(19) SUOMI - FINLAND
(FI)

PATENTTI- JA REKISTERIHALLITUS
PATENT- OCH REGISTERSTYRELSEN
FINNISH PATENT AND REGISTRATION OFFICE

(10) **FI 973142 A7**

(12) **JULKISEKSI TULLUT PATENTTIHAKEMUS**
PATENTANSÖKAN SOM BLIVIT OFFENTLIG
PATENT APPLICATION MADE AVAILABLE TO THE PUBLIC

(21) Patenttihakemus - Patentansökan - Patent application 973142

(51) Kansainvälinen patenttiluokitus - Internationell patentklassifikation - International patent classification
G10L 21/02 (2006.01)

(22) Tekemispäivä - Ingivningsdag - Filing date **12.01.1996**

(23) Saapumispäivä - Ankomstdag - Reception date **29.07.1997**

(41) Tullut julkiseksi - Blivit offentlig - Available to the public **30.09.1997**

(43) Julkaisupäivä - Publiceringsdag - Publication date **13.06.2019**

(86) Kansainvälinen hakemus - Internationell ansökan - International application **12.01.1996 PCT/SE1996/000024**

(32) (33) (31) Etuoikeus - Prioritet - Priority

30.01.1995 SE 9500321

(71) Hakija - Sökande - Applicant

1 • Telefonaktiebolaget L M Ericsson, 126 25 Stockholm, SVERIGE, (SE)

(72) Keksijä - Uppfinnare - Inventor

1 • Händel, Peter, Sverige, SVERIGE, (SE)

(74) Asiamies - Ombud - Agent

Boco IP Oy Ab, Itämerenkatu 5, 00180 Helsinki

(54) Keksinnön nimitys - Uppfinningens benämning - Title of the invention

Spektraalisesti vähentävä kohinan vaimennusmenetelmä

Spektralt subtraherande brusdämpningsförfarande

Spektrisesti vähentävä kohinanvaimennusmenetelmä
Spektral subtraktion brusundertryckningsförfarande

Tekniikan ala

Esillä oleva keksintö koskee kohinan vaimentamista digitaalisissa kehyspohjaisissa liikennöintijärjestelmissä ja erityisesti spektrisesti vähentävää kohinanvaimennusmenetelmää tällaisissa järjestelmissä.

Keksinnön tausta

Yleinen ongelma puhesignaalin käsittelyssä on puhesignaalin korostus (engl. enhancement) kohinaympäristöstään. Eräs mahdollisuus korostaa puhetta perustuen yksittäiskanava (mikrofoni) -mittauksiin on taajuusalueen suodatus käyttäen spektrivähennystekniikoita [1], [2]. Oletetaan, että taustakohina on pitkällä aikavälillä vakio (verrattuna puheeseen), estimoidaan yleensä taustakohinalle malli aikaväleillä, joilla ei esiinny puheaktiviteettia. Tämän jälkeen, datakehyksissä, joissa on puheaktiviteettia, tätä estimoitua kohinamallia käytetään yhdessä estimoidun kohinapuhemallin kanssa puheen parantamiseksi. Spektrivähennystekniikoiden yhteydessä nämä mallit ilmoitetaan yleensä tehospektritiheyden (PSD; Power Spectral Density) avulla, joka estimoidaan käyttäen klassisia FFT-menetelmiä.

Mitkään yllä mainituista tekniikoista eivät anna perusmuodossaan antosignaalia, jolla on hyväksyttävä äänilaatu matkapuhelinsovelluksissa, eli

1. ei-vääristetty puheanto

2. kohinatason riittävä pieneneminen

3. jäljellä oleva kohina ilman häiritseviä tekotuotteita (engl. artifacts).

Erityisesti spektrivähennysmenetelmien tiedetään loukkaavan 1:tä, mikäli 2 täyttyy tai loukkaavan 2:ta, mikäli 1 täyttyy. Lisäksi tapausta 3 loukataan enemmän tai vähemmän, koska menetelmät aikaansaavat nk. musiikkikohinaa.

Yllä mainitut spektrivähennysmenetelmien haitat ovat yleisesti tunnettuja ja kirjallisuudessa esiintyy useita erityisiä modifikaatioita perusalgoritmeille koskien erityisesti puhetta-kohinassa skenaarioissa. Kuitenkin ongelma, miten muodostaa spektrivähennysmenetelmä, joka yleisillä skenaarioilla täyttää ehdot 1 - 3, on jäänyt ratkaisematta.

Puheen korostamisen kohinadatasta ongelmien tähdentämiseksi on huomattava, että spektrivähennysmenetelmät perustuvat suodatukseen, jossa käytetään tulevan datan estimoituja malleja. Mikäli nämä estimoidut mallit ovat lähellä alla olevia "todellisia" malleja, tämä on hyvin toimiva malli. Kuitenkin, johtuen puheen lyhytaikaisesta stationaarisuudesta (ajasta riippumattomuus) (10 - 40 ms), kuten myös johtuen fyysisestä todellisuudesta, joka ympäröi matkaviestinsovellusta (8000 Hz näytteenottotaajuus, 0,5 - 2,0 s kohinan stationaarisuus, jne.) on oletettavaa, että estimoidut mallit olennaisesti poikkeavat alla olevasta todellisuudesta, ja täten aikaansaavat suodatetun annon, jolla on matala äänilaatu.

Julkaisussa EP-A1-0.588.526 esitetään menetelmä, jossa spektrianalyysi suoritetaan joko käyttäen nopeata Fourier-muunnosta (FFT; Fast Fourier Transformation) tai lineaarista ennustavaa koodausta (LPC; Linear Predictive Coding).

Yhteenveto keksinnöstä

Esillä olevan keksinnön tarkoituksena on aikaansaada spektrivähennystä käyttävä kohinan vaimennusmenetelmä, joka antaa

paremman kohinan vähenemisen ilman, että se vaikuttaa negatiivisesti äänen laatuun.

Tämä tarkoitus ratkaistaan patenttivaatimuksen 1 tunnusomaisten piirteiden avulla.

Lyhyt kuvioiden selostus

Keksintöä, sekä sen tarkoituksia ja etuja, voidaan parhaiten ymmärtää viittaamalla seuraavaan selostukseen yhdessä oheisien piirustuksien kanssa, joissa:

kuvio 1 on lohkokaavio spektrivähennys kohinan vaimennusjärjestelmästä, joka sopii esillä olevan keksinnön menetelmän suorittamiseen;

kuvio 2 on ääniaktiviteetti-ilmaisimen (VAD; Voice Activity Detector) tilakaavio, jota voidaan käyttää kuvion 1 järjestelmässä;

kuvio 3 on kaavio, jossa esitetään kaksi eri tehospektritiheyden estimaattia puhekehyksessä;

kuvio 4 on aikakaavio näytteenotetusta äänisignaalista, joka sisältää puhetta ja taustakohinaa;

kuvio 5 on kuvion 3 signaalin aikakaavio kun spektrikohinavähennys on suoritettu tunnetun tekniikan mukaisesti;

kuvio 6 on kuvion 3 signaalin aikakaavio kun spektrikohinavähennys on suoritettu esillä olevan keksinnön mukaisesti; ja

kuvio 7 on virtauskaavio, jossa esitetään esillä olevaa keksintöä.

Edullisten suoritusmuotojen yksityiskohtainen selostus

Spektrivähennystekniikka

Ajatellaan puhekehystä, joka on huonontunut lisäkohinasta

$$x(k) = s(k) + v(k) \quad k = 1, \ldots, N \tag{1}$$

jossa $x(k)$, $s(k)$ ja $v(k)$ merkitsevät vastaavasti puheen kohina-arvoa, puhetta ja lisäkohinaa ja N tarkoittaa kehyksessä olevien näytteiden lukumäärää.

Oletetaan, että puhe on stationaarinen kehyksessä ja oletetaan, että kohina on pitkäaikaisesti stationaarinen, eli stationaarinen useiden kehysten yli. Kehysten lukumäärä, jossa $v(k)$ on stationaarinen saadaan kaavasta $\tau >> 1$. Edelleen oletetaan, että puheaktiviteetti on riittävän alhainen siten, että kohinan malli voidaan täsmälleen estimoida ei-puheaktiviteetin aikana.

Merkitään tehospektritiheydet (PSD:t) mittaukselle, puheelle ja kohinalle vastaavasti $\Phi_X(\omega)$, $\Phi_S(\omega)$ ja $\Phi_V(\omega)$, jossa

$$\Phi_x(\omega) = \Phi_s(\omega) + \Phi_v(\omega) \tag{2}$$

Kun tiedetään $\Phi_X(\omega)$ ja $\Phi_V(\omega)$, suureet $\Phi_S(\omega)$ ja $s(k)$ voidaan estimoida käyttäen vakiospektrivähennysmenetelmiä, kts. esim. [2], johon lyhyesti viitataan alla.

Annetaan $\hat{s}(k)$ merkitä $s(k)$:n estimaatin. Tällöin

$$\begin{aligned} \hat{s}(k) &= \mathcal{F}^{-1}(H(\omega)\, X(\omega)) \\ X(\omega) &= \mathcal{F}(x(k)) \end{aligned} \tag{3}$$

jossa $\mathcal{F}(\cdot)$ merkitsee jotain lineaarista muunnosta, esim. diskreettiä Fourier-muunnosta (DFT; Discrete Fourier Trans-

form) ja jossa $H(\omega)$ on reaali-arvoinen tasainen funktio, jossa $\omega \in (0,2\pi)$ ja siten, että $o \leq H\ \omega) \leq 1$. Funktio $H(\omega)$ riippuu $\Phi_X(\omega)$:stä ja $\Phi_V(\omega)$:stä. Koska $H(\omega)$ on reaaliarvoinen, $\hat{S}(\omega) = H(\omega)X(\omega)$:n vaihe on yhtä suuri kuin huonontuneen puheen vaihe. Reaaliarvoisen $H(\omega)$:n käyttö motivoidaan ihmiskorvan epäherkkyydellä vaihevääristymälle.

Yleensä $\Phi_X(\omega)$ ja $\Phi_V(\omega)$ ovat tuntemattomia ja ne on korvattava $H(\omega)$:ssä estimoiduilla suureilla $\hat{\Phi}_X(\omega)$ ja $\hat{\Phi}_V(\omega)$. Johtuen puheen ei-stationaarisuudesta $\Phi_X(\omega)$ estimoidaan yhdestä ainoasta datakehyksestä kun toisaalta $\Phi_V(\omega)$ estimoidaan käyttäen dataa τ puhevapaassa kehyksessä. Yksinkertaisuuden vuoksi oletetaan, että ääniaktivoitu ilmaisin (VAD) on käytettävissä erottamaan kehysten, jotka sisältävät kohinallista puhetta ja kehysten, jotka sisältävät ainoastaan kohinaa, välillä. Oletetaan, että $\Phi_V(\omega)$ estimoidaan ei-puheaktiviteetin aikana laskemalla keskiarvo useiden kehysten yli, esimerkiksi kaavalla

$$\hat{\Phi}_v(\omega)^{\ell} = \rho\,\hat{\Phi}_v(\omega)^{\ell-1} + (1-\rho)\bar{\Phi}_v(\omega) \qquad (4)$$

Kaavassa (4) $\hat{\Phi}_V(\omega)^{\ell}$ on keskimääräistetty PSD-estimaatti perustuen dataan aina kehysnumeroon ℓ asti ja sisältäen tämän, ja $\bar{\Phi}_V(\omega)$ on estimaatti, joka perustuu nykyiseen kehykseen. Skalaari $\rho \in (0,1)$ viritetään $v(k)$:n oletettuun stationaarisuuteen nähden. Keskimääräinen arvo τ kehystä vastaa karkeasti arvoa ρ seuraavan kaavan mukaisesti:

$$\frac{2}{1-\rho} = \tau \qquad (5)$$

Sopiva PSD-estimaatti (olettaen, että taustakohinan spektrimuodolle ei ole olemassa olevia oletuksia) saadaan kaavasta:

$$\bar{\Phi}_v(\omega) = \frac{1}{N}V(\omega)V^*(\omega) \qquad (6)$$

jossa "*" merkitsee kompleksikonjugaattia ja jossa $V(\omega)$ =

$\mathcal{F}(v(k))$. Kun $\mathcal{F}(\cdot)$ = FFT$(\cdot)$ (nopea Fourier-muunnos), $\bar{\Phi}_V(\omega)$ on periodogrammi ja $\hat{\Phi}_V(\omega)$ kaavassa (4) on keskimääräinen periodogrammi, jotka molemmat johtavat asymptoottisesti (N >> 1) tasapuolisiin (engl. unbiased) PSD-estimaatteihin likimääräisillä variansseilla

$$\mathrm{Var}(\bar{\Phi}_v(\omega)) \approx \Phi_v^2(\omega)$$

$$\mathrm{Var}(\hat{\Phi}_v(\omega)) \approx \frac{1}{\tau}\Phi_v^2(\omega) \qquad (7)$$

Vastaava kaava kuin (7) pätee $\Phi_X(\omega)$:lle puheaktiviteetin aikana (korvataan kaavassa (7) $\Phi_v^2(\omega)$ arvolla $\Phi_x^2(\omega)$).

Kuviossa 1 on esitetty spektrivähennys kohinan vaimennusmenetelmä, joka soveltuu suorittamaan esillä olevan keksinnön mukaista menetelmää. Mikrofonista 10 äänisignaali $x(t)$ viedään A/D-muuntimelle 12. A/D- muunnin 12 lähettää eteenpäin digitalisoituja ääninäytteitä kehysmuodossa $\{x(k)\}$ muunnoslohkoon 14, esim. FFT-lohkoon (nopea Fourier-muunnos), joka muuntaa kunkin kehyksen vastaavaksi taajuusmuunnetuksi kehykseksi $\{X(\omega)\}$:n. Muunnettu kehys suoritetaan $\hat{H}(\omega)$:n toimesta lohkossa 16. Tässä askeleessa suoritetaan todellinen spektrivähennys. Näin saatu signaali $\{\hat{S}(\omega)\}$ muunnetaan takaisin aika-alueelle käänteismuunnoslohkossa 18. Tämän tuloksena on kehys $\{\hat{s}(k)\}$, jossa kohina on vaimennettu. Tämä kehys voidaan lähettää eteenpäin kaiunkumoajaan 20 ja tämän jälkeen puheen koodaajaan 22. Puhekoodattu signaali viedään tämän jälkeen kanavan koodaajaan ja modulaattoriin lähetettäväksi (näitä elementtejä ei ole esitetty).

$\hat{H}(\omega)$:n todellinen muoto lohkossa 16 riippuu estimaateista $\Phi_X(\omega)$, $\Phi_V(\omega)$, jotka on muodostettu PSD-estimaattorissa 24, ja näiden estimaattien analyyttisesta lausekkeesta, jota käytetään. Kuviossa 2 seuraavassa osassa esitetään esimerkkejä eri lausekkeista. Seuraavan selostuksen pääosa keskittyy eri

menetelmiin estimaattien $\Phi_X(\omega)$, $\Phi_V(\omega)$ muodostamiseksi ottokehyksestä $\{x(k)\}$.

PSD-estimaattoria 24 ohjataan ääniaktiviteetti-ilmaisimella (VAD) 26, joka käyttää ottokehystä $\{x(k)\}$ määrittämään sisältääkö kehys puhetta (S) tai taustakohinaa (B). Sopiva VAD esitetään viitejulkaisuissa [5], [6]. VAD voidaan implementoida tilakoneena, jossa on 4 tilaa, kuten on esitetty kuviossa 2. Näin saatu ohjaussignaali S/B viedään eteenpäin PSD-estimaattorille 24. Mikäli VAD 26 osoittaa puhetta (S), tilat 21 ja 22, PSD-estimaattori 24 muodostaa $\Phi_X(\omega)$:n. Toisaalta mikäli VAD 26 osoittaa ei-puheaktiviteettia, tila 20, PSD-estimaattori 24 muodostaa $\Phi_V(\omega)$:n. Jälkimmäistä estimaattia käytetään muodostamaan $\hat{H}(\omega)$ seuraavan puhekehyssekvenssin aikana (yhdessä $\Phi_X(\omega)$:n kanssa sekvenssin kullakin kehyksellä).

Signaali S/B lähetetään myös eteenpäin spektrivähennyslohkoon 16. Tällä tavalla lohko 16 voi käyttää eri suodattimia puhe- ja ei-puhekehysten aikana. Puhekehysten $\hat{H}(\omega)$ aikana vallitsee yllä mainittu lauseke $\Phi_X(\omega), \Phi_V(\omega)$. Toisaalta ei-puhekehysten $\hat{H}(\omega)$ aikana voi olla vakio H ($0 \leq H \leq 1$), joka vähentää taustaäänitasoa samalle tasolle kuin taustaäänitaso, joka jää puhekehyksiin kohinan vaimennuksen jälkeen. Tällä tavalla vastaanotettu kohinataso on sama sekä puhe- että ei-puhekehysten aikana.

Ennen kuin antosignaali $\hat{s}(k)$ lasketaan kaavassa (3), $\hat{H}(\omega)$ voidaan edullisessa suoritusmuodossa jälkisuodattaa seuraavan kaavan mukaisesti:

$$H_p(\omega) = \max\left(0.1, W(\omega)\bar{H}(\omega)\right) \qquad \forall\omega \tag{8}$$

jossa $\bar{H}(\omega)$ lasketaan taulukon 1 mukaisesti. Skalaari 0.1 tarkoittaa, että kohinataso on -20 dB.

Taulukko 1. Jälkisuodatustoiminnot.

Tila (st)	$\bar{H}(\omega)$	Kommentti
0	1 ($\forall\omega$)	$\hat{s}(k) = x(k)$
20	0.316 ($\forall\omega$)	vaimennus -10 dB
21	0.7 $\hat{H}(\omega)$	varovainen suodatus (-3 dB)
22	$\hat{H}(\omega)$	

Edelleen signaali S/B viedään myös eteenpäin puhekooderille 22. Tämä mahdollistaa puheen ja taustaäänien eri koodauksen.

PSD-virheanalyysi

On ilmeistä, että stationaarioletukset, jotka on kohdistettu s(k):hon ja v(k):hon aiheuttavat rajoja koskien sitä, miten tarkka estimaatti $\hat{s}(k)$ on verrattuna kohinattomaan puhesignaaliin s(k). Tässä osassa esitetään analyysitekniikka spektrivähennysmenetelmiä varten. Se perustuu PSD-estimaattien $\Phi_X(\omega)$ ja vastaavasti $\Phi_V(\omega)$ (katso (11) alla) ensimmäisen asteen approksimaatioon, yhdistettynä järjestelmään tulleiden vaihtelujen tarkkuuden approksimoituihin lausekkeisiin (nolla-asteen approksimointi). Erityisesti, johdetaan seuraavassa lauseke taajuusaluevirheelle estimoidulle signaalille $\hat{s}(k)$, joka on peräisin käytetystä menetelmästä (siirtofunktion $H(\omega)$ valinta) ja johtuen käytettyjen PSD-estimaattoreiden tarkkuudesta. Koska ihmiskorva on epäherkkä vaihevääristymään on relevanttia käsitellä PSD- virhe, joka on kaavan 9 mukainen:

$$\tilde{\Phi}_s(\omega) = \hat{\Phi}_s(\omega) - \Phi_s(\omega) \qquad (9)$$

jossa

$$\hat{\Phi}_s(\omega) = \hat{H}^2(\omega)\,\Phi_x(\omega) \qquad (10)$$

Huomaa, että $\Phi_S(\omega)$ rakenteeltaan on virhetermi, joka esittää

erotuksen (taajuusalueella) suodatetun kohinamittauksen suuruuden ja puheen suuruuden välillä. Siksi $\Phi_S(\omega)$ voi saada sekä positiivisia että negatiivisia arvoja, eikä ole minkä tahansa aika-aluesignaalin PSD. Kaavassa (10) $\hat{H}(\omega)$ merkitsee $H(\omega)$:n estimaattia perustuen $\Phi_X(\omega)$:hen ja $\Phi_V(\omega)$:hen. Tässä osassa analyysi rajoitetaan tapaukseen tehovähennys (PS; Power Subtraction), [2]. Muita $\hat{H}(\omega)$:n valintoja voidaan analysoida samalla tavalla (katso liitteet A - C). Lisäksi uusia $\hat{H}(\omega)$:n valintoja otetaan käyttöön ja analysoidaan (katso liitteet D - G). Yhteenveto eri sopivista $\hat{H}(\omega)$:n valinnoista saadaan taulukosta 2.

Taulukko 2:

Esimerkkejä eri spektrivähennysmenetelmistä: Tehovähennys (PS) (vakio PS, $\hat{H}_{PS}(\omega)$ arvolla $\delta = 1$), suuruusvähennys (MS; Magnitude Subtraction), spektrivähennysmenetelmiä perustuen Wiener-suodatukseen (WF; Wiener Filtering) ja suurimpaan todennäköisyyteen (ML; Maximum Likelihood) -menetelmiin ja parannettuun tehovähennykseen (IPS; Improved Power Subtraction) esillä olevan keksinnön edullisen suoritusmuodon mukaisesti.

$$\hat{H}(\omega)$$

$$\hat{H}_{\delta PS}(\omega) = \sqrt{1 - \delta\hat{\Phi}_v(\omega)/\hat{\Phi}_x(\omega)}$$

$$\hat{H}_{MS}(\omega) = 1 - \sqrt{\hat{\Phi}_v(\omega)/\hat{\Phi}_x(\omega)}$$

$$\hat{H}_{WF}(\omega) = \hat{H}^2_{PS}(\omega)$$

$$\hat{H}_{ML}(\omega) = \tfrac{1}{2}(1 + \hat{H}_{PS}(\omega))$$

$$\hat{H}_{IPS}(\omega) = \sqrt{\hat{G}(\omega)}\hat{H}_{PS}(\omega)$$

Määritellään, että $H(\omega)$ kuuluu väliin $0 \leq H(\omega) \leq 1$, joka ei välttämättä päde taulukon 2 vastaaville estimoiduille suu-

reille, ja siksi käytetään käytännössä puoliaalto- tai kokoaaltosuuntausta [1].

Analyysin suorittamiseksi oletetaan, että kehyspituus N on riittävän suuri (N >> 1) jotta $\Phi_X(\omega)$ ja $\Phi_V(\omega)$ ovat olennaisesti tasapuolisia. Otetaan käyttöön ensimmäisen asteen poikkeamat

$$\hat{\Phi}_x(\omega) = \Phi_x(\omega) + \Delta_x(\omega)$$

$$\hat{\Phi}_v(\omega) = \Phi_v(\omega) + \Delta_v(\omega) \qquad (11)$$

jossa $\Delta_X(\omega)$ ja $\Delta_V(\omega)$ ovat nolla-keskiarvon omaavia stokastisia muuttujia siten, että $E[\Delta_x(\omega)/\Phi_x(\omega)]^2 << 1$ ja $E[\Delta_V(\omega)/\Phi_V(\omega)]^2 << 1$. Tässä ja jatkossa tarkoitetaan merkinnällä $E[\cdot]$ tilastollista olettamaa. Edelleen, mikäli kohinan korrelointiaika on lyhyt verrattuna kehyspituuteen $E[(\bar{\Phi}_V(\omega)^{\ell} - \Phi_V(\omega))\ (\bar{\Phi}_V(\omega)^{k} - \Phi_V(\omega))] \approx 0$ kun $\ell \neq k$, jossa $\bar{\Phi}_V(\omega)^{\ell}$ on estimointi, joka perustuu ℓ:nnen kehyksen dataan. Tämä tarkoittaa, että $\Delta_X(\omega)$ ja $\Delta_V(\omega)$ ovat suunnilleen toisistaan riippumattomat. Muutoin, mikäli kohina on hyvin korreloitua, oletetaan, että $\Phi_V(\omega)$:llä on rajoitettu (<< N) lukumäärä (voimakkaita) huippuja, jotka sijaitsevat taajuuksilla $\omega_1, \ldots, \omega_n$. Tällöin $E[\bar{\Phi}_v(\omega)^{\ell} - \Phi_V(\omega))(\bar{\Phi}_V(\omega)^{k} - \Phi_V(\omega))] \approx 0$ pätee arvoilla $\omega \neq \omega_{jj} = 1, \ldots, n$ ja $\ell \neq k$ ja analyysi on edelleen tosi arvoilla $\omega \neq w_{jj} = 1, \ldots, n$.

Kaava (11) tarkoittaa, että asymptoottisia (N >> 1) tasapuolisia PSD-estimaattoreita, kuten periodogrammia tai keskimääräistettyä periodogrammia, käytetään. Kuitenkin, käyttämällä asymptoottisesti suunnattuja PSD-estimaattoreita, kuten Blackman-Turkey PSD-estimaattoreita, vastaava analyysi on tosi mikäli korvataan kaava (11) seuraavalla kaavalla:

$$\hat{\Phi}_x(\omega) = \Phi_x(\omega) + \Delta_x(\omega) + B_x(\omega)$$

ja

$$\hat{\Phi}_v(\omega) = \bar{\Phi}_v(\omega) + \Delta_v(\omega) + B_v(\omega)$$

jossa vastaavasti $B_X(\omega)$ ja $B_V(\omega)$ ovat deterministisiä termejä, jotka esittävät PSD-estimaattoreiden asymptoottista suuntautumista.

Edelleen kaava (11) tarkoittaa, että $\Phi_S(\omega)$ kaavassa (9) on (ensimmäisen asteen approksimaatiolla) lineaarinen funktio $\Delta_X(\omega)$:ssä ja $\Delta_V(\omega)$:ssä. Seuraavassa eri menetelmien suoritus suuntaumavirheen (bias error) ($E[\Phi_S(\omega)]$) suhteen ja virhevarianssin ($Var(\Phi_S(\omega))$) suhteen tutkitaan. Täydellinen $\hat{H}_{PS}(\omega)$:n johtaminen suoritetaan seuraavassa osassa. Vastaavia johtamisia muille spektrivähennysmenetelmille taulukossa 1 esitetään liitteissä A - G.

$\hat{H}_{PS}(\omega)$ ($\hat{H}_{\delta PS}(\omega)$ analyysi kun $\delta = 1$

Sovittamalla kaava (10) ja $\hat{H}_{PS}(\omega)$ taulukosta 2 kaavaan (9), käyttämällä Taylor-sarjalaajennusta $(1+x)^{-1} \simeq 1 - x$ ja jättämällä huomioitta suuremman asteen kuin ensimmäisen asteen poikkeamat, suora laskeminen antaa

$$\tilde{\Phi}_s(\omega) \simeq \frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega) - \Delta_v(\omega) \qquad (12)$$

jossa "$\simeq$" käytetään merkitsemään likimääräistä yhtäläisyyttä, jossa ainoastaan dominantit termit ovat. Suureet $\Delta_X(\omega)$ ja $\Delta_V(\omega)$ ovat nolla-keskiarvon omaavia stokastisia muuttujia. Täten

$$E[\tilde{\Phi}_s(\omega)] \simeq 0 \qquad (13)$$

ja

$$Var(\tilde{\Phi}_s(\omega)) \simeq \frac{\Phi_v^2(\omega)}{\Phi_x^2(\omega)}Var(\hat{\Phi}_x(\omega)) + Var(\hat{\Phi}_v(\omega)) \qquad (14)$$

Jotta voisimme jatkaa, käytetään yleistä tulosta, että asymptoottisella tasapuolisella spektriestimaattorilla $\hat{\Phi}(\omega)$, kts. esim. (7)

$$\mathrm{Var}(\hat{\Phi}(\omega)) \simeq \gamma(\omega)\,\Phi^2(\omega) \tag{15}$$

muutamalla (mahdollisesti taajuusriippuvalla) muuttujalla $\gamma(\omega)$. Esimerkiksi periodogrammi vastaa $\gamma(\omega) = 1 + (\sin\omega N / N\sin\omega)^2$, joka arvolla N >> 1 yksinkertaistuu muotoon $\gamma \approx 1$. Yhdistämällä kaavat (14) ja (15) saadaan

$$\mathrm{Var}(\bar{\Phi}_s(\omega)) \simeq \gamma\,\Phi_v^2(\omega) \tag{16}$$

$\hat{H}_{MS}(\omega)$:n tuloksia

Vastaavat laskennat $\hat{H}_{MS}(\omega)$:lle antavat (yksityiskohdat on esitetty liitteessä A):

$$E[\bar{\Phi}_s(\omega)] \simeq 2\Phi_v(\omega)\left(1-\sqrt{\frac{\Phi_x(\omega)}{\Phi_v(\omega)}}\right)$$

ja

$$\mathrm{Var}(\bar{\Phi}_s(\omega)) \simeq \left(1-\sqrt{1+\frac{\Phi_s(\omega)}{\Phi_v(\omega)}}\right)^2 \gamma\,\Phi_v^2(\omega)$$

$\hat{H}_{WF}(\omega)$:n tuloksia

Laskennat $\hat{H}WF_{MS}(\omega)$:lle antavat (yksityiskohdat on esitetty liitteessä B):

$$E[\bar{\Phi}_s(\omega)] \simeq -\left(1-\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\right)\Phi_v(\omega)$$

ja

$$\mathrm{Var}(\bar{\Phi}_s(\omega)) \simeq 4\left(1-\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\right)^2 \gamma\Phi_v^2(\omega)$$

$\hat{H}_{ML}(\omega)$:n tuloksia

Laskennat $\hat{H}_{ML}(\omega)$:lle antavat (yksityiskohdat on esitetty liitteessä C):

$$E[\tilde{\Phi}_s(\omega)] \simeq \frac{1}{2}\Phi_v(\omega) - \frac{1}{4}\left(\sqrt{\Phi_x(\omega)} - \sqrt{\Phi_s(\omega)}\right)^2$$

ja

$$\mathrm{Var}(\tilde{\Phi}_s(\omega)) \simeq \frac{1}{16}\left(1 + \sqrt{\frac{\Phi_x(\omega)}{\Phi_s(\omega)}}\right)^2 \gamma\Phi_v^2(\omega)$$

$\hat{H}_{IPS}(\omega)$:n tuloksia

Laskennat $\hat{H}_{IPS}(\omega)$:lle antavat $\hat{H}_{IPS}(\omega)$ on johdettu liitteessä D ja analysoitu liitteessä E):

$$E[\tilde{\Phi}_s(\omega)] \simeq (\bar{G}(\omega) - 1)\Phi_s(\omega)$$

ja

$$\mathrm{Var}(\tilde{\Phi}_s(\omega)) \simeq \bar{G}^2(\omega) \times \left(\bar{G}(\omega) + \gamma\Phi_v(\omega)\frac{\Phi_v(\omega) + 2\Phi_x(\omega)}{\Phi_s^2(\omega) + \gamma\Phi_v^2(\omega)}\right)^2 \gamma\Phi_v^2(\omega)$$

Yhteiset ominaisuudet

Esitetyillä menetelmillä todetaan, että suuntaumavirhe riippuu ainoastaan $\hat{H}(\omega)$:n valinnasta, kun toisaalta virhevarianssi riippuu sekä $\hat{H}(\omega)$:n valinnasta että käytettyjen PSD-estimaattoreiden varianssista. Esimerkiksi keskimääräisellä periodogrammi-estimaatilla koskien $\Phi_Y(\omega)$:tä saadaan kaavasta (7), että $\gamma_Y \approx 1/\tau$. Toisaalta, käyttäen yksittäistä kehysperiodogrammia $\Phi_X(\omega)$:n estimaatille saadaan $\gamma_x \approx 1$. Täten arvolla $\tau >> 1$ dominoiva termi kaavassa $\gamma = \gamma_x + \gamma_v$, joka esiintyy yllä olevissa varianssikaavoissa, on γ_x, ja täten päävirhelähde on yksittäinen kehys PSD-estimaatti, joka perustuu kohinalliseen puheeseen.

Yllä olevista huomautuksista seuraa, että spektrivähennystekniikoiden parantamiseksi on toivottavaa pienentää γ_x:n arvo (valita sopiva PSD-estimaattori, joka on lähes tasapuolinen estimaattori, jolla on mahdollisimman hyvät suoritusarvot) ja valita "hyvä" spektrivähennystekniikka (valita $\hat{H}(\omega)$). Esillä olevan keksinnön perusideana on, että γ_x:n arvo voidaan pienentää käyttäen fyysistä mallinnusta (vähennetään N:n vapausasteiden lukumäärä (näytteiden lukumäärä kehyksessä) arvoksi, joka on pienempi kuin N) äänialueella. On yleisesti tunnettua, että s(k) voidaan esittää tarkasti autoregressiivisellä (AR) mallilla (tyypillisesti kertalukua p ≈ 10). Tätä käsitellään seuraavassa kahdessa osassa.

Lisäksi $\Phi_s(\omega)$:n tarkkuus (ja täten $\tilde{s}(k)$:n tarkkuus) riippuu $\hat{H}(\omega)$:n valinnasta. Uusia edullisia $\hat{H}(\omega)$:n valintoja johdetaan ja analysoidaan liitteissä D - G.

Puheen AR mallinnus

Esillä olevan keksinnön edullisessa suoritusmuodossa s(k) mallinnetaan autoregressiivisenä (AR) prosessina

$$s(k) = \frac{1}{A(q^{-1})} w(k) \quad k = 1, \ldots, N \tag{17}$$

jossa $A(q^{-1})$ on p:nnen asteen polynomi, jossa ensimmäinen kerroin on yksi taaksepäin siirto operaattorissa ($q^{-1}\omega(k) = \omega(k-1)$, jne.)

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_p q^{-p} \tag{18}$$

ja ω(k) on valkoinen kohina, jolla on nolla-keskiarvo ja varianssi σ_w^2. Ensiksi voi vaikuttaa rajoittavalta tutkia ainoastaan AR-malleja. Kuitenkin AR-mallien käyttö puhemallinnusta varten on motivoitua sekä äänialueen fyysisen mallinnuksen kannalta, että, joka tässä on tärkeämpää, kohinallisen puheen fyysisten rajoitusten kannalta estimoitujen mallien tarkkuuteen.

Puhesignaaliprosessoinnissa N:n kehyspituus ei ehkä ole riittävän suuri sallimaan keskimääräistämistekniikoiden soveltamista kehyksen sisällä varianssin vähentämiseksi ja edelleen PSD-estimaattorin tasapuolisuuden säilyttämiseksi. Täten, ensimmäisen termin vaikutuksen vähentämiseksi esimerkiksi kaavassa (12) on käytettävä äänialueen fyysistä mallinnusta. AR-rakenne (17) sovitetaan s(k):hon. Tällöin

$$\Phi_x(\omega) = \frac{\sigma_w^2}{|A(e^{i\omega})|^2} + \Phi_v(\omega) \qquad (19)$$

Lisäksi $\Phi_V(\omega)$ voidaan esittää parametrimallina

$$\Phi_v(\omega) = \sigma_v^2 \frac{|B(e^{i\omega})|^2}{|C(e^{i\omega})|^2} \qquad (20)$$

jossa $B(q^{-1})$ ja $C(q^{-1})$ ovat vastaavasti q:nnen ja r:nnen asteen polynomeja, jotka on määritetty vastaavasti kuin $A(q^{-1})$ kaavassa (18). Yksinkertaisuuden vuoksi käytetään parametrista kohinamallia (20) esityksessä alla, jossa estimoidaan parametrimallin kertalukua. On kuitenkin ymmärrettävää, että taustakohinan myös muut mallit ovat mahdollisia. Yhdistämällä kaavat (19 ja (20) saadaan

$$x(k) = \frac{D(q^{-1})}{A(q^{-1})\,C(q^{-1})}\eta(k) \quad k = 1, \ldots, N \qquad (21)$$

jossa $\eta(k)$ on valkoinen kohina, jonka keskiarvo on nolla ja jonka varianssi on σ_η^2 ja jossa $D(q^{-1})$ saadaan seuraavasta kaavasta:

$$\sigma_\eta^2 |D(e^{i\omega})|^2 = \sigma_w^2 |C(e^{i\omega})|^2 + \sigma_v^2 |B(e^{i\omega})|^2 |A(e^{i\omega})|^2 \qquad (22)$$

Puheen parametrin estimointi

Parametrien estimointi kaavoissa (17) - (18) on yksinkertaista mikäli lisäkohinaa ei esiinny. Huomaa, että kohinattomassa tapauksessa kaavan (22) oikealla puolella oleva toinen termi

häviää, ja täten kaava (21) yksinkertaistuu kaavaksi (17) yksinkertaistettaessa.

Tässä etsitään PSD-estimaattoria perustuen autokorrelaatiomenetelmään. Tähän on neljä syytä.

- Autokorrelaatiomenetelmä on yleisesti tunnettu. Erityisesti estimoiduilla parametreilla on minimivaihe, joka varmistaa aikaansaatavan suodatuksen stabiilisuuden.
- Käyttämällä Levinson-algoritmia menetelmä on helppo implementoida, ja sillä on pieni laskennallinen monimutkaisuus.
- Optimaalinen proseduuri sisältää ei-lineaarisen optimoinnin, joka vaatii jonkun initialisointiproseduurin. Autokorrelaatiomenetelmä ei tarvitse tällaista.
- Käytännön syistä on edullista mikäli samaa estimointiproseduuria voidaan käyttää huononnettua puhetta ja vastaavasti puhdasta puhetta varten, kun tällainen on saatavana. Toisin sanoen estimointimenetelmän tulisi olla riippumaton toimintoskenaariosta, eli riippumaton puhe/kohina- suhteesta.

On yleisesti tunnettua, että ARMA-malli (kuten kaava (21)) voidaan mallintaa äärettömän kertaluvun AR- prosessilla. Kun äärellinen lukumäärä dataa on käytettävissä parametrin estimointiin, äärettömän kertaluvun AR-malli on katkaistava. Tässä käytettävä malli on

$$x(k) = \frac{1}{F(q^{-1})}\eta(k) \tag{23}$$

jossa $F(q^{-1})$ on kertalukua $\bar{p}$. Sopiva mallikertaluku ilmenee selostuksesta alla. Likimääräinen malli (23) on lähellä puhetta kohinaprosessissa mikäli niiden PSD:t ovat suurin piirtein yhtä suuret, eli

$$\frac{|D(e^{i\omega})|^2}{|A(e^{i\omega})|^2\,|C(e^{i\omega})|^2} \approx \frac{1}{|F(e^{i\omega})|^2} \tag{24}$$

Perustuen äänialueen fysikaaliseen mallinnukseen on yleistä, että $p = \deg(A(q^{-1})) = 10$. Kaavasta (24) seuraa myös, että $\bar{p} = \deg(F(q^{-1}) >> \deg(A(q^{-1})) + \deg(C(q^{-1})) = p + r$, jossa $p + r$ karkeasti on yhtä suuri kuin huippujen lukumäärä $\Phi_x(\omega)$:ssä. Toisaalta kohinallisten kapeakaistaprosessien mallinnus käyttäen AR-mallia vaatii, että $\bar{p} << N$ jotta varmistetaan PSD-estimaattien luotettavuus. Yhteenvetona

$$p + r \ll \bar{p} \ll N$$

Sopiva peukalosääntö on $\bar{p} \sim \sqrt{N}$. Yllä olevasta esityksestä voidaan todeta, että parametrinen näkemys on onnistunut mikäli $N >> 100$. Kaavasta (22) voidaan myös todeta, että mitä tasaisempi kohinaspektri on, sitä pienempiä N:n arvoja sallitaan. Vaikkakaan $\bar{p}$ ei ole tarpeeksi suuri, parametrinen suhtautuminen antaa todennäköisesti riittävän hyviä tuloksia. Tämä johtuu siitä, että koskien virhevarianssia parametrinen käsittely antaa olennaisesti enemmän tarkkoja PSD-estimaatteja kuin periodogrammipohjainen käsittely (tyypillisessä esimerkissä varianssien välinen suhde on 1:8, katso alla), joka olennaisesti vähentää tekotuotteita äänikohinana annossa.

Parametrisesta PSD-estimaattorista tehdään yhteenveto seuraavasti. Käytetään autokorrelaatiomenetelmää ja korkean asteen AR-mallia (mallin kertoluku $\bar{p} >> p$ ja $\bar{p} \sim \sqrt{N}$) AR-parametrien laskemiseksi $\{\hat{f}_1, \ldots, \hat{f}_{\bar{p}}\}$, ja kohinan varianssin $\hat{\sigma}_\eta^2$ laskemiseksi kaavassa (23). Estimoidusta AR-mallista lasketaan (erillisessä N-pisteessä, jotka vastaavat $X(\omega)$:n taajuussäiliöitä kaavassa (3)) $\hat{\Phi}_x(\omega)$ seuraavan kaavan mukaisesti:

$$\hat{\Phi}_x(\omega) = \frac{\hat{\sigma}_\eta^2}{|\hat{F}(e^{i\omega})|^2} \qquad (25)$$

Tämän jälkeen käytetään jotain spektrivähennystekniikoista taulukossa 2 puheen s(k) parantamiseksi.

Seuraavaksi käytetään alhaisen kertaluvun approksimaatiota parametrisen PSD-estimaattorin varianssille (vastaava kuin (7)

ei-parametrisille menetelmille) ja täten käytetään s(k):n Fourier-sarjalaajennusta olettaen, että kohina on valkoista. Tämän jälkeen asymptoottinen (sekä datalukumäärälle (N >> 1) ja malli kertaluvulle ($\bar{p}$ >> 1) varianssi $\hat{\Phi}_x(\omega)$):lle saadaan kaavasta

$$\mathrm{Var}(\hat{\Phi}_x(\omega)) \simeq \frac{2\bar{p}}{N}\Phi_x^2(\omega) \qquad (26)$$

Yllä oleva lauseke pätee myös puhtaalle (korkean asteen) AR-prosessille. Kaavasta (26) seuraa suoraan, että $\gamma_x \approx 2\bar{p}/N$, että yllä mainitun peukalosäännön mukaisesti pätee lähes $\gamma_x \simeq 2/\sqrt{N}$, joka on verrattava siihen, että $\gamma_x \approx 1$, joka pätee periodogrammi-pohjaiselle PSD-estimaattorille.

