NORGE

(12) **PATENT**

(19) NO (11) **306965** (13) B1

(51) Int Cl[6] G 10 L 9/10

Patentstyret

(21) Søknadsnr	19924782	(86) Int. inng. dag og søknadsnummer	29.04.1992, PCT/FI92/00128
(22) Inng. dag	10.12.1992	(85) Videreføringsdag	10.12.1992
(24) Løpedag	29.04.1992	(30) Prioritet	1991.04.30, FI, 912088
(41) Alm. tilgj.	26.02.1993		
(45) Meddelt dato	17.01.2000		

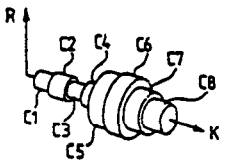
(71) Patenthaver Nokia Telecommunications OY, Mäkkylän puistotie 1, SF-02600 Esbo, FI
(72) Oppfinner Marko Vänskä, Nummela, FI
(74) Fullmektig Bryn & Aarflot AS, 0104 Oslo

(54) Benevnelse **Fremgangsmåte for gjenkjennelse av en taler**

(56) Anførte publikasjoner Ingen

(57) Sammendrag

Oppfinnelsen vedrører en fremgangsmåte for gjenkjennelse av en taler, som omfatter sammenligning av en modell beregnet på grunnlag av sampler utledet fra et talesignal med en lagret modell av minst en kjent taler. Ved fremgangsmåten ifølge oppfinnelsen blir gjennomsnittene av tverrsnittsarealene eller andre tverrsnittsdimensjoner av deler C1 - C8 av en tapsfri rørmodell av talerens taletrakt som er beregnet ved hjelp av samplene utledet fra talesignalet, sammenlignet med de tilsvarende gjennomsnitt av delene av den lagrede taletraktmodellen til minst en kjent taler.



Foreliggende oppfinnelse vedrører en fremgangsmåte for gjenkjennelse av en taler der en taler blir gjenkjent ved å sammenligne en modell beregnet på grunnlag av sampler utledet fra talerens talesignal med en lagret modell for minst en kjent taler.

En kjent måte til gjenkjennelse og verifisering av en bruker i forskjellige systemer, slik som datamaskinsystemer eller telefonsystemer, er å benytte talesignalet. Alle slike systemer for talegjenkjennelse forsøker å finne taleparametere for automatisk gjenkjennelse av taleren og skjelne forskjellige talere fra hverandre. En sampel utledet fra talen til hver taler blir brukt til å skape modeller som inneholder visse parametere som er karakteristiske for stemmen, og modellene blir så lagret i minnet til taler-gjenkjennelsessystemet. For å gjenkjenne en anonym taler blir dennes talesignal samplet, og en modell som omfatter de samme parametere blir skapt fra talesamplene og sammenlignet med referansemodellene eller mønsteret lagret i systemets minne. Hvis modellen som er skapt fra det talesignalet som skal identifiseres, passer med ett av mønstrene til kjente talere som er lagret i minnet med tilstrekkelig nøyaktighet når det benyttes et forutbestemt kriterium, blir den anonyme taleren gjenkjent som den person fra hvis talesignal det mønsteret som passer, hadde blitt skapt. Generelt sagt er dette hovedprinsippet ved alle systemer for gjenkjennelse av talere, mens parameterene og løsningene som brukes for å modellere talerens stemme, avviker sterkt fra hverandre. Eksempler på kjente fremgangsmåter og systemer for gjenkjennelse av talere er beskrevet i US patent 4 720 863 og 4 837 830, britisk patentsøknad 2 169 120 og EP-patentsøknad 369 485.

Formålet med oppfinnelsen er en fremgangsmåte for talegjenkjennelse av en ny type, hvor taleren blir gjenkjent på grunnlag av et vilkårlig talesignal med større nøyaktighet og med en enklere algoritme enn tidligere.

Dette blir oppnådd ved hjelp av en fremgangsmåte av den type som er beskrevet i innledningen, hvor ifølge oppfinnelsen gjennomsnittene av tverrsnittsarealene eller andre tverr-

snittsdimensjoner til deler av en tapsfri rørmodell av en talers taletrakt (taleorganer), bestemt ved hjelp av samplene som er utledet fra talesignalet, blir sammenlignet i sammenligningstrinnet med de tilsvarende gjennomsnitt av delene av den lagrede taletraktmodell for minst en kjent taler.

Den grunnleggende ide ved oppfinnelsen er å gjenkjenne taleren på grunnlag av den taletrakten som er spesifikk for hver taler. Taletrakten slik den omtales i denne forbindelse, refererer til den passasjen som omfatter de menneskelige stemmebånd, strupehodet, svelg, munn og lepper, ved hjelp av hvilke den menneskelige tale blir frembrakt. Profilen til talerens taletrakt varierer kontinuerlig med tiden og den nøyaktige form av taletrakten er vanskelig å bestemme bare på grunnlag av informasjon utledet fra talesignalet som er utsatt for de kompliserte vekselvirkninger mellom de forskjellige deler av taletrakten og den forskjellige sammensetning av veggmaterialet i taletrakten for forskjellige personer. Ifølge fremgangsmåten i henhold til oppfinnelsen er det imidlertid ikke nødvendig med en nøyaktig form på taletrakten. Ifølge oppfinnelsen blir talerens taletrakt modellert ved hjelp av en tapsfri rørmodell, idet formen på røret er spesifikk for hver taler. Selv om profilen på talerens taletrakt og dermed den tapsfrie rørmodellen varierer kontinuerlig under tale, er i tillegg ytterdimensjonene og gjennomsnittsverdien av taletrakten og den tapsfrie rørmodellen konstante verdier som er spesifikke for hver taler. Derfor kan taleren gjenkjennes ved hjelp av fremgangsmåten ifølge oppfinnelsen med rimelig nøyaktighet på grunnlag av den gjennomsnittlige formen eller formene til den tapsfrie rørmodellen av talerens taletrakt. I en utførelsesform av oppfinnelsen blir gjenkjennelsen utført på grunnlag av de gjennomsnittlige tverrsnittsarealer av sylinderdelene til den tapsfrie rørmodellen og ekstremverdiene eller ytterverdiene, d.v.s. maksimums- og minimumsverdiene av tverrsnittsarealet til sylinderdelene.

I en annen utførelsesform av oppfinnelsen blir nøyaktigheten av talergjenkjennelsen ytterligere forbedret ved å

definere et gjennomsnitt av den tapsfrie rørmodellen for individuelle lyder. Under en viss lyd forblir formen på taletrakten nesten uendret og representerer bedre talerens taletrakt. Når flere lyder blir brukt ved gjenkjennelsen, blir en meget høy nøyaktighet oppnådd.

Tverrsnittsarealene av sylinderdelene til den tapsfrie rørmodellen kan lett bestemmes ved hjelp av såkalte refleksjons-koeffisienter frembrakt ved konvensjonelle talekodingsalgoritmer og systemer. Det er selvsagt også mulig å bruke andre tverrsnitts-dimensjoner av det tapsfrie røret som referanseparameter, slik som radien eller diameteren. På den annen side er ikke tverrsnittet av røret nødvendigvis sirkulært.

I det følgende vil oppfinnelsen bli beskrevet mer detaljert ved hjelp av en illustrerende utførelsesform under henvisning til de vedføyde tegninger, hvor:

Fig. 1 og 2 illustrerer modelleringen av en talers taletrakt ved hjelp av et tapsfritt rør som omfatter påfølgende sylinderdeler;

Fig. 3 viser et flytskjema som illustrerer en fremgangsmåte til talergjenkjennelse ifølge oppfinnelsen;

Fig. 4 illustrerer endringer i modellene med det tapsfrie røret under tale; og

Fig. 5 viser et blokkskjema som illustrerer gjenkjennelsen av en taler på ett lydnivå.

Figur 1 viser en perspektivskisse av en tapsfri rørmodell som omfatter påfølgende sylinderdeler C1-C8 og som hovedsakelig representerer en menneskelig taletrakt. På figur 2 er den tapsfrie rørmodellen på figur 1 vist fra siden. Den menneskelige taletrakt refererer vanligvis til en passasje som utgjøres av stemmebåndene, strupen, svelget og leppene, ved hjelp av hvilke mennesker frembringer talelyder. På figur 1 og 2 representerer sylinderdelen C1 formen av en stemmetraktseksjon som befinner seg etter glottis, d.v.s. åpningen mellom stemmebåndene, og sylinderdelen C8 representerer formen på stemmetrakten i området ved leppene, og sylinderdelene C2-C7 mellom disse representerer formen på de diskrete stemme-

traktseksjoner mellom glottis og leppene. Formen på taletrakten varierer typisk kontinuerlig under tale når forskjellige lyder blir frembrakt. Likeledes varierer tverrsnittsdiametrene og arealene til de diskrete sylinderne C1-C8 som representerer forskjellige stemmetraktseksjoner, under tale. Oppfinneren har nå oppdaget at en gjennomsnittlig taletraktform beregnet fra et forholdsvis stort antall umiddelbare taletraktformer er en talerspesifikk konstant som kan brukes til gjenkjennelse av taleren. Det samme gjelder sylinderdelene C1-C8 til den tapsfrie rørmodellen av taletrakten, d.v.s. de langtidige gjennomsnittlige tverrsnittsarealer av sylinderdelene C1-C8 beregnet fra de umiddelbare verdiene av tverrsnittsarealene til sylinderdelene C1-C8 er konstanter med en forholdsvis høy nøyaktighet. Videre blir ekstremverdiene til tverrsnittsdimensjonene for sylinderne i den tapsfrie rørmodellen bestemt av ekstremdimensjonene til den aktuelle taletrakten og er således talerspesifikke, forholdsvis nøyaktige konstanter.

Fremgangsmåten ifølge oppfinnelsen benytter såkalte refleksjonskoeffisienter, d.v.s. PARCOR-koeffisienter r_k oppnådd som et foreløpig resultat i den lineære prediktive koding (LPC) som er velkjent på området. Disse koeffisientene har en viss forbindelse med formen og strukturen til taletrakten. Forbindelsen mellom refleksjonskoeffisientene r_k og arealene A_k til sylinderdelene C_k i den tapsfrie rørmodellen av talertrakten, er beskrevet ved hjelp av ligning (1)

$$- r(k) = \frac{A(k+1) - A(k)}{A(k+1) + A(k)} \qquad (1)$$

hvor k = 1, 2, 3, ...

Den LPC-analysen som frembringer refleksjonskoeffisientene som brukes i oppfinnelsen, blir benyttet i mange kjente talekodingsmetoder. En fordelaktig anvendelse av fremgangsmåten ifølge oppfinnelsen er ventet å bli funnet ved gjenkjennelse av abonnenter i mobilradiosystemer, spesielt i det Pan-europeiske digitale mobilradiosystemet GSM. GSM-spesifikasjonen 06.10 definerer meget nøyaktig den regulære

pulseksitasjon - langtids prediksjon (RPE-LTP) talekodingsmetoden som benyttes i systemet. Bruken av fremgangsmåten ifølge oppfinnelsen i forbindelse med denne talekodingsmetoden er fordelaktig ettersom de refleksjonskoeffisientene som er nødvendig ifølge oppfinnelsen, blir oppnådd som et foreløpig resultat i den ovennevnte RPE-LPC-kodingsmetoden. I den foretrukne utførelsesform av oppfinnelsen følger alle fremgangsmåtetrinn opp til beregningen av refleksjonskoeffisientene og anordningene som utfører trinnene den talekodingsalgoritmen som stemmer overens med GSM-spesifikasjon 06.10, som herved inntas som referanse. I det følgende vil disse fremgangsmåtetrinnene bli beskrevet bare generelt så langt det er nødvendig for å forstå oppfinnelsen under henvisning til flytskjemaet på figur 3.

På figur 3 blir et inngangssignal IN samplet i blokk 10 ved en samplingsfrekvens på 8 kHz, og det blir dannet en 8 bits sampelsekvens s_0. Når inngangssignalet IN er et analogt talesignal, kan blokk 10 realiseres ved hjelp av en konvensjonelle analog/digital-omformerkrets som brukes i telekommunikasjons-anordninger til å digitalisere tale. Når inngangssignalet IN er et digitalt signal, for eksempel PCM, kan blokken 10 være et digitalt inngangsbuffer synkronisert med signalet IN. I blokk 11 blir likestrømkomponenten (DC-komponenten) trukket ut fra samplene for å eliminere en forstyrrende sidetone som muligens opptrer under koding. Blokken 11 kan realiseres ved hjelp av et digitalt lavpassfilter. Deretter blir samplesignalet forforsterket i blokk 12 ved å veie høye signalfrekvenser ved hjelp av et digitalt førsteordens FIR-filter. I blokk 13 blir samplene segmentert i rammer på 160 sampler, idet varigheten av hver ramme er omkring 20 ms.

I blokk 14 blir talesignalets spektrum modellert ved å utføre en LPC-analyse på hver ramme ved hjelp av en autokorrelasjonsmetode med en åtte ordens autokorrelasjonsfunksjon. p + 1 verdier av autokorrelajonsfunksjonen ACF blir derved beregnet fra rammen

$$ACF(k) = \sum_{i=1}^{160} s(i)s(i-k) \qquad (2)$$

hvor k = 0, 1, ..., 8.

I steden for autokorrelasjons-funksjonen er det mulig å bruke enhver annen passende funksjon, slik som en kovariansfunksjon. Verdiene av åtte refleksjonskoeffisienter r_k for et korttids analysefilter som brukes i talekoderen, blir beregnet fra de oppnådde verdier av autokorrelasjonsfunksjonen ved hjelp av Schurs rekursjon, definert i GSM-spesifikasjon 06.10, eller enhver annen rekursjonsmetode. Schurs rekursjon frembringer nye refleksjonskoeffisienter for hvert 20. ms. I den foretrukne utførelsesform av oppfinnelsen omfatter koeffisientene 16 biter og deres antall er 8. Ved å anvende Schurs rekursjon for en utvidet tidsperiode kan antallet refleksjonskoeffisienter økes om ønskelig.

Refleksjonskoeffisientene (PARCOR) kan også bestemmes på en måte og ved hjelp av et apparat som er beskrevet i US patent nr. 4 837 830 som herved inntas som referanse.

I blokk 16 blir et tverrsnittsareal A_k av hver sylinderdel C_k i den tapsfrie rørmodellen av taletrakten til taleren beregnet og lagret i en lagringsanordning på grunnlag av de refleksjonskoeffisientene r_k som er beregnet fra hver ramme. Ettersom Schurs rekursjon frembringer et nytt sett med refleksjonskoeffisienter for hver 20. ms, vil det være 50 arealverdier pr. sekund for hver sylinderdel C_k. I blokk 17 etter at sylinderarealene i den tapsfrie rørmodellen er blitt beregnet for N rammer i blokk 16, blir gjennomsnittsverdiene $A_{k,ave}$ for arealene av sylinderdelene C_k til de N tapsfrie rørmodellene som er lagret i lagringsanordningen, beregnet, og det maksimale tverrsnittsarealet $A_{k,max}$ som inntraff i løpet av de nevnte N rammer for hver sylinderdel C_k, blir bestemt. I blokk 18 blir de oppnådde gjennomsnittsarealer $A_{k,ave}$ og de maksimale arealer $A_{k,max}$ av sylinderdelene i den tapsfrie rørmodellen av talerens taletrakt så sammenlignet med gjennomsnitts- og maksimums-arealene til den forut bestemte tapsfrie rørmodellen av minst en kjent taler. Hvis sammenlig-

ningen av parametrene viser at den beregnede gjennomsnitts-formen på det tapsfrie røret stemmer overens med en av de forut bestemte modellene som er lagret i lageret, så følges beslutningsblokken 19 av blokk 21 hvor det blir bekreftet at den analyserte taleren er den person som representeres av denne modellen (positiv gjenkjennelse). Hvis de beregnede parameterene ikke svarer til eller stemmer overens med de tilsvarende parametere for noen av de forut bestemte modellene som er lagret i lageret, blir beslutningsblokken 19 fulgt av blokk 20 hvor det vises at taperen er ukjent (negativ gjenkjennelse).

I et mobilradiosystem for eksempel, kan blokk 21 tillate opprettelse av en forbindelse eller bruk av en spesiell tjeneste, mens blokk 20 forhindrer disse prosedyrene. Blokk 21 kan innbefatte en generering av et signal som indikerer en positiv gjenkjennelse.

Nye modeller kan beregnes og lagres i lageret for gjenkjennelse ved hjelp av en prosedyre som hovedsakelig er maken til den som er vist i flytskjemaet på figur 3, bortsett fra at etter at gjennomsnitts- og maksimums-arealene er blitt beregnet i blokk 18, blir de beregnede arealdata lagret i lageret til gjenkjennelsessystemet som talerspesifikke filer sammen med andre nødvendige person-identifikasjonsdata, slik som navn, telefonnummer, o.s.v.

I en annen utførelsesform av oppfinnelsen blir den analysen som benyttes ved gjenkjennelsen, utført på nivået av lyder, slik at gjennomsnittet av tverrsnittsarealene av sylinderdelene i den tapsfrie rørmodellen av taletrakten blir beregnet fra talesignalet som analyseres, fra arealene av sylinderdelene til de umiddelbare tapsfrie rørmodeller som skapes under en forut bestemt lyd. Varigheten av en lyd er ganske lang slik at flere, selv titalls påfølgende modeller kan beregnes fra en enkelt lyd som er til stede i tale-signalet. Dette er illustrert på figur 4 som viser fire påfølgende umiddelbare tapsfrie rørmodeller S1 - S4. Som det fremgår klart av figur 4 varierer radiene (og tverrsnitts-arealene) til de enkelte sylindere i det tapsfrie røret med

tiden. For eksempel kan de umiddelbare modellene S1, S2 og S3 grovt klassifiseres som å ha blitt frembrakt under den samme lyden, slik at deres gjennomsnitt kan beregnes. Istedet er modellen S4 forskjellig og tilordnet en annen lyd, og derfor blir den ikke tatt i betraktning under dannelsen av gjennomsnittet.

I det følgende vil gjenkjennelse ved hjelp av lydnivået bli beskrevet under henvisning til blokkskjemaet på figur 5. Selv om gjenkjennelsen kan utføres ved hjelp av en enkelt lyd, blir det foretrukket å bruke minst to forskjellige lyder ved gjenkjennelsen, for eksempel en vokal og/eller en konsonant. De forut bestemte tapsfrie rørmodellene av en kjent taler som svarer til disse lydene, utgjør en såkalt kombinasjonstabell 58 som er lagret i lageret. En enkel kombinasjonsmodell kan for eksempel omfatte gjennomsnittsarealene av sylinderne i de tapsfrie rørmodellene som er beregnet for tre lyder "a", "e" og "i", d.v.s. tre forskjellige gjennomsnittlige tapsfrie rørmodeller. Denne kombinasjonstabellen blir lagret i lageret i den ovenfor nevnte talerspesifikke filen. Når den umiddelbare eller øyeblikkelige tapsfrie rørmodellen som er beregnet (blokk 51) fra den samplede talen til en ukjent taler som skal gjenkjennes ved hjelp av refleksjonskoeffisientene, blir detektert (kvantisert) til grovt å svare til en av de forut bestemte modellene (blokk 52), blir den lagret i lageret (blokk 53) for påfølgende gjennomsnittsdannelse. Når et tilstrekkelig antall N med umiddelbare tapsfrie rørmodeller er blitt oppnådd for hver lyd, blir gjennomsnittene A_{ij} av tverrsnittsarealene for sylinderdelene i den tapsfrie rørmodellen beregnet separat for hver lyd (blokk 55). Gjennomsnittene A_{ij} blir så sammenlignet med tverrsnittsarealene A_{ij} for sylinderne i de tilsvarende modeller som er lagret i kombinasjonstabellen 58. Hver modell i kombinasjonstabellen 58 har sin egen referansefunksjon 56 og 57, slik som en krysskorrelasjonsfunksjon ved hjelp av hvilken overensstemmelsen eller korrelasjonen mellom den modellen som er beregnet fra talen og den lagrede modellen av kombinasjonstabellen 58, blir evaluert. Den ukjente taleren blir gjenkjent hvis der er

en tilstrekkelig nøyaktig korrelasjon mellom modellene som er beregnet for alle lyder eller et tilstrekkelig antall med lyder og de lagrede modellene.

Den umiddelbare eller øyeblikkelige tapsfrie rørmodellen 59 som er frembrakt fra det samplede talesignalet, kan detekteres i blokk 52 til å tilsvare en viss lyd hvis tverrsnittsdimensjonen av hver sylinderdel i den umiddelbare tapsfrie rørmodellen 59 er innenfor de forut bestemte, lagrede grenseverdier for den tilsvarende lyd fra den kjente taleren. Disse lydspesifikke og sylinderspesifikke grenseverdiene blir lagret i en såkalt kvantiseringstabell 54 i lageret. På figur 5 viser henvisningstallene 60 og 61 hvordan de lyd- og sylinder-spesifikke grenseverdier skaper en maske eller modell for hver lyd, innenfor det tillatte området 60A og 61A (uskraverte områder) av hvilke den umiddelbare taletrakt-modellen 59 som skal identifiseres, må falle. På figur 5 kan den umiddelbare taletraktmodellen 59 passes inn innefor lydmasken 60, mens det er klart at den ikke kan innpasses innenfor lydmasken 61. Blokken 52 virker derfor som et slags lydfilter som klassifiserer taletraktmodellene i de respektive talegrupper A, E, I, o.s.v.

Fremgangsmåten kan for eksempel realiseres ved hjelp av programvare i et konvensjonelt signalbehandlingssystem som har minst ett ROM- eller EEPROM-lager for lagring av programvaren, et arbeidslager av RAM-typen og en tilpasningskrets for inngangssignalet IN eller taledigitalisereren 10. De forut-bestemte modellene og andre forutbestemte gjenkjennelses-parametere kan også være lagret i et ikke-flyktig lager, slik som et EEPROM. Fremgangsmåten kan også realiseres som programvare i det apparatet som er beskrevet i US patent nr. 4 837 830.

Figurene og beskrivelsen vedrørende disse er bare ment å illustrere foreliggende oppfinnelse. Når det gjelder detaljene kan fremgangsmåten ifølge oppfinnelsen variere innenfor rammen av de vedføyde krav.

P A T E N T K R A V

1. Fremgangsmåte for gjenkjennelse av en taler der en taler blir gjenkjent ved å sammenligne en modell som er beregnet på grunnlag av sampler utledet fra talerens talesignal med en lagret modell av minst en kjent taler,
k a r a k t e r i s e r t v e d at gjennomsnittene av tverrsnittsarealene eller andre tverrsnittsdimensjoner av deler av en tapsfri rørmodell av talerens taletrakt, bestemt ved hjelp av samplene som er utledet fra talesignalet, blir sammenlignet i sammenligningstrinnet med de tilsvarende gjennomsnitt av delene av den lagrede taletraktmodellen til minst en kjent taler.

2. Fremgangsmåte ifølge krav 1, omfattende

a) gruppering av de samplene som er utledet fra talesignalet i rammer som inneholder M sampler;

b) beregning av verdiene av en forut bestemt auto-korrelasjonsfunksjon eller en tilsvarende funksjon fra samplene i rammene;

c) beregning av refleksjonskoeffisienter rekursivt fra verdiene av autokorrelasjonsfunksjonen eller lignende;

d) sammenligning av parametere som er beregnet ved hjelp av refleksjonskoeffisientene med tilsvarende lagrede parametere for minst en kjent taler,
k a r a k t e r i s e r t v e d at trinn d) omfatter

beregning av arealet av hver sylinderdel i den tapsfrie rørmodellen av taletrakten ved hjelp av sylinderdeler på grunnlag av rammens refleksjonskoeffisienter;

gjentakelse av beregningen av arealene i N rammer og beregning av gjennomsnittet av de oppnådde arealer separat for hver sylinderdel; og

sammenligning av de beregnede gjennomsnittsverdier med gjennomsnittsverdiene av sylinderdelene til den lagrede modellen for minst en kjent taler.

3. Fremgangsmåte ifølge krav 1 eller 2, k a r a k t e r i s e r t v e d at ekstremverdien av arealet til hver sylinderdel blir bestemt i løpet av N rammer, og at gjennomsnitts- og maksimal-arealene til sylinderdelene blir sammenlignet med gjennomsnitts- og maksimal-arealene til sylinderdelene i den lagrede taletraktmodellen for minst en kjent taler.

4. Fremgangsmåte ifølge krav 1, 2 eller 3, k a r a k t e r i s e r t v e d at de gjennomsnittlige tverrsnittsdimensjoner av sylinderdelene i den tapsfrie rørmodellen av taletrakten blir dannet ved hjelp av gjennomsnittene av tverrsnittsdimensjonene for sylinderdelene i umiddelbare tapsfrie rørmodeller som er frembrakt under en forut bestemt lyd.

5. Fremgangsmåte ifølge krav 4, k a r a k t e r i s e r t v e d at

gjennomsnittene av tverrsnittsdimensjonene for sylinderdelene til den tapsfrie rørmodellen blir beregnet separat for minst to forskjellige lyder;

gjennomsnittene for sylinderdelene til den tapsfrie rørmodellen for hver lyd blir sammenlignet med tverrsnittsdimensjonene til sylinderdelene i den lagrede tapsfrie rørmodellen av den kjente talerens respektive lyd; og

taleren blir gjenkjent hvis den beregnede tapsfrie rørmodellen for et tilstrekkelig antall lyder korrelerer nøyaktig nok med den respektive lagrede tapsfrie rørmodellen.

6. Fremgangsmåte ifølge krav 4 eller 5, k a r a k t e r i s e r t v e d at den umiddelbare tapsfrie rørmodellen som er frembrakt fra det samplede talesignalet, blir detektert å tilsvare en forut bestemt lyd hvis tverrsnittsdimensjonen til hver del av den umiddelbare tapsfrie rørmodellen faller innenfor de forut bestemte grenseverdier for den respektive lyd som er lagret i en kvantiseringstabell.

7. Fremgangsmåte ifølge krav 4, 5 eller 6, k a r a k t e r i s e r t v e d at lydene er vokaler og/eller konsonanter.

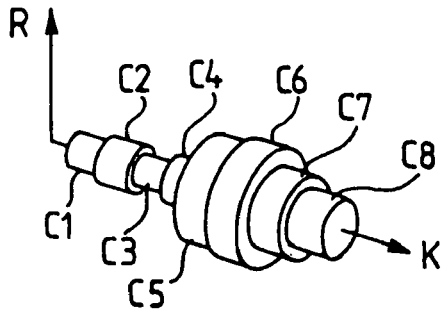


FIG. 1

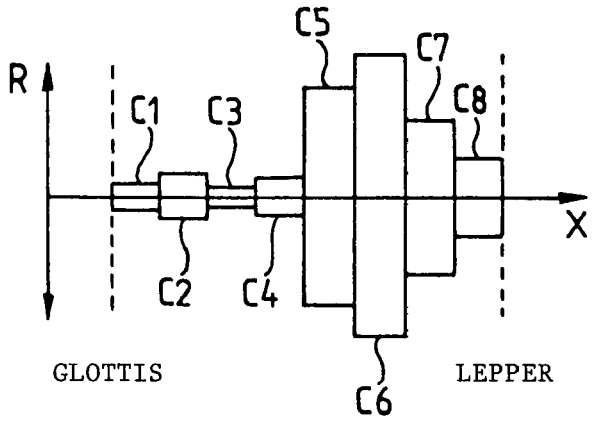


FIG. 2

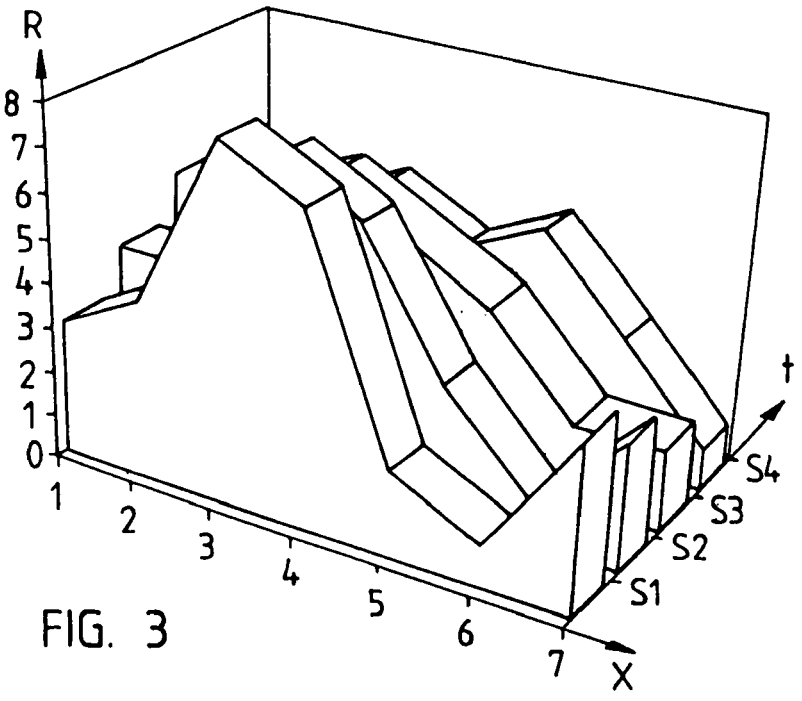


FIG. 3

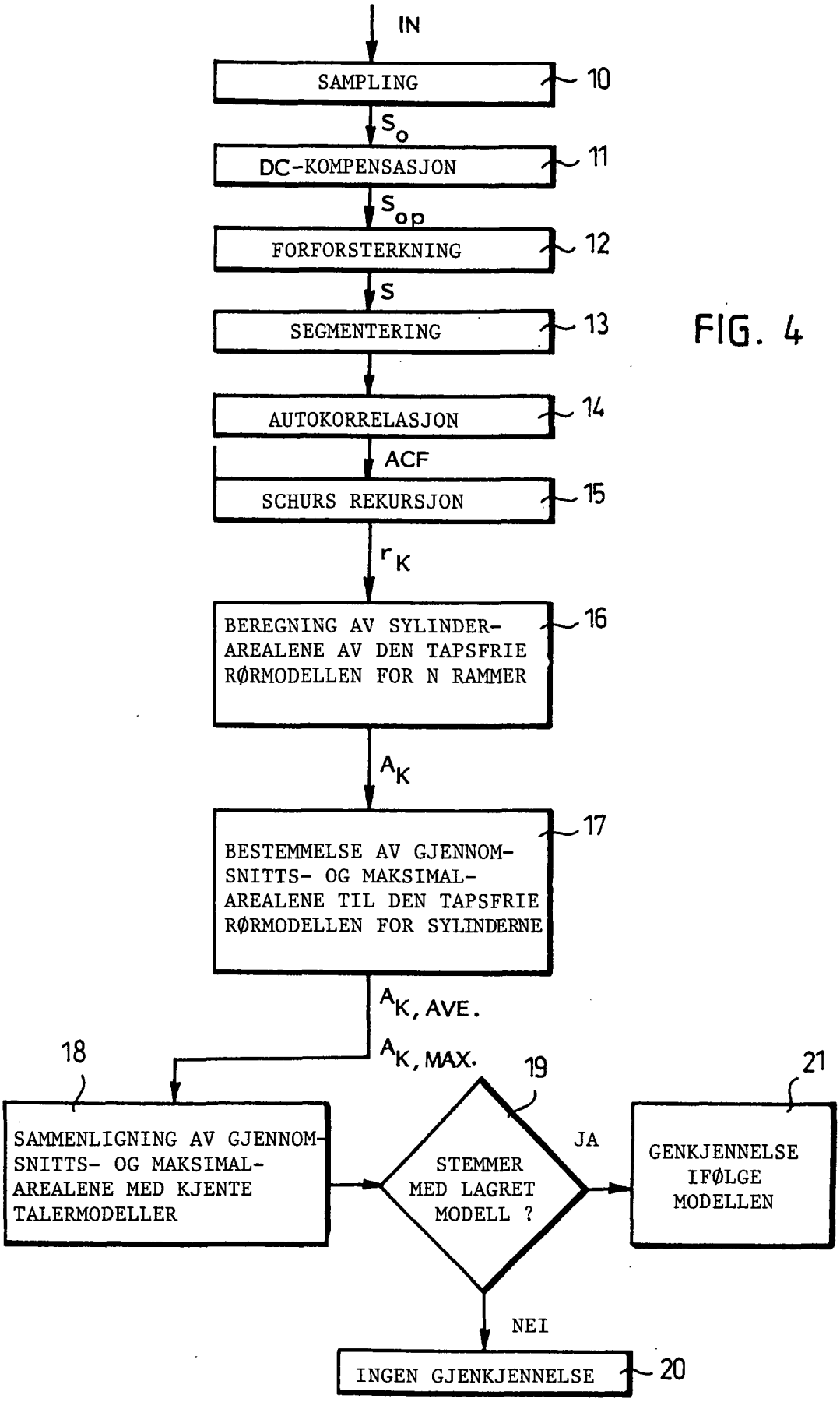

IN
SAMPLING
10
S_o
DC-KOMPENSASJON
11
S_{op}
FORFORSTERKNING
12
S
SEGMENTERING
13
AUTOKORRELASJON
14
ACF
SCHURS REKURSJON
15
r_K
BEREGNING AV SYLINDER-AREALENE AV DEN TAPSFRIE RØRMODELLEN FOR N RAMMER
16
A_K
BESTEMMELSE AV GJENNOM-SNITTS- OG MAKSIMAL-AREALENE TIL DEN TAPSFRIE RØRMODELLEN FOR SYLINDERNE
17
$A_{K,AVE.}$
$A_{K,MAX.}$
18
SAMMENLIGNING AV GJENNOM-SNITTS- OG MAKSIMAL-AREALENE MED KJENTE TALERMODELLER
19
STEMMER MED LAGRET MODELL ?
JA
21
GENKJENNELSE IFØLGE MODELLEN
NEI
INGEN GJENKJENNELSE
20

FIG. 4

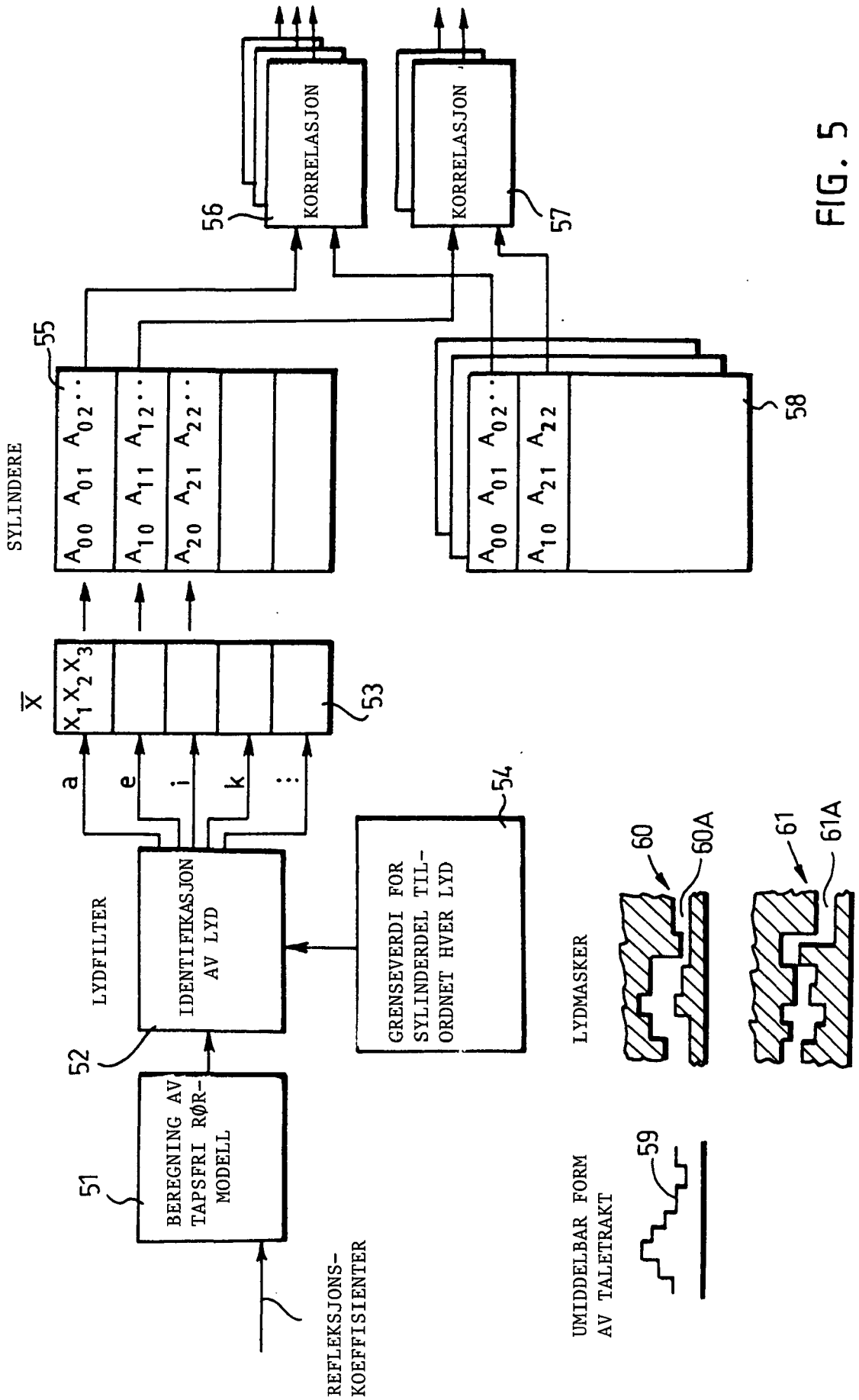

REFLEKSJONS-
KOEFFISIENTER
51
BEREGNING AV
TAPSFRI RØR-
MODELL
52
LYDFILTER
IDENTIFIKASJON
AV LYD
GRENSEVERDI FOR
SYLINDERDEL TIL-
ORDNET HVER LYD
54
a
e
i
k
:
$\bar{X}$
$X_1 X_2 X_3$
53
SYLINDERE
55
A_{00} A_{01} A_{02} ..
A_{10} A_{11} A_{12} ..
A_{20} A_{21} A_{22} ..
56
KORRELASJON
57
KORRELASJON
58
A_{00} A_{01} A_{02} ..
A_{10} A_{21} A_{22}
LYDMASKER
60
60A
61
61A
UMIDDELBAR FORM
AV TALETRAKT
59

FIG. 5