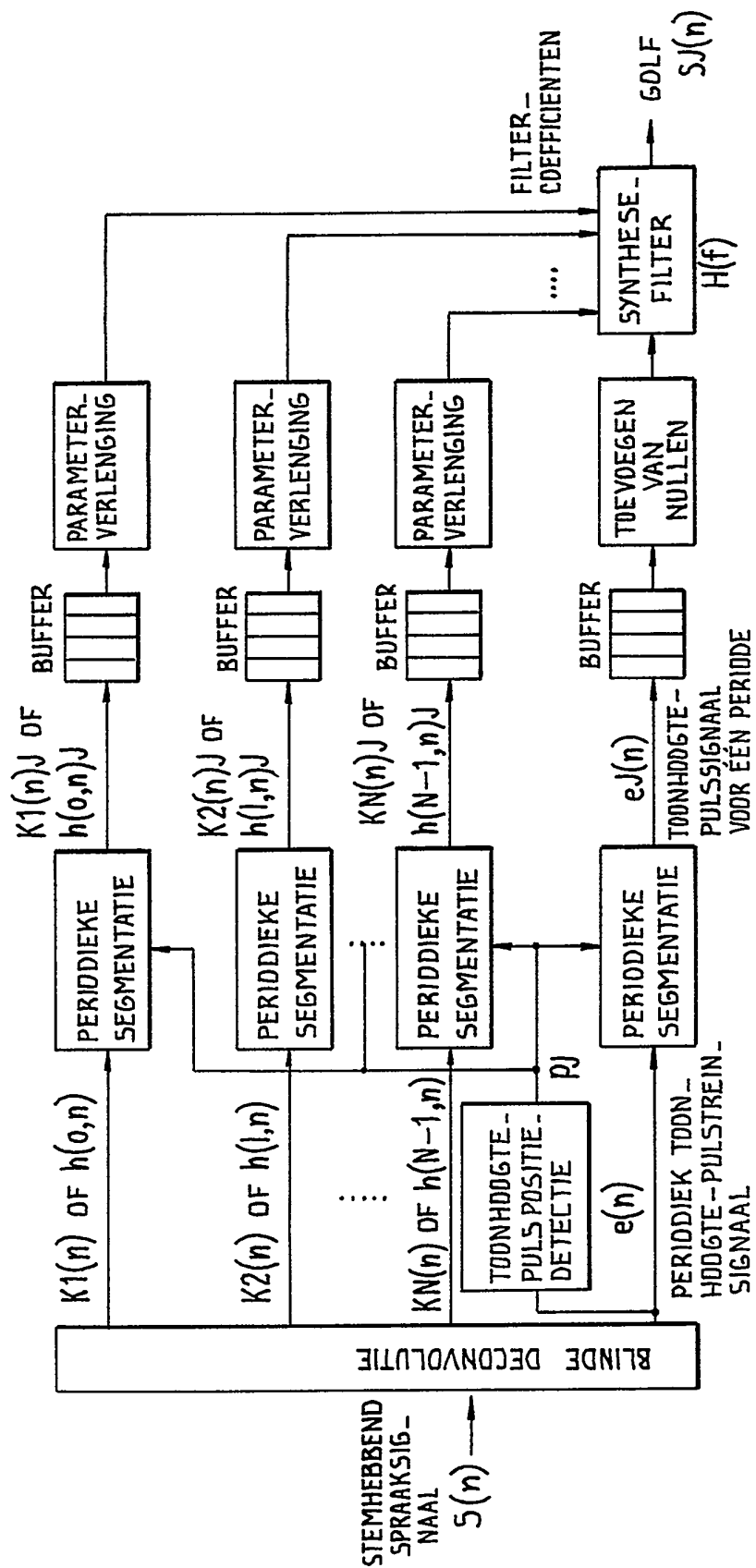
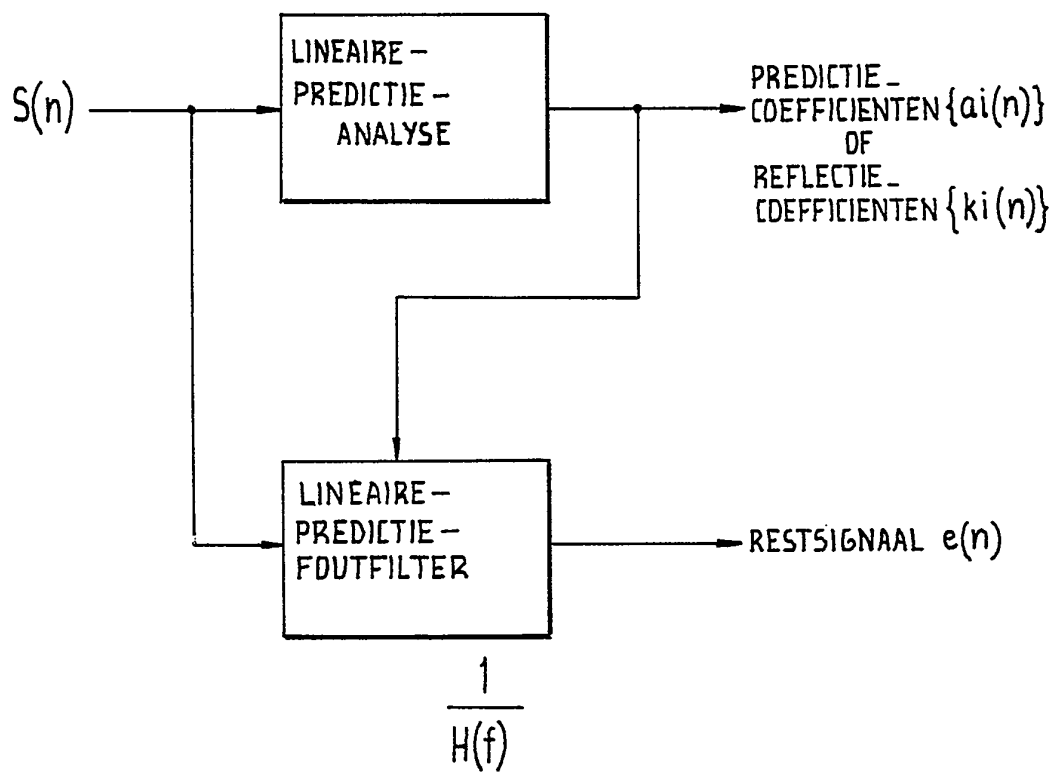


FIG.5A



9201941

FIG. 5B

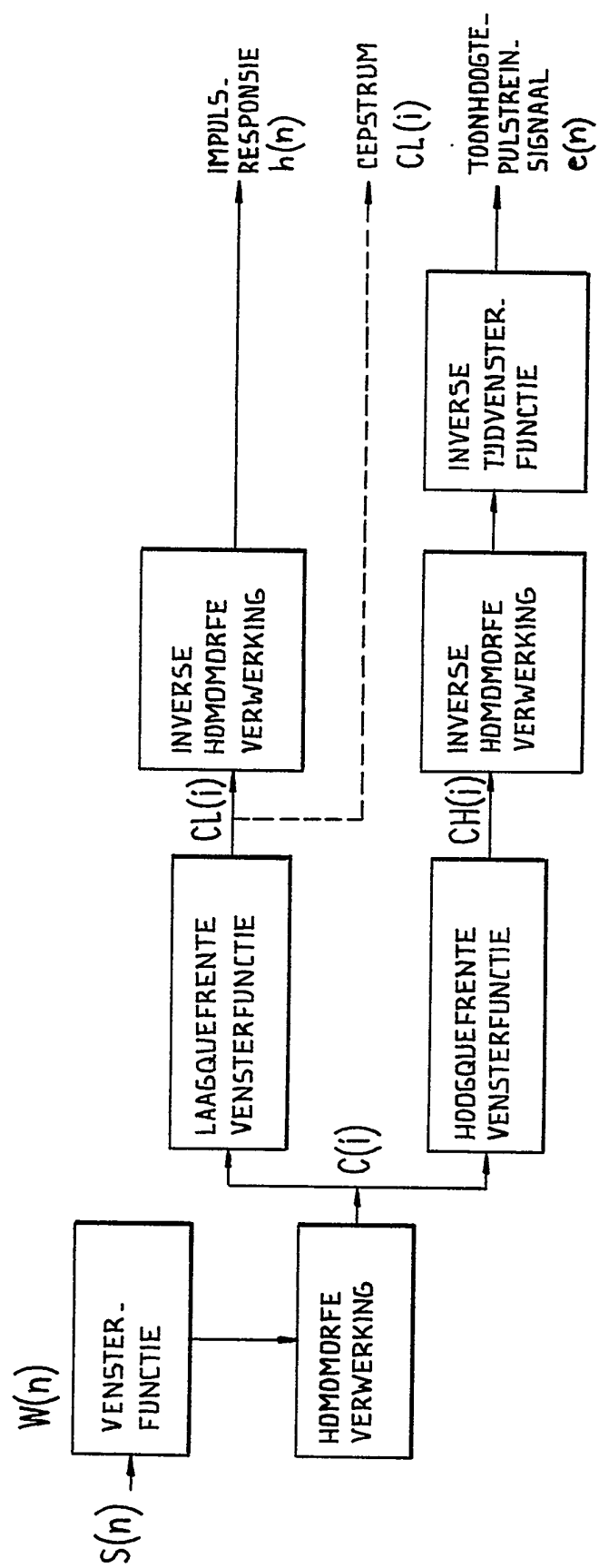


FIG. 5C

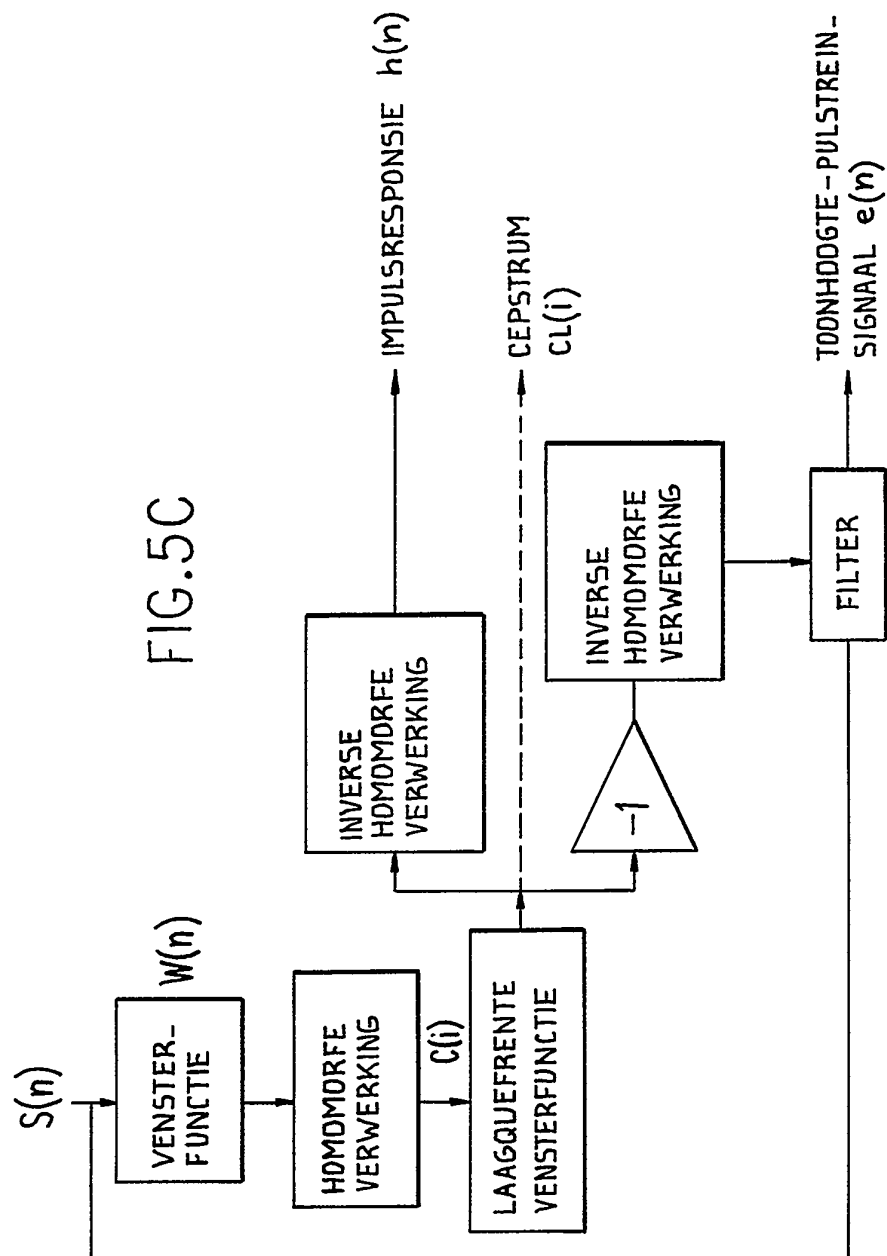


FIG.5D

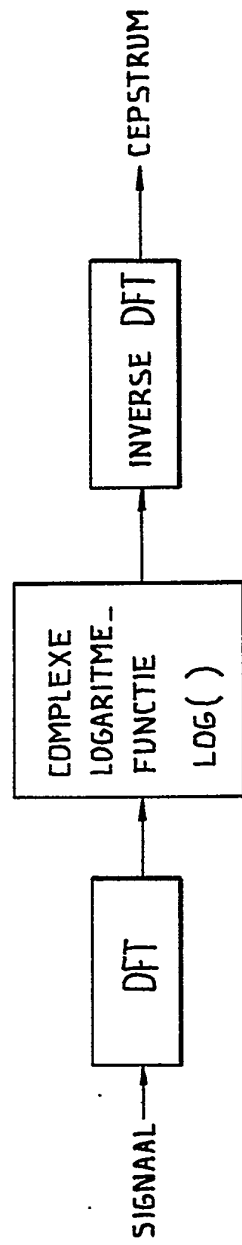


FIG.5E

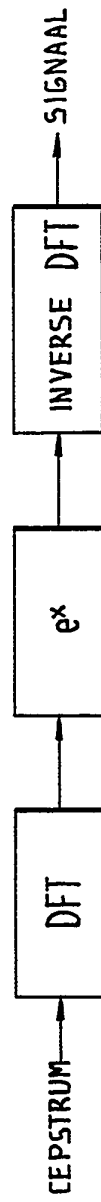


FIG.6A

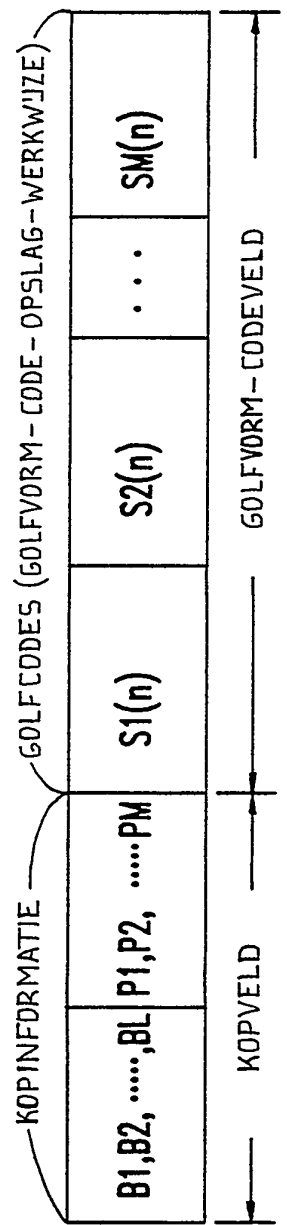


FIG.6B

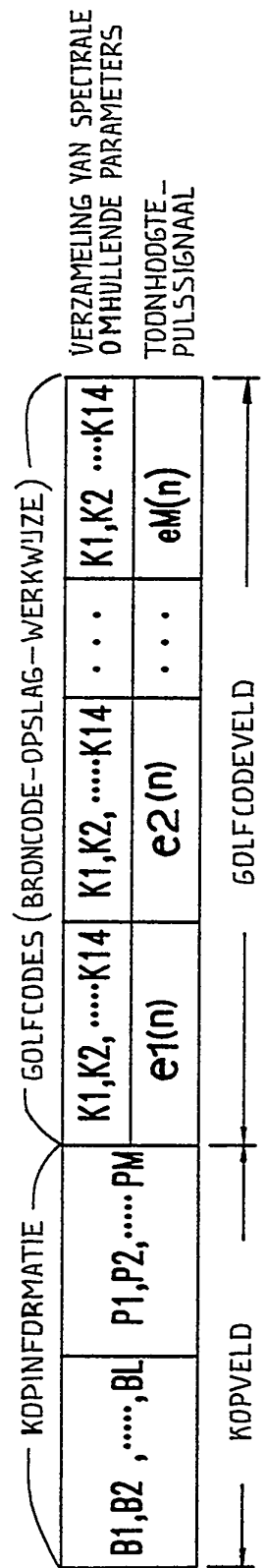
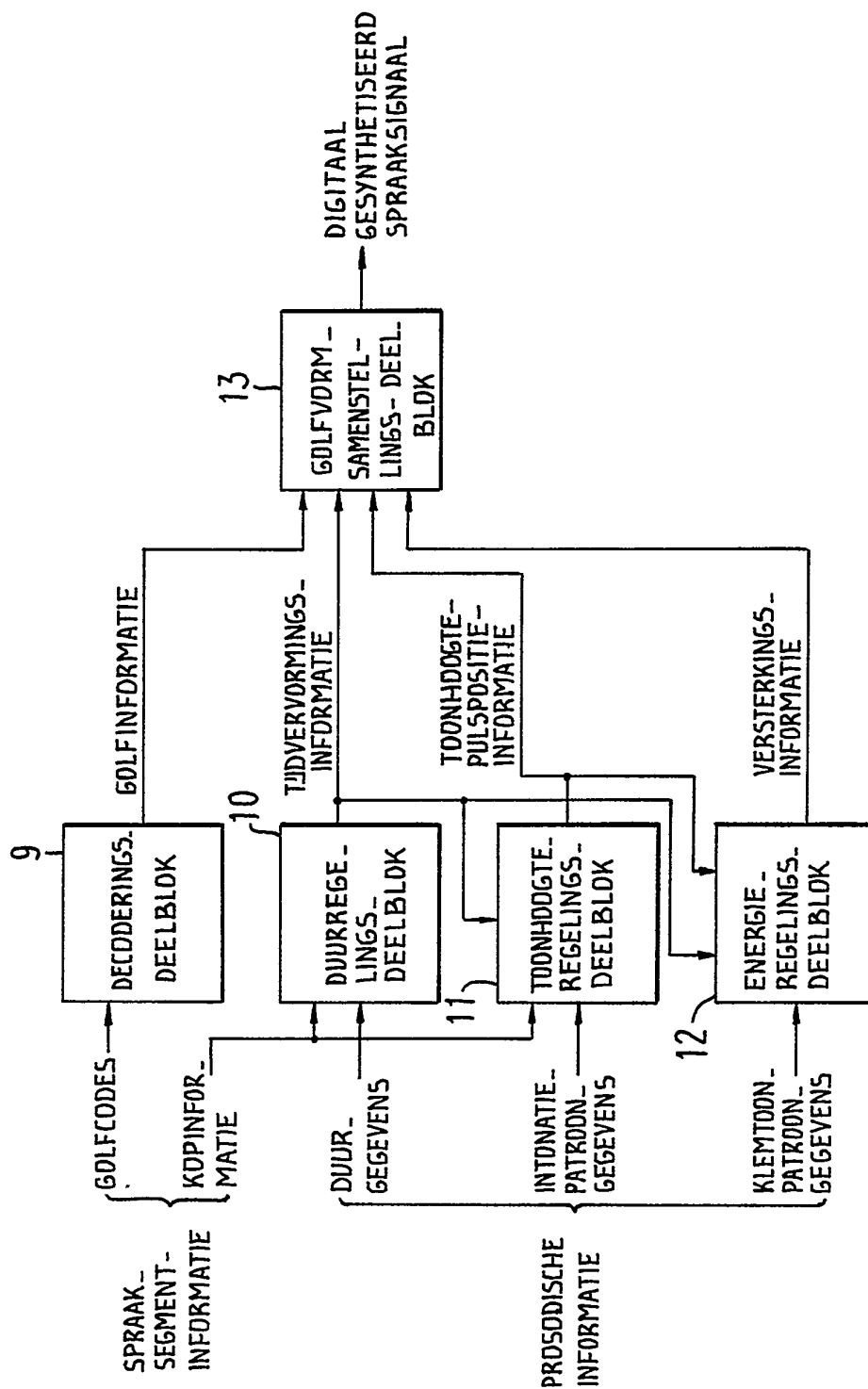


FIG.7



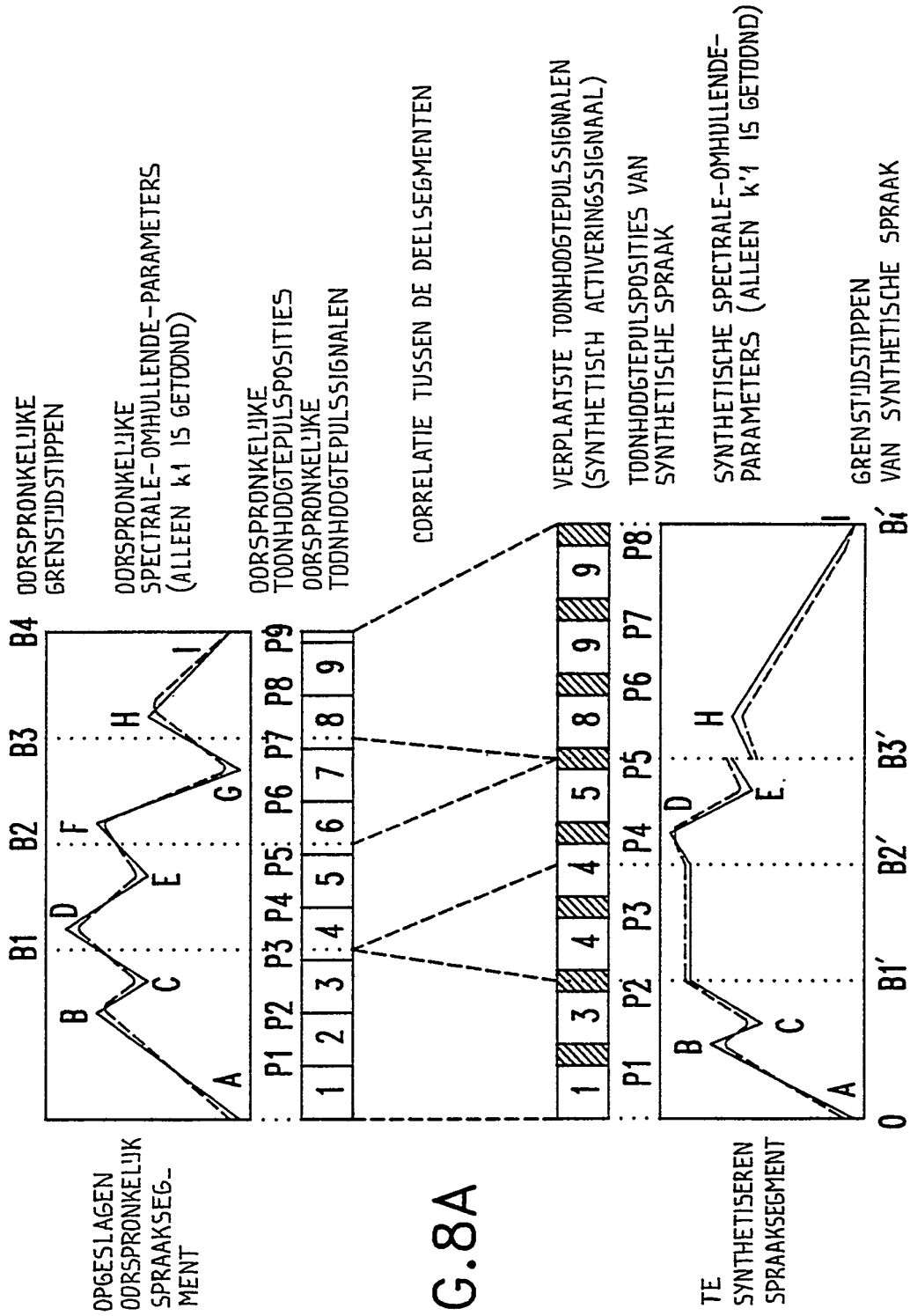


FIG.8A

FIG.8B