Esimerkiksi matkapuhelimien kädet vapaana ympäristössä on todennäköistä olettaa, että kohina on stationaarista noin 0,5 s (8000 Hz:n näytteenottotaajuudella ja kehyspituudella N = 256), josta saadaan $\tau \approx 15$ ja täten $\gamma_v \simeq 1/15$. Edelleen arvolla $\bar{p} = \sqrt{N}$ saadaan, että $\gamma_x = 1/8$.

Kuviossa 3 esitetään ero periodogrammi PSD-estimaatin ja parametrisen PSD-estimaatin, joka on esillä olevan keksinnön mukainen, välillä, tyypillisellä puhekehyksellä. Tässä esimerkissä N = 256 (256 näytettä) ja siinä käytetään AR-mallia, jossa on 10 parametria. Todetaan, että parametrin PSD-estimaatti $\hat{\Phi}_x(\omega)$ on paljon tasaisempi kuin vastaava periodogrammi PSD-estimaatti.

Kuviossa 4 esitetään 5 sekuntia näytteenotettua äänisignaalia, joka sisältää puhetta kohinallisessa taustassa. Kuviossa 5 esitetään kuvion 4 signaali kun spektrivähennys on suoritettu perustuen periodogrammi PSD-estimaattiin, joka priorisoi korkean äänilaadun. Kuviossa 6 esitetään kuvion 4 signaali kun spektrivähennys on suoritettu perustuen parametri PSD-estimaattiin esillä olevan keksinnön mukaisesti.

Vertaamalla kuvioita 5 ja 6 nähdään, että aikaansaadaan huomattava kohinan vaimeneminen (suuruusluokkaa 10 dB) esillä olevan keksinnön mukaisesti. (Kuten todettiin yllä kuvion 1 selostuksen yhteydessä pienentyneet kohinatasot ovat samat sekä puhe- että ei-puhekehyksissä. Toinen ero, joka ilmenee kuviosta 6, on, että näin saatu puhesignaali on vähemmän vääristynyt kuin kuvion 5 puhesignaali.

Teoreettiset tulokset, koskien suuntaumaa ja virhevarianssia PSD:n virheellä, kaikilla tutkituilla menetelmillä, on koottu taulukkoon 3.

On mahdollista järjestää eri menetelmät paremmuusjärjestykseen. Voidaan ainakin löytää kaksi kriteeriä sopivan menetelmän valitsemiseksi.

Ensinnä, pienillä hetkellisillä SNR:illä on toivottavaa, että menetelmällä on alhainen varianssi tekotuoteäänien estämiseksi $\hat{s}(k)$:ssa. Tämä ei ole mahdollista ilman, että suuntauma lisääntyy, ja tämän suuntaumatermin tulisi, jotta taajuusalueet vaimenisivat (eikä vahvistuisi) pienellä hetkellisellä SNR:llä, olla negatiivinen etumerkki (täten pakottaen $\Phi_S(\omega)$:tä kaavassa (9) kohti nollaa). Kandidaatit, jotka täyttävät tämän ehdon, ovat vastaavasti MS, IPS ja WF.

Toiseksi, korkeilla hetkellisillä SNR:illä toivotaan pientä puhevääristymän osuutta. Edelleen, mikäli suuntaumatermi on dominoiva, sillä tulee olla positiivinen etumerkki. ML, $\bar{\delta}$PS, PS, IPS ja (mahdollisesti) WF täyttävät tämän ehdon. Suuntaumatermi dominoi MSE-lausekkeesta ainoastaan ML:llä ja WF:llä, jossa suuntaumatermien etumerkki on positiivinen ML:llä ja vastaavasti negatiivinen WF:llä. Täten ML, $\bar{\delta}$PS, PS ja IPS täyttävät tämän kriteerin.

Algoritmisia näkökohtia

Tässä osassa esitetään edullisia suoritusmuotoja koskien

spektrivähennysmenetelmää esillä olevan keksinnön mukaisesti viittaamalla kuvioon (7).

1. Otto: $x = \{x(k) | k = 1, \ldots, N\}$.
2. Muuttujat

Taulukko 3:

Suuntauma ja varianssilausekkeet tehovähennystä (PS) (vakio PS, $\hat{H}_{PS}(\omega)$ arvolla $\delta = 1$), suuruusvähennystä (MS), parannettua tehovähennystä (IPS) ja spektrivähennysmenetelmiä varten, jotka perustuvat Wiener-suodatukseen (WF) ja maksimitodennäköisyys (ML) metodeihin. Hetkellinen SNR määritetään $SNR = \Phi_S(\omega)/\Phi_V(\omega)$. PS:llä optimaalinen vähennyskerroin $\bar{\delta}$ saadaan kaavasta (58) ja IPS, $\hat{G}(\omega)$ saadaan kaavasta (45), jossa $\Phi_x(\omega)$ ja $\Phi_V(\omega)$ korvataan vastaavasti $\hat{\Phi}_x(\omega)$:llä ja $\hat{\Phi}_V(\omega)$:llä.

$\hat{H}(\omega)$	Suuntauma	Varianssi
	$E[\bar{\Phi}_s(\omega)]/\Phi_v(\omega)$	$\mathrm{Var}(\bar{\Phi}_s(\omega))/\gamma\Phi_v^2(\omega)$
δPS	$1-\delta$	δ^2
MS	$-2(\sqrt{1+\mathrm{SNR}}-1)$	$(\sqrt{1+\mathrm{SNR}}-1)^2$
IPS	$-\frac{\gamma \mathrm{SNR}}{\gamma+\mathrm{SNR}^2}$	$\left(\frac{\mathrm{SNR}^2}{\mathrm{SNR}^2+\gamma}\right)^2\left(1+2\gamma\frac{1+\mathrm{SNR}}{\mathrm{SNR}^2+\gamma}\right)^2$
WF	$-\frac{\mathrm{SNR}}{\mathrm{SNR}+1}$	$4\left(\frac{\mathrm{SNR}}{\mathrm{SNR}+1}\right)^2$
ML	$\frac{1}{2}-\frac{1}{4}(\sqrt{\mathrm{SNR}+1}-\sqrt{\mathrm{SNR}})^2$	$\frac{1}{16}\left(1+\sqrt{1+\frac{1}{\mathrm{SNR}}}\right)^2$

$\overline{p}$ puhekohinassa mallin kertaluku

ρ $\hat{\Phi}_V(\omega)$:n keskimääräinen päivityskerroin

3. Kullakin ottodatan kehyksellä suoritetaan:

(a) Puheen tunnistus (askel 110)
Muuttujan "puhe" arvoksi asetetaan "tosi" mikäli VAD-anto on yhtä kuin st = 21 tai st = 22. "Puhe" saa arvon "epätosi" mikäli st = 20. Mikäli VAD-anto on yhtä kuin st = 0, algoritmi käynnistetään uudestaan.

(b) Spektriestimointi
Mikäli puhe, niin estimoidaan $\hat{\Phi}_x(\omega)$:
i. Estimoidaan kertoimet (polynomikertoimet $\{\hat{f}_1, \ldots, \hat{f}_p\}$ ja varianssi $\hat{\sigma}_\eta^2$) kaikki-napamallille (23) käyttäen autokorrelointimenetelmää, jota on sovellettu nolla-keskiarvon säädetylle ottodatalle $\{x(k)\}$ (askel 120).
ii. Lasketaan $\hat{\Phi}_x(\omega)$ kaavan (25) mukaisesti (askel 130).
Muuten estimoidaan $\hat{\Phi}_V(\omega)$ (askel 140).
i. Päivitä taustakohinan spektrimalli $\hat{\Phi}_V(\omega)$ käyttäen kaavaa (4), jossa $\bar{\Phi}_V(\omega)$ periodogrammi perustuen nolla-keskiarvon säädettyyn ja Hanning/Hamming- ikkunoituun ottodataan x. Koska tässä käytetään ikkunoitua dataa, kun toisaalta $\hat{\Phi}_x(\omega)$ perustuu ei-ikkunoituun dataan, on $\hat{\Phi}_V(\omega)$ sopivasti normalisoitava. $\hat{\Phi}_V(\omega)$:n sopiva alkuarvo saadaan ensimmäisen kehyksen periodogrammin "keskimääräisestä" arvosta (taajuussäiliössä), joka on skaalattu esimerkiksi kertoimella 0,25, joka tarkoittaa sitä, että alkuperäinen valkoinen kohinaoletus kohdistetaan taustakohinaan.

(c) Spektrivähennys (askel 150)
i. Lasketaan taajuuspainotusfunktio $\hat{H}(\omega)$ taulukon 1 mukaisesti.
ii. Suoritetaan mahdollisesti jälkisuodatus, vaimennus ja kohinatason säätö.
iii. Lasketaan anto käyttäen kaavaa (3) ja nolla-keskiarvon säädettyä dataa $\{x(k)\}$. Data $\{x(k)\}$ voidaan ikkunoida tai ei, riippuen todellisesta kehyksen limittymisestä (suorakaiteen muotoista ikkunaa käytetään ei-limittyvillä kehyk-

sillä, kun toisaalta Hanning-ikkunaa käytetään 50 %:n limittymisellä.

Yllä esitetystä selostuksesta on selvää, että esillä oleva keksintö aikaansaa huomattavaan kohinan pienenemisen vaikuttamatta huonontavasti äänen laatuun. Tätä parannusta voidaan selittää erillisillä tehospektriestimointimenetelmillä, joita käytetään puheelle ja ei-puhekehyksille. Nämä menetelmät hyödyntävät puhe- ja ei-puhesignaalien (taustakohinan) eri luonnetta vastaavien tehospektriestimaattien varianssin minimoimiseksi.

- Ei-puhekehyksillä $\hat{\Phi}_v(\omega)$ estimoidaan ei-parametrisella tehospektri-estimointimenetelmällä, esimerkiksi FFT-pohjaisella periodogrammi-estimoinnilla, joka käyttää kunkin kehyksen kaikkia N-näytteitä. Varaamalla kaikki N vapausastetta ei-puhekehyksestä voidaan mallintaa laajempi taustakohinan joukko. Koska oletetaan, että taustakohina on stationaarista useiden kehysten yli, varianssin $\hat{\Phi}_v(\omega)$ pieneneminen voidaan aikaansaada suorittamalla keskiarvo tehospektri- estimaatille useiden ei-puhekehysten yli.

- Puhekehyksillä $\hat{\Phi}_x(\omega)$ estimoidaan parametrisella tehospektri-estimointimenetelmällä, joka perustuu puheen parametrimalliin. Tässä tapauksessa puheen erityisominaisuutta käytetään vapausasteiden lukumäärän pienentämiseksi (parametrisen mallin parametrien lukumäärään) puhekehyksessä. Malli, joka perustuu harvempiin parametreihin, vähentää tehospektriestimaatin varianssia. Tämä tapa on edullinen puhekehyksillä, koska oletetaan, että puhe on stationaarista ainoastaan yhdessä kehyksessä.

Alan ammattimies ymmärtää, että keksintöä voidaan modifioida ja muuttaa poikkeamatta keksinnöstä, jonka suojan määräävät seuraavat patenttivaatimukset.

Liite A

$\hat{H}_{MS}(\omega)$:n analyysi

Sijoittamalla $\hat{H}_{MS}(\omega)$:n laskennat rinnan saadaan

$$\tilde{\Phi}_s(\omega) = \left(1 - \sqrt{\frac{\hat{\Phi}_v(\omega)}{\hat{\Phi}_x(\omega)}}\right)^2 \Phi_x(\omega) - \Phi_s(\omega) \tag{27}$$

$$\simeq \left(1 - \sqrt{\frac{\Phi_x(\omega)}{\Phi_v(\omega)}}\right)\left(2\Phi_v(\omega) - \frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega) + \Delta_v(\omega)\right)$$

jossa toisessa yhtäläisyydessä käytetään myös Taylor- sarjan laajennusta $\sqrt{1+x} \simeq 1 + x/2$. Kaavasta (27) seuraa, että odotusarvo $\Phi_S(\omega)$ on nollasta poikkeava, ja saadaan kaavasta

$$E[\tilde{\Phi}_s(\omega)] \simeq 2\Phi_v(\omega)\left(1 - \sqrt{\frac{\Phi_x(\omega)}{\Phi_v(\omega)}}\right) \tag{28}$$

Edelleen $\mathrm{Var}(\tilde{\Phi}_s(\omega)) \simeq$

$$\left(1 - \sqrt{\frac{\Phi_x(\omega)}{\Phi_v(\omega)}}\right)^2 \left(\frac{\Phi_v^2(\omega)}{\Phi_x^2(\omega)}\mathrm{Var}(\hat{\Phi}_x(\omega)) + \mathrm{Var}(\hat{\Phi}_v(\omega))\right) \tag{29}$$

Yhdistämällä (29) ja (15) saadaan:

$$\mathrm{Var}(\tilde{\Phi}_s(\omega)) \simeq \left(1 - \sqrt{1 + \frac{\Phi_s(\omega)}{\Phi_v(\omega)}}\right)^2 \gamma\, \Phi_v^2(\omega) \tag{30}$$

LIITE B

$\hat{H}_{WF}(\omega)$:n analyysi

Tässä liitteessä PSD-virhe johdetaan puheen korostamiseksi perustuen Wiener-suodatukseen [2]. Tässä tapauksessa $\hat{H}(\omega)$ saadaan kaavasta:

$$\hat{H}_{WF}(\omega) = \frac{\hat{\Phi}_s(\omega)}{\hat{\Phi}_s(\omega) + \hat{\Phi}_v(\omega)} = \hat{H}^2_{PS}(\omega) \tag{31}$$

Tässä $\hat{\Phi}_S(\omega)$ on $\Phi_S(\omega)$:n estimaatti ja toinen yhtäläisyys seuraa siitä, että $\hat{\Phi}_S(\omega) = \hat{\Phi}_x(\omega) - \hat{\Phi}_V(\omega)$. Toteamalla, että

$$\hat{H}^2_{WF}(\omega) \simeq \frac{\Phi_s(\omega)}{\Phi^2_x(\omega)} \left(\Phi_s(\omega) + 2 \left\{ \frac{\Phi_v(\omega)}{\Phi_x(\omega)} \Delta_x(\omega) - \Delta_v(\omega) \right\} \right) \tag{32}$$

saadaan:

$$\begin{aligned} \tilde{\Phi}_s(\omega) \simeq & \left(1 - \frac{\Phi_v(\omega)}{\Phi_x(\omega)} \right) \\ & \times \left(-\Phi_v(\omega) + 2 \left\{ \frac{\Phi_v(\omega)}{\Phi_x(\omega)} \Delta_x(\omega) - \Delta_v(\omega) \right\} \right) \end{aligned} \tag{33}$$

Kaavasta (33) saadaan:

$$E[\tilde{\Phi}_s(\omega)] \simeq - \left(1 - \frac{\Phi_v(\omega)}{\Phi_x(\omega)} \right) \Phi_v(\omega) \tag{34}$$

ja

$$\mathrm{Var}(\tilde{\Phi}_s(\omega)) \simeq 4 \left(1 - \frac{\Phi_v(\omega)}{\Phi_x(\omega)} \right)^2 \gamma \Phi^2_v(\omega) \tag{35}$$

LIITE C

$\hat{H}_{ML}(\omega)$:n analyysi

Kuvaamalla puhe deterministisellä aaltomuodolla, jolla on tuntematon amplitudi ja vaihe, määritetään maksimitodennäköisyyden (ML) spektrivähennysmenetelmä seuraavasti:

$$\begin{aligned}\hat{H}_{ML}(\omega) &= \frac{1}{2}\left(1+\sqrt{1-\frac{\hat{\Phi}_v(\omega)}{\hat{\Phi}_x(\omega)}}\right) \\ &= \frac{1}{2}\left(1+\hat{H}_{PS}(\omega)\right)\end{aligned} \tag{36}$$

Sovittamalla kaava (11) kaavaan (36) saadaan:

$$\begin{aligned}\hat{H}_{ML}(\omega) &\simeq \frac{1}{2}\left(1+\sqrt{\frac{\Phi_s(\omega)}{\Phi_x(\omega)}}\left(1-\frac{\Delta_v(\omega)}{\Phi_s(\omega)}+\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\frac{\Delta_x(\omega)}{\Phi_s(\omega)}\right)^{\frac{1}{2}}\right) \\ &\simeq \frac{1}{2}\left(1+\sqrt{\frac{\Phi_s(\omega)}{\Phi_x(\omega)}}\right) \\ &\quad +\frac{1}{4}\frac{1}{\sqrt{\Phi_x(\omega)\Phi_s(\omega)}}\left(\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega)-\Delta_v(\omega)\right)\end{aligned} \tag{37}$$

jossa ensimmäisessä yhtäläisyydessä käytetään Taylor-sarjalaajennusta $(1+x)^{-1} \simeq 1-x$ ja toisessa $\sqrt{1+x} \simeq 1+x/2$. Tällöin voidaan laskea PSD-virhe. Sovittamalla kaava (37) kaavoihin (9)-(10) saadaan, jättämällä huomioitta korkeammat kuin ensimmäisen asteen poikkeamat $\hat{H}^2_{ML}(\omega)$:n laajennuksessa

$$\begin{aligned}\tilde{\Phi}_s(\omega) &\simeq \frac{1}{4}\left(1+\sqrt{\frac{\Phi_s(\omega)}{\Phi_x(\omega)}}\right)^2\Phi_x(\omega)-\Phi_s(\omega) \\ &\quad +\frac{1}{4}\left(1+\sqrt{\frac{\Phi_x(\omega)}{\Phi_s(\omega)}}\right)\left(\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega)-\Delta_v(\omega)\right)\end{aligned} \tag{38}$$

Kaavasta (38) seuraa, että

$$E[\tilde{\Phi}_s(\omega)] \simeq \frac{1}{4}\left(1+\sqrt{\frac{\Phi_s(\omega)}{\Phi_x(\omega)}}\right)^2 \Phi_x(\omega) - \Phi_s(\omega)$$
$$= \frac{1}{2}\Phi_v(\omega) - \frac{1}{4}\left(\sqrt{\Phi_x(\omega)} - \sqrt{\Phi_s(\omega)}\right)^2 \tag{39}$$

jossa toisessa yhtäläisyydessä käytetään kaavaa (2). Edelleen

$$\mathrm{Var}(\tilde{\Phi}_s(\omega)) \simeq \frac{1}{16}\left(1+\sqrt{\frac{\Phi_x(\omega)}{\Phi_s(\omega)}}\right)^2 \gamma\Phi_v^2(\omega) \tag{40}$$

LIITE D

$\hat{H}_{IPS}(\omega)$:n johtaminen

Kun $\Phi_x(\omega)$ ja $\Phi v(\omega)$ tunnetaan tarkasti, neliöidy PSD-virhe minimoimaan $H_{PS}(\omega)$:llä, eli $\hat{H}_{PS}(\omega)$ $\Phi_X(\omega)$:n ja $\Phi_V(\omega)$:n kanssa korvataan $\Phi_x(\omega)$:llä ja $\Phi_V(\omega)$:llä vastaavasti. Tämä tosiasia seuraa suoraan kaavoista (9) ja (10), eli $\Phi_S(\omega) = [H^2(\omega)\Phi_x(\omega) - \Phi_S(\omega)]^2 = 0$, jossa (2) käytetään viimeisessä yhtäläisyydessä. Huomaa, että tässä tapauksessa $H(\omega)$ on deterministinen suure, kun toisaalta $\hat{H}(\omega)$ on stokastinen suure. Huomioimalla PSD- estimaattien epävarmuus, tämä tosiasia ei yleisesti ottaen enää päde, ja tässä osassa johdetaan datariippumaton painotusfunktio $\hat{H}_{PS}(\omega)$:n suorituksen parantamiseksi. Muodostetaan varianssilauseke

$$\mathrm{Var}(\bar{\Phi}_s(\omega)) \simeq \xi\gamma\Phi_v^2(\omega) \qquad (41)$$

($\xi = 1$ PS:lle ja $\xi = (1-\sqrt{1+\mathrm{SNR}})^2$ MS:llä ja $\gamma = \gamma_x + \gamma_v$) . Muuttuja γ riippuu ainoastaan PSD-estimointimenetelmästä, jota käytetään, ja siihen ei vaikuta siirtofunktion $\hat{H}(\omega)$:n valinta. Ensimmäinen kerroin ξ riippuu kuitenkin $\hat{H}(\omega)$:n valinnasta. Tässä osassa etsitään datariippumaton painotusfunktio $\bar{G}(\omega)$ siten, että $\hat{H}(\omega) = \sqrt{\bar{G}(\omega)}\,\hat{H}_{PS}(\omega)$ minimoi neliöiden PSD-virheen oletusta, eli

$$\bar{G}(\omega) = \arg\min_{G(\omega)} E[\bar{\Phi}_s(\omega)]^2$$
$$\bar{\Phi}_s(\omega) = G(\omega)\hat{H}_{PS}^2(\omega)\Phi_x(\omega) - \Phi_s(\omega) \qquad (42)$$

Kaavassa (42), $G(\omega)$ on geneerinen painotusfunktio. Ennen kuin jatkamme, huomaa, että mikäli sallitaan, että painotusfunktio $G(\omega)$ on datariippuvainen, saadaan yleinen luokka spektrivähennysmenetelmiä, jotka sisältävät erityistapauksina useita yleisesti käytettyjä menetelmiä, esim. suuruusvähennystä käyttäen $G(\omega) = \hat{H}_{MS}^2(\omega)/\hat{H}_{PS}^2(\omega)$. Tällä havainnolla ei kuitenkaan ole suurta merkitystä, koska kaavan (42) optimointi

datariippuvalla G(ω):llä suuresti riippuu G (ω):n muodosta. Siksi menetelmät, jotka käyttävät datariippuvaista painotusfunktiota on analysoitava yksitellen, koska yleistä tulosta ei voida johtaa sellaisessa tapauksessa.

Kaavan (42) minimoimiseksi saadaan suoralla laskennalla

$$\begin{aligned}\bar{\Phi}_s(\omega) \;&\simeq\; (G(\omega)-1)\,\Phi_s(\omega) \\ &+G(\omega)\left(\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega)-\Delta_v(\omega)\right)\end{aligned} \tag{43}$$

Olettamalla neliöity PSD-virhe ja käyttämällä kaavaa (41) saadaan:

$$E[\bar{\Phi}_s(\omega)]^2 \simeq (G(\omega)-1)^2\Phi_s^2(\omega)+G^2(\omega)\,\gamma\,\Phi_v^2(\omega) \tag{44}$$

Kaavassa (44) G(ω) on korotettu toiseen potenssiin ja se voidaan analyyttisesti minimoida. Näin saadaan:

$$\begin{aligned}\bar{G}(\omega) &= \frac{\Phi_s^2(\omega)}{\Phi_s^2(\omega)+\gamma\Phi_v^2(\omega)} \\ &= \frac{1}{1+\gamma\left(\frac{\Phi_v(\omega)}{\Phi_x(\omega)-\Phi_v(\omega)}\right)^2}\end{aligned} \tag{45}$$

jossa toisessa yhtäläisyydessä käytetään kaavaa (2). Ei ole yllättävää, että $\bar{G}(\omega)$ riippuu (tuntemattomista) PSD:istä ja muuttujasta y. Kuten mainittiin yllä, ei voida suoraan korvata tuntemattomia PSD:itä kaavassa (45) vastaavilla estimaateilla ja väittää, että näin saatu modifioitu PS-menetelmä on optimaalinen, eli minimoi kaavaa (42). Voidaan kuitenkin olettaa, että ottamalla huomioon $\Phi_X(\omega)$:n ja $\Phi_V(\omega)$:n epävarmuusproseduurissa, modifioitu PS-menetelmä suoriutuu "paremmin" kuin vakio-PS. Johtuen ylläsanotusta modifioitua PS-menetelmää kutsutaan parannetuksi PS-menetelmäksi (IPS, Improved Power

Subtraction). Ennen IPS-menetelmän analysointia liitteessä E haluamme huomauttaa seuraavaa.

Korkeilla, hetkellisillä SNR-arvoilla (ω-arvolla siten, että $\Phi_S(\omega)/\Phi_V(\omega)$ >>1) seuraa kaavasta (45), että $\bar{G}(\omega) \simeq 1$, ja johtuen siitä, että normalisoitu virhevarianssi $\mathrm{Var}(\Phi_S(\omega))/\Phi_s^2(\omega)$ kts. kaava (41), on pieni tässä tapauksessa, voidaan päätellä, että IPS:n suoritus on lähellä vakio-PS:n suoritusta. Toisaalta, pienillä, hetkellisillä SNR:llä (ω siten, että $\gamma\Phi_v^2(\omega) \gg \Phi_s^2(\omega)$), $\bar{G}(\omega) \approx \Phi_s^2(\omega)/(\gamma\Phi_v^2(\omega))$, saadaan, kts. esim. kaava (43)

$$E[\tilde{\Phi}_s(\omega)] \approx -\Phi_s(\omega) \tag{46}$$

ja

$$\mathrm{Var}(\tilde{\Phi}_s(\omega)) \approx \frac{\Phi_s^4(\omega)}{\gamma\Phi_v^2(\omega)} \tag{47}$$

Kuitenkin pienellä SNR:llä ei voida päätellä, että kaavat (46)-(47) ovat edes lähimain voimassa mikäli $\bar{G}(\omega)$ kaavassa (45) korvataan $\hat{G}(\omega)$:llä, eli korvataan kaavassa $\Phi_x(\omega)$ ja $\Phi_V(\omega)$ niiden estimoiduilla arvoilla $\hat{\Phi}_x(\omega)$ ja vastaavasti $\hat{\Phi}_V(\omega)$.

LIITE E

$\hat{H}_{IPS}(\omega)$:n analysointi

Tässä liitteessä analysoidaan IPS-menetelmää. Määritetään $\bar{G}(\omega)$ kaavan (45) mukaan, ja siten, että $\Phi_v(\omega)$ ja $\Phi_x(\omega)$ korvataan siinä vastaavilla estimoiduilla suureilla. Voidaan osoittaa, että

$$\begin{aligned}\bar{\Phi}_s(\omega) \simeq\ & (\bar{G}(\omega)-1)\Phi_s(\omega) \\ & +\bar{G}(\omega)\left(\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega)-\Delta_v(\omega)\right) \\ & \times\left(\bar{G}(\omega)+\gamma\Phi_v(\omega)\frac{\Phi_v(\omega)+2\Phi_x(\omega)}{\Phi_s^2(\omega)+\gamma\Phi_v^2(\omega)}\right)\end{aligned} \tag{48}$$

joka voidaan verrata kaavaan (43). Tällöin

$$E[\bar{\Phi}_s(\omega)] \simeq (\bar{G}(\omega)-1)\Phi_s(\omega) \tag{49}$$

ja

$$\begin{aligned}\mathrm{Var}(\bar{\Phi}_s(\omega)) \simeq\ & \bar{G}^2(\omega) \\ & \times\left(\bar{G}(\omega)+\gamma\Phi_v(\omega)\frac{\Phi_v(\omega)+2\Phi_x(\omega)}{\Phi_s^2(\omega)+\gamma\Phi_v^2(\omega)}\right)^2\gamma\Phi_v^2(\omega)\end{aligned} \tag{50}$$

Korkeilla SNR-arvoilla siten, että $\Phi_S(\omega)/\Phi_V(\omega)$ >>1, voidaan tarkastella kaavoja (49)-(50). Täss tapauksessa voidaan johtaa

$$E[\bar{\Phi}_s(\omega)] \simeq 0 \tag{51}$$

ja

$$\mathrm{Var}(\bar{\Phi}_s(\omega)) \simeq \left(1+4\gamma\frac{\Phi_v(\omega)}{\Phi_s(\omega)}\right)\gamma\Phi_v^2(\omega) \tag{52}$$

Poisjätetyt termit kaavoissa (51) ja (52) ovat suuruusluokkaa $0((\Phi_V(\omega)/\Phi_S(\omega))^2)$. Täten, kuten jo mainittiin, IPS:n suoritus vastaa PS:n suoritusta korkealla SNR:llä. Toisaalta pienellä SNR:llä (ω siten, että $\Phi_s^2(\omega)/(\gamma\Phi_v^2(\omega)) \ll 1$), $\bar{G}(\omega) \simeq \Phi_s^2(\omega)/(\gamma\Phi_v^2(\omega))$, ja

$$E[\tilde{\Phi}_s(\omega)] \simeq -\Phi_s(\omega) \tag{53}$$

ja

$$\mathrm{Var}(\tilde{\Phi}_s(\omega)) \simeq 9\,\frac{\Phi_s^4(\omega)}{\gamma\Phi_v^2(\omega)} \tag{54}$$

Vertaamalla kaavoja (53)-(54) vastaaviin PS-tuloksiin (13) ja (16) voidaan todeta, että pienellä hetkellisillä SNR-arvolla IPS-menetelmä vähentää olennaisesti $\Phi_S(\omega)$:n varianssia verrattuna PS-menetelmään pakottamalla $\hat{\Phi}_S(\omega)$:tä kaavassa (9) kohti nollaa. Erityisesti IPS:n ja PS-varianssien välinen suhde on luokkaa $O(\Phi_s^4(\omega)/\Phi_v^4(\omega))$. Voidaan myös verrata kaavoja (53)-(54) likimääräiseen lausekkeeseen (47), ja todeta, että niiden välinen suhde on yhtä kuin 9.

LIITE F

PS optimaalisella vähennyskertoimella δ

Usein käytetty tehovähennyksen modifiointimenetelmä on:

$$\hat{H}_{\delta PS}(\omega) = \sqrt{1 - \delta(\omega)\frac{\hat{\Phi}_v(\omega)}{\hat{\Phi}_x(\omega)}} \qquad (55)$$

missä $\delta(\omega)$ on mahdollinen taajuusriippuvainen funktio. Erityisesti, kun $\delta(\omega) = \delta$ jollakin vakiolla $\delta > 1$, tätä menetelmää kutsutaan usein tehovähennykseksi ylivähennyksellä. Tämä modifiointi vähentää olennaisesti kohinatasoa ja vähentää äänitekotuotteita (engl. tonal artifacts). Lisäksi se väärentää puhetta olennaisesti, joka tekee modifioinnin käyttökelvottomaksi laadukkaassa puheen korostamisessa. Tämä seikka nähdään helposti kaavasta (55) kun $\delta >> 1$. Täten, keskisuurilla ja pienillä puhe/kohinasuhteilla (ω-alueella) lauseke neliöjuuren alla on usein hyvin negatiivinen ja tasasuuntaajalaite asettaa sen siksi usein nollaksi (puoliaaltotasasuuntaus), joka tarkoittaa, että ainoastaan taajuuskaistat, joissa SNR on korkea, esiintyvät antosignaalissa $\hat{s}(k)$ kaavassa (3). Johtuen ei-lineaarisesta tasasuuntauslaitteesta esillä oleva analyysimenetelmä ei ole suoraan sovellettavissa tähän tapaukseen, ja koska $\delta > 1$ johtaa antoon, jolla on huono äänilaatu, tätä modifiointia ei tarkemmin tutkita.

Mielenkiintoinen tapaus esiintyy kuitenkin kun $\delta(\omega) \leq 1$, joka ilmenee seuraavasta selostuksesta. Kuten mainittiin aikaisemmin, kun $\Phi_x(\omega)$ ja $\Phi_V(\omega)$ tunnetaan täsmälleen, kaava (55) jossa $\delta(\omega) = 1$ on optimaalinen neliöidyn PSD-virheen minimoimiseksi. Toisaalta, mikäli $\Phi_x(\omega)$ ja $\Phi_V(\omega)$ ovat täysin tuntemattomia, eli niistä ei ole käytettävissä estimaatteja, voidaan korkeintaan estimoida puhe itse kohinamittauksesta, eli $\hat{s}(k) = x(k)$, joka vastaa kaavan (55) soveltamista arvolla $\delta = 0$. Johtuen yllämainitusta kahdesta ääritapauksesta voidaan olettaa, että

kun tuntemattomat $\Phi_x(\omega)$ ja $\Phi_V(\omega)$ korvataan vastaavasti $\hat{\Phi}_x(\omega)$:llä ja $\hat{\Phi}_V(\omega)$:llä, virhe $E[\Phi_S(\omega)]^2$ minimoidaan jollakin $\delta(\omega)$-arvolla alueella $0 < \delta(\omega) < 1$.

Lisäksi empiirisessä suureessa, keskimääräistetty spektrivääristymäparaneminen, joka on samanlainen kuin PSD-virhe, tutkittiin kokeellisesti koskien MS:n vähennyskerrointa. Perustuen useisiin kokeisiin pääteltiin, että optimaalinen vähennyskerroin edullisesti on alueella 0,5 - 0,9.

Erityisesti PSD-virheen laskeminen tässä tapauksessa antaa:

$$\bar{\Phi}_s(\omega) \simeq (1-\delta(\omega))\Phi_v(\omega) + \delta(\omega)\left(\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega) - \Delta_v(\omega)\right) \tag{56}$$

Käyttämällä neliöidyn PSD-virheen oletusta saadaan:

$$E[\bar{\Phi}_s(\omega)]^2 \simeq (1-\delta(\omega))^2\,\Phi_v^2(\omega) + \delta^2\gamma\Phi_v^2(\omega) \tag{57}$$

jossa kaavaa (41) käytetään. Kaavassa (57) $\delta(\omega)$ esiintyy neliössä, ja se voidaan minimoida analyyttisesti. Merkitsemällä optimaalinen arvo $\bar{\delta}$:lla, saadaan:

$$\bar{\delta} = \frac{1}{1+\gamma} < 1 \tag{58}$$

Huomaa, että koska γ kaavassa (58) on likimain taajuusriippumaton (ainakin arvoilla N >> 1) myös $\bar{\delta}$ on taajuudesta riippumaton. Erityisesti $\bar{\delta}$ on riippumaton $\Phi_x(\omega)$:stä ja $\Phi_V(\omega)$:stä, joka tarkoittaa, että varianssi ja $\Phi_S(\omega)$:n suuntauma seuraa suoraan kaavasta (57).

$\bar{\delta}$:n arvo voi olla huomattavasti pienempi kuin jossain (realistisessa) tapauksessa. Esimerkiksi mikäli jälleen sijoitetaan $\gamma_v = 1/\tau$ ja $\gamma_x = 1$ saadaan:

$$\bar{\delta} = \frac{1}{2}\frac{1}{1+1/2\tau}$$

joka selvästi kaikilla τ-arvoilla on pienempi kuin 0,5. Tässä tapauksessa tosiasia $\bar{\delta} << 1$ osoittaa, että PSD-estimaattoreiden epävarmuus (ja erityisesti $\Phi_x(\omega)$:n epävarmuus) vaikuttavat suuresti annon laatuun (PSD-virheen termeinä). Erityisesti käyttämällä $\bar{\delta} << 1$ saadaan, että puhe/kohinasuhteen paraneminen ottosignaalista antosignaaliin, on pieni.

Tässä herää kysymys, onko olemassa, vastaavasti kuin liitteen D:n IPS-menetelmän painotusfunktiolla, datariippumaton painotusfunktio $\bar{G}(\omega)$. Liitteessä G johdetaan tällainen menetelmä (ja se merkitään δIPS).

LIITE G

$\hat{H}_{\delta IPS}(\omega)$:n johtaminen

Tässä liitteessä etsitään datariippumaton painotuskerroin $\bar{G}(\omega)$ siten, että $\hat{H}(\omega) = \sqrt{\bar{G}}(\omega)\hat{H}_{\delta IPS}(\omega)$ eräällä vakiolla $\delta\,(0\leq\delta\leq1)$ joka minimoi neliöidyn PSD-virheen oletuksen, kts. esim. kaava (42). Tästä saadaan:

$$\begin{aligned}\tilde{\Phi}_s(\omega) &= (G(\omega)-1)\Phi_s(\omega)+G(\omega)(1-\delta)\Phi_v(\omega)\\ &\quad G(\omega)\delta\left(\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega)-\Delta_v(\omega)\right)\end{aligned} \tag{59}$$

Neliöidyn PSD-virheen oletusarvo saadaan kaavasta:

$$\begin{aligned}E[\tilde{\Phi}_s(\omega)]^2 &= (G(\omega)-1)^2\Phi_s^2(\omega)+G^2(\omega)(1-\delta)^2\Phi_v^2(\omega)\\ &\quad 2(G(\omega)-1)\Phi_s(\omega)G(\omega)(1-\delta)\Phi_v(\omega)+G^2(\omega)\delta^2\gamma\Phi_v^2(\omega)\end{aligned} \tag{60}$$

Kaavan (60) oikealla puolella on $G(\omega)$ neliöitynä, ja se voidaan analyyttisesti minimoida. Näin saatu $\bar{G}(\omega)$ on:

$$\begin{aligned}\bar{G}(\omega) &= \\ &\frac{\Phi_s^2(\omega)+\Phi_s(\omega)\Phi_v(\omega)(1-\delta)}{\Phi_s^2(\omega)+2\Phi_s(\omega)\Phi_v(\omega)(1-\delta)+(1-\delta)^2\Phi_v^2(\omega)+\delta^2\gamma\Phi_v^2(\omega)}\\ &= \frac{1}{1+\beta\left(\frac{\Phi_v(\omega)}{\Phi_x(\omega)-\Phi_v(\omega)}\right)^2}\end{aligned} \tag{61}$$

jossa ß toisessa yhtäläisyydessä saadaan:

$$\beta = \frac{(1-\delta)^2+\delta^2\gamma+(1-\delta)\Phi_s(\omega)/\Phi_v(\omega)}{1+(1-\delta)\Phi_v(\omega)/\Phi_s(\omega)} \tag{62}$$

Kun $\delta = 1$, (61)-(62) yllä yksinkertaistavat IPS-menetelmää, (45), ja kun $\delta = 0$ päädymme vakio-PS:ään. Korvaamalla $\Phi_S(\omega)$ ja $\Phi_V(\omega)$ kaavoissa (61)-(62) vastaavilla estimoiduilla arvoilla $\hat{\Phi}_x(\omega) - \Phi_V(\omega)$ ja vastaavasti $\Phi_V(\omega)$, saadaan menetelmä, joka IPS-menetelmän näkökannalta merkitään δIPS. δIPS-menetelmän analyysi on vastaava kuin IPS-menetelmän analyysi, mutta se vaatii paljon vaivaa ja laskentaa, ja jätetään siksi huomioimatta.

VIITEJULKAISUJA

[1] S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-27, Huhtikuu 1979, ss. 113-120.

[2] J. S. Lim ja A. V. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech", Proceedings of the IEEE, Vol. 67, No. 12, Joulukuu 1979, ss. 1586-1604.

[3] J. D. Gibson, B. Koo ja S. D. Gray, "Filtering of Colored Noise for Speech Enhancement and Coding", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-39, No. 8, Elokuu 1991, ss. 1732-1742.

[4] J. H. L. Hansen ja M. A. Clements, "Constrained Iterative Speech Enhancement with Applicatioin to Speech Recognition", IEEE Transactions on Signal Processing, Vol. 39, No. 4, Huhtikuu 1991, ss. 795-805.

[5] D. K. Freeman, G. Cosier, C. B. Southcott ja I. Boid, "The Voice Activity Detector for the Pan-European Digital Cellular Mobile Telephone Service", 1989 IEEE International Conference Acoustics, Speech and Signal Processing, Glasgow, Scotland, 23-26 maaliskuuta 1989, ss. 369-372.

[6] PCT application WO 89/08910, British Telecommunications PLC.

Patenttivaatimukset

1. Spektrivähennys kohinan vaimennusmenetelmä kehyspohjaisessa digitaalisessa liikennöintijärjestelmässä, jossa kuhunkin kehykseen kuuluu ennaltamäärätty lukumäärä N ääninäytteitä, jolloin saadaan kullekin kehykselle N vapausastetta, jossa spektrivähennysfunktio $\hat{H}(\omega)$ perustuu ei-puhekehysten taustakohinan tehospektritiheyden estimaattiin $\hat{\Phi}_v(\omega)$ ja puhekehyksen tehospektritiheyden estimaattiin $\hat{\Phi}_x(\omega)$, t u n n e t t u siitä, että

kukin puhekehys approksimoidaan parametrisella mallilla, joka vähentää vapausasteiden lukumäärää arvoksi, joka on pienempi kuin N; ja

kunkin puhekehyksen tehospektritiheyden mainittu estimaatti $\hat{\Phi}_x(\omega)$ estimoidaan parametrisella tehospektriestimointimenetelmällä perustuen approksimatiiviseen parametrimalliin

kunkin ei-puhekehyksen tehospektritiheyden mainittu estimaatti $\hat{\Phi}_v(\omega)$ estimoidaan ei-parametrisella tehospektriestimointimenetelmällä.

2. Patenttivaatimuksen 1 mukainen menetelmä, t u n n e t t u siitä, että approksimatiivinen parametrinen malli on autoregressiivinen (AR) malli.

3. Patenttivaatimuksen 2 mukainen menetelmä, t u n n e t t u siitä, että autoregressiivinen (AR) malli on likimain kertalukua $\sqrt{N}$.

4. Patenttivaatimuksen 3 mukainen menetelmä, t u n n e t t u siitä, että autoregressiivinen (AR) malli on likimain kertalukua 10.

5. Patenttivaatimuksen 3 mukainen menetelmä, t u n n e t t u spektrivähennysfunktiosta $\hat{H}(\omega)$ seuraavan kaavan mukaisesti:

$$\hat{H}(\omega) = \sqrt{\hat{G}(\omega)\left(1 - \delta(\omega)\frac{\hat{\Phi}_v(\omega)}{\hat{\Phi}_x(\omega)}\right)}$$

jossa $\hat{G}(\omega)$ on painotusfunktio ja $\delta(\omega)$ on vähennyskerroin.

6. Patenttivaatimuksen 5 mukainen menetelmä, t u n n e t t u siitä, että $\hat{G}(\omega) = 1$.

7. Patenttivaatimuksen 5 tai 6 mukainen menetelmä, t u n n e t t u siitä, että $\delta(\omega)$ on vakio ≤ 1.

8. Patenttivaatimuksen 3 mukainen menetelmä, t u n n e t t u spektrivähennysfunktiosta $\hat{H}(\omega)$ seuraavan kaavan mukaisesti:

$$\hat{H}(\omega) = 1 - \sqrt{\frac{\hat{\Phi}_v(\omega)}{\hat{\Phi}_x(\omega)}}$$

9. Patenttivaatimuksen 3 mukainen menetelmä, t u n n e t t u spektrivähennysfunktiosta $\hat{H}(\omega)$ seuraavan kaavan mukaisesti:

$$\hat{H}(\omega) = \left(1 - \frac{\hat{\Phi}_v(\omega)}{\hat{\Phi}_x(\omega)}\right)$$

10. Patenttivaatimuksen 3 mukainen menetelmä, t u n n e t t u siitä, että spektrivähennysfunktiosta $\hat{H}(\omega)$ seuraavan kaavan mukaisesti:

$$\hat{H}(\omega) = \frac{1}{2}\left(1 + \sqrt{\left(1 - \frac{\hat{\Phi}_v(\omega)}{\hat{\Phi}_x(\omega)}\right)}\right)$$

Patentkrav

1. Spektrumreduktion i en brusdämpningsmetod i ett rambaserat digitalt kommunikationssystem, där i var ram ingår ett i förväg bestämt antal N ljudprov, varvid man för var ram får N frihetsgrader, där spektrumreduktionsfunktionen $\hat{H}(\omega)$ är baserad på ett estimat $\Phi_v(\omega)$ av effektspektrumsdensiteten i bakgrundsbruset av icke-ljudramar och ett estimat $\Phi_x(\omega)$ av effektspektrumsdensiteten i ljudramen, **kännetecknad** av att

var ljudram approximeras med en parametrisk modell, som reducerar antalet frihetsgrader till ett värde, som är mindre än N; och

estimatet $\Phi_x(\omega)$ av effektspektrumsdensiteten i var ljudram estimeras med en parametrisk effektspektrumsestimeringsmetod baserad på en approximativ parametermodell

estimatet $\Phi_v(\omega)$ av effektspektrumsdensiteten i var icke-ljudram estimeras med en icke-parametrisk effektspektrumsestimeringsmetod.

2. En metod enligt patentkrav 1, **kännetecknad** av att den approximativa parametriska modellen är en autoregressiv (AR) modell.

3. En metod enligt patentkrav 2, **kännetecknad** av att den autoregressiva (AR) modellen är av cirka ordningen $\sqrt{N}$.

4. En metod enligt patentkrav 3, **kännetecknad** av att den autoregressiva (AR) modellen är av cirka ordningen 10.

5. En metod enligt patentkrav 3, **kännetecknad** av spektrumreduktionsfunktionen $\hat{H}(\omega)$ enligt följande formel:

$$\hat{H}(\omega) = \sqrt{\hat{G}(\omega)\left(1 - \delta(\omega)\frac{\hat{\Phi}_v(\omega)}{\hat{\Phi}_x(\omega)}\right)}$$

där $\hat{G}(\omega)$ är en vägningsfunktion och $\delta(\omega)$ är en reduktionsfaktor.

6. En metod enligt patent krav 5, **kännetecknad** av att $\hat{G}(\omega) = 1$.

7. En metod enligt patent krav 5 eller 6, **kännetecknad** av att $\delta(\omega)$ är en konstant ≤ 1.

8. En metod enligt patent krav 3, **kännetecknad** av spektrumreduktionsfunktionen $\hat{H}(\omega)$ enligt följande formel:

$$\hat{H}(\omega) = 1 - \sqrt{\frac{\hat{\Phi}_v(\omega)}{\hat{\Phi}_x(\omega)}}$$

9. En metod enligt patent krav 3, **kännetecknad** av spektrumreduktionsfunktionen $\hat{H}(\omega)$ enligt följande formel:

$$\hat{H}(\omega) = \left(1 - \frac{\hat{\Phi}_v(\omega)}{\hat{\Phi}_x(\omega)}\right)$$

10. En metod enligt patent krav 3, **kännetecknad** av spektrumreduktionsfunktionen $\hat{H}(\omega)$ enligt följande formel:

$$\hat{H}(\omega) = \frac{1}{2}\left(1 + \sqrt{\left(1 - \frac{\hat{\Phi}_v(\omega)}{\hat{\Phi}_x(\omega)}\right)}\right)$$

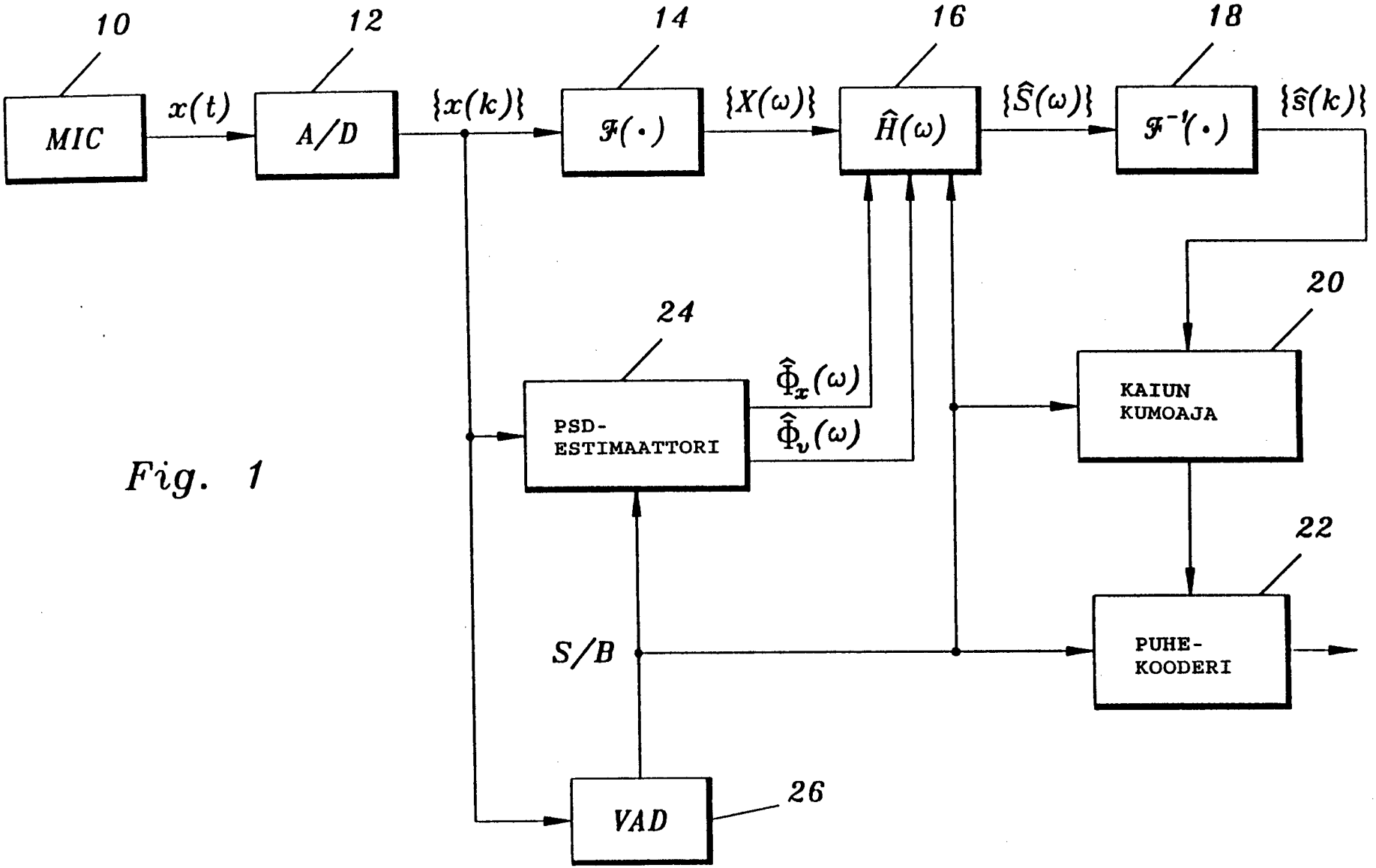
10
12
14
16
18
MIC
x(t)
A/D
{x(k)}
$\mathcal{F}(\cdot)$
{X(ω)}
$\hat{H}(\omega)$
{$\hat{S}(\omega)$}
$\mathcal{F}^{-1}(\cdot)$
{$\hat{s}(k)$}
20
24
$\hat{\Phi}_x(\omega)$
$\hat{\Phi}_v(\omega)$
PSD-
ESTIMAATTORI
KAIUN
KUMOAJA
22
S/B
PUHE-
KOODERI
VAD
26

Fig. 1

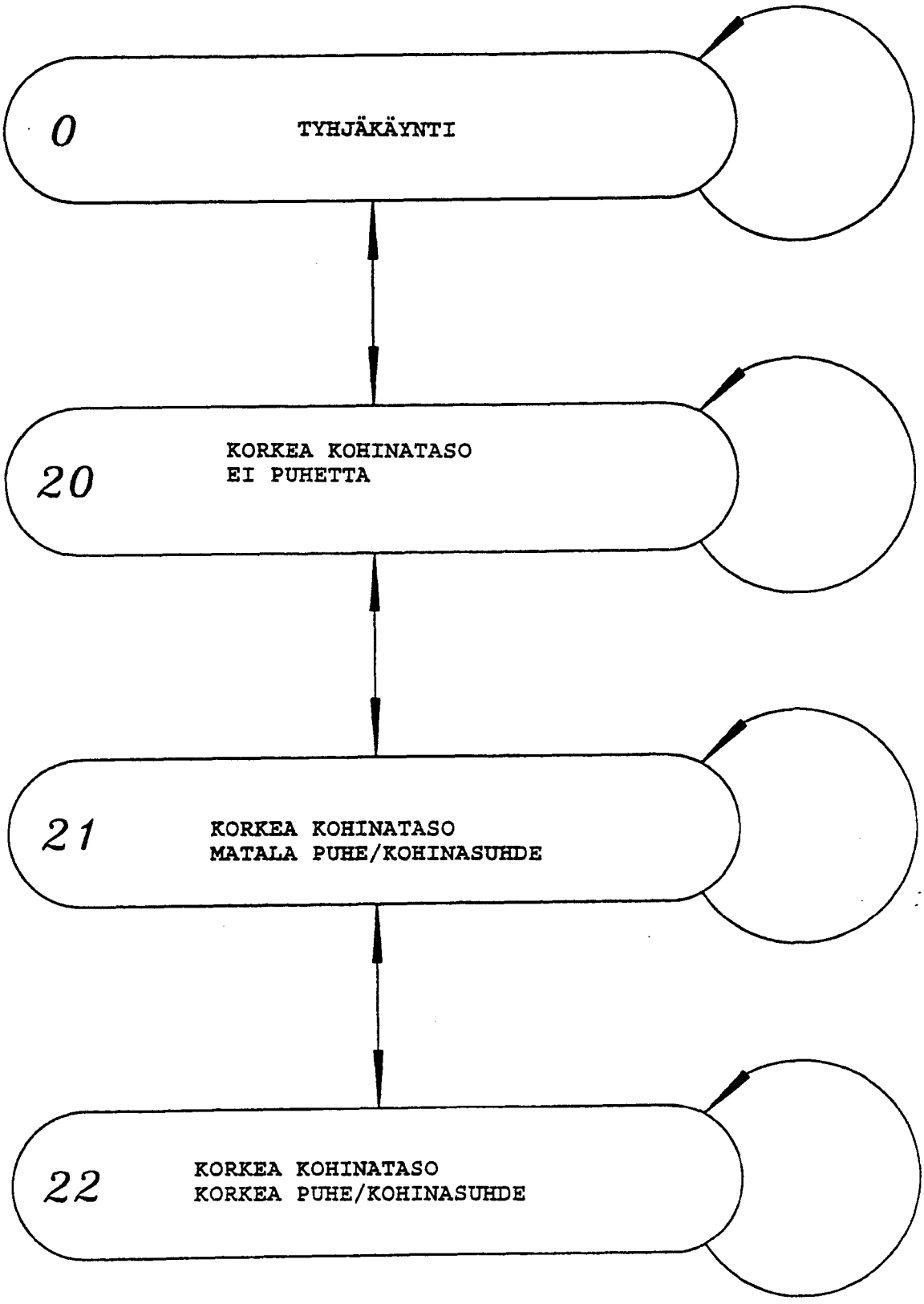
0
TYHJÄKÄYNTI
20
KORKEA KOHINATASO
EI PUHETTA
21
KORKEA KOHINATASO
MATALA PUHE/KOHINASUHDE
22
KORKEA KOHINATASO
KORKEA PUHE/KOHINASUHDE

Fig. 2

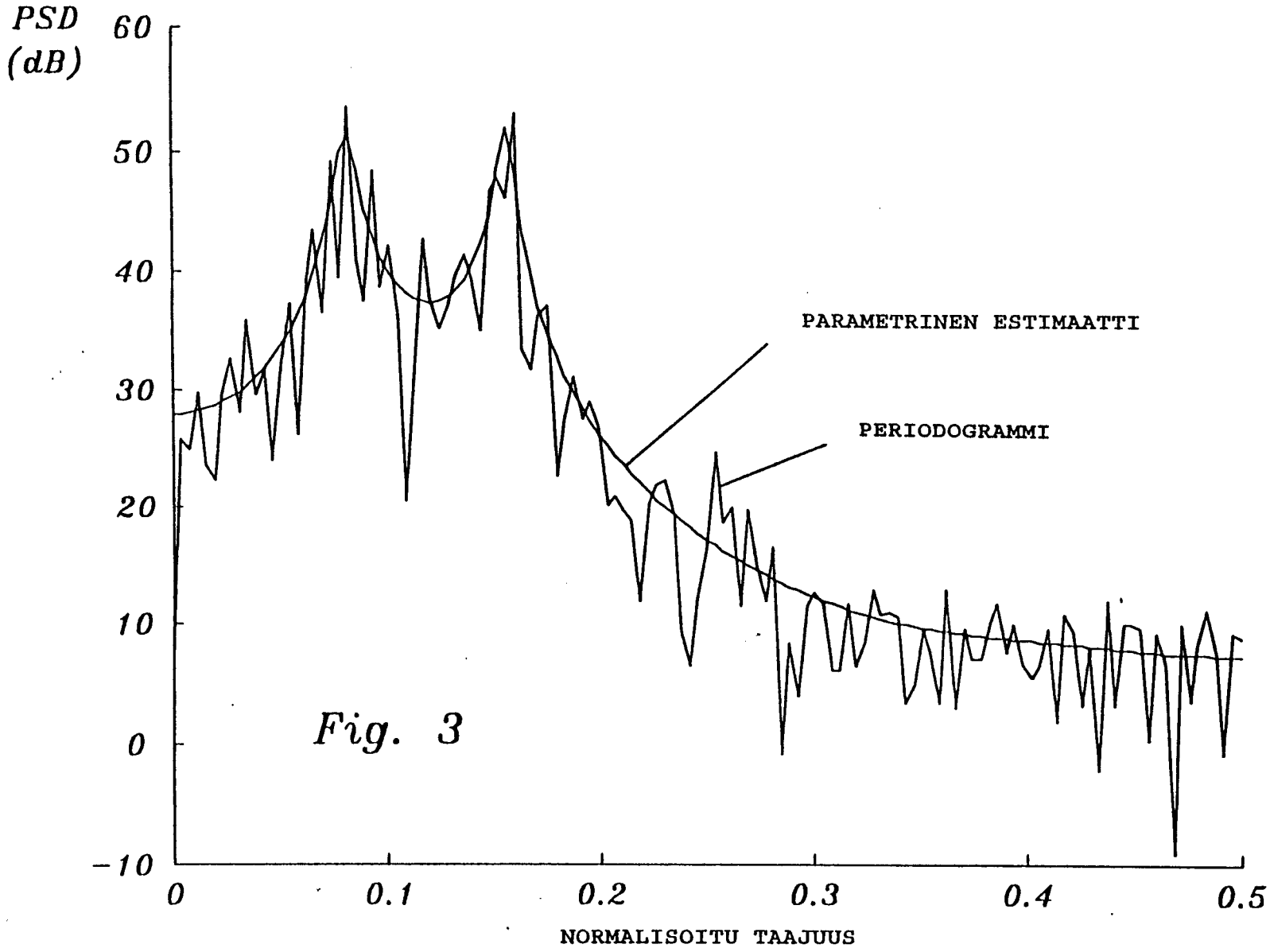
PSD
(dB)
60
50
40
30
20
10
0
-10
0
0.1
0.2
0.3
0.4
0.5
NORMALISOITU TAAJUUS
PARAMETRINEN ESTIMAATTI
PERIODOGRAMMI

Fig. 3

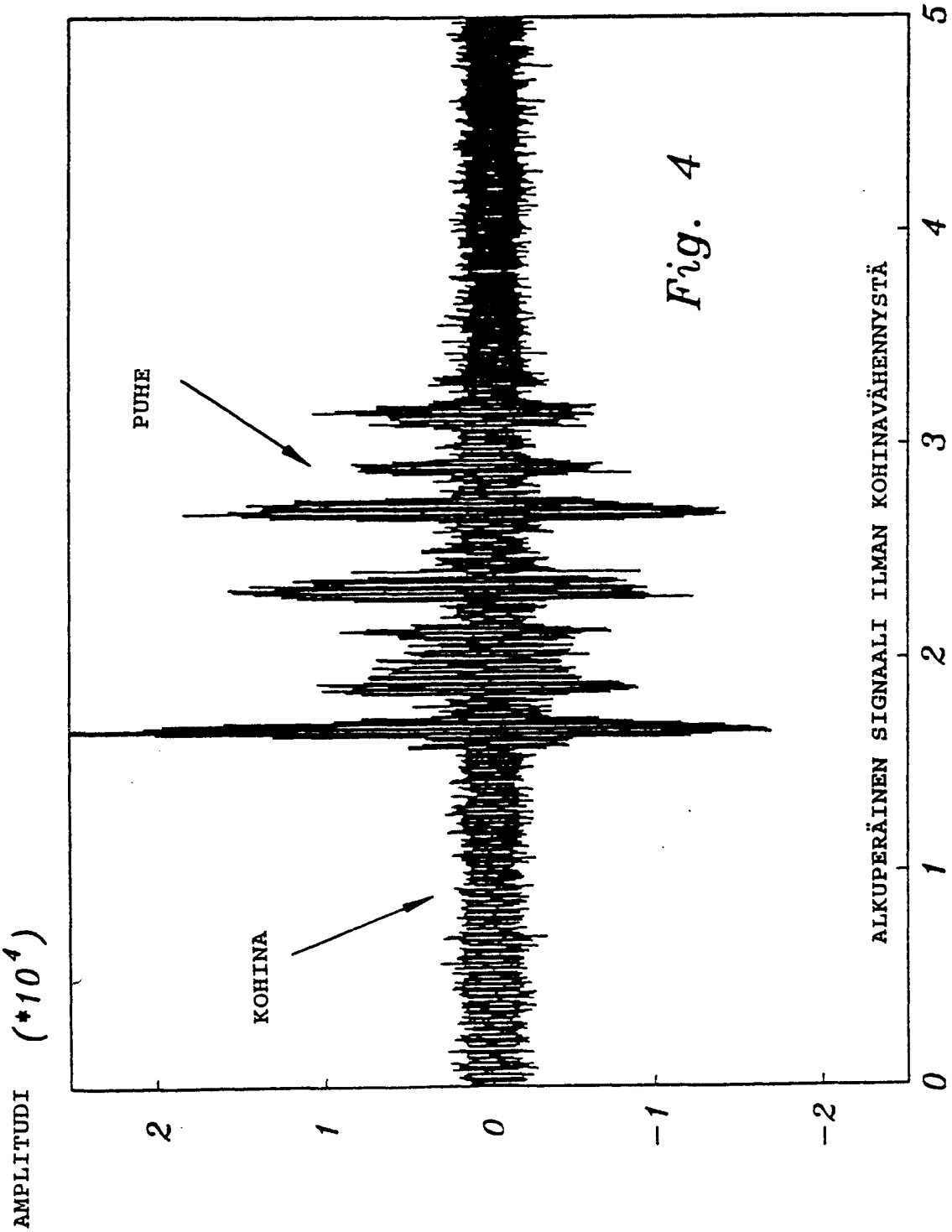
AMPLITUDI (*10^4)
2
1
0
−1
−2
KOHINA
PUHE
ALKUPERÄINEN SIGNAALI ILMAN KOHINAVÄHENNYSTÄ
0
1
2
3
4
5
TIME (s)

Fig. 4

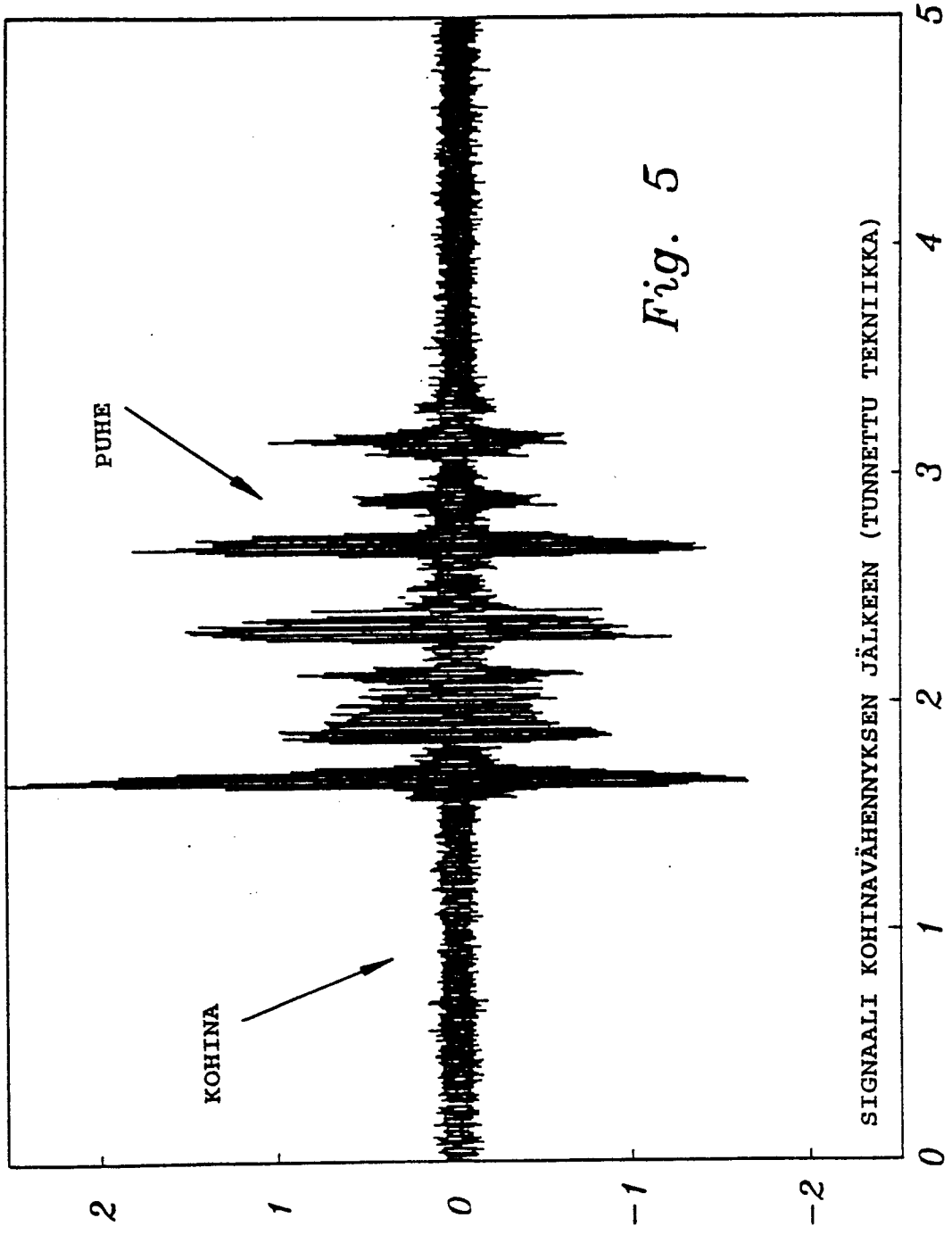
AMPLITUDI
(*10^4)
2
1
0
-1
-2
KOHINA
PUHE
0
1
2
3
4
5
AIKA (s)
SIGNAALI KOHINAVÄHENNYKSEN JÄLKEEN (TUNNETTU TEKNIIKKA)

Fig. 5

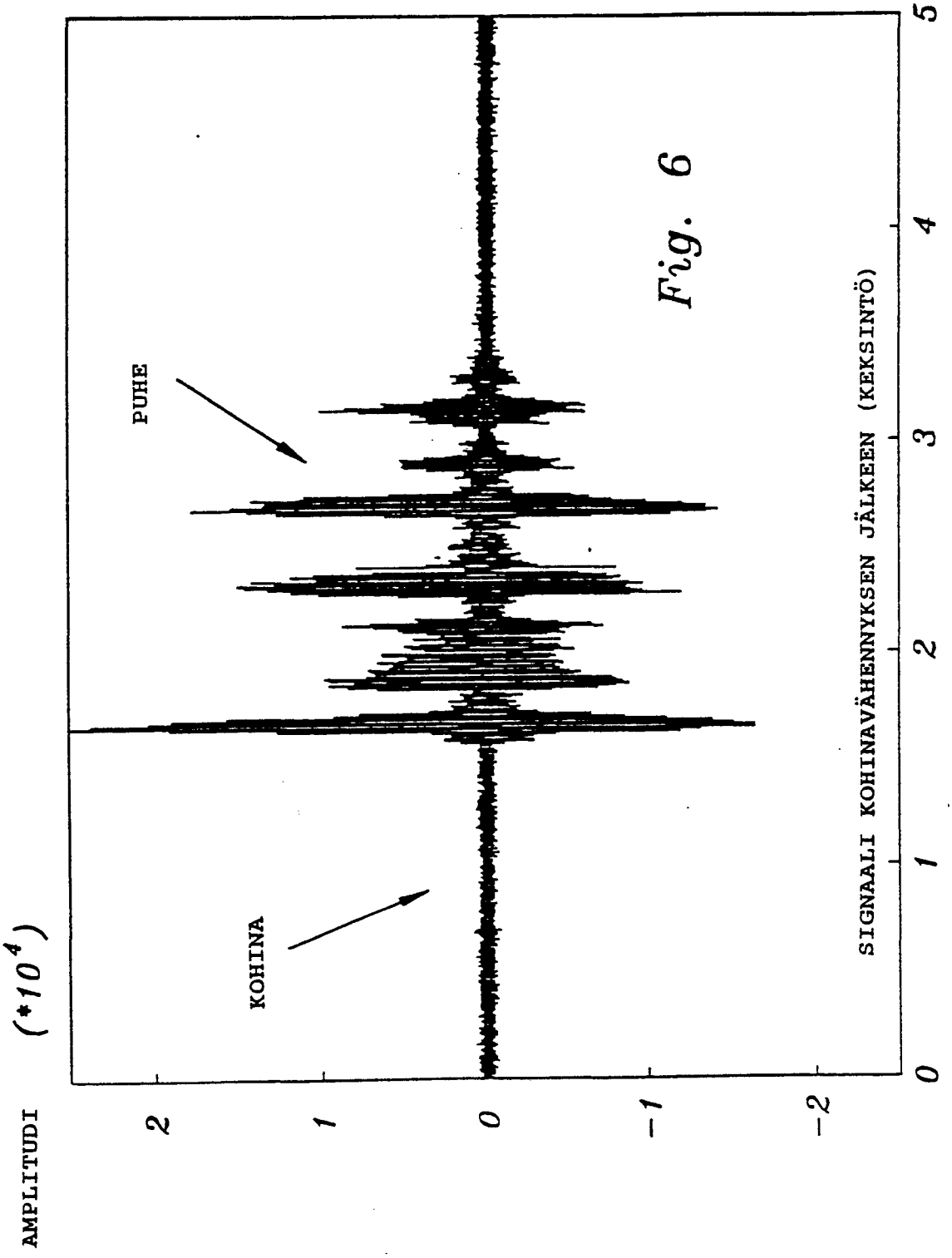
AMPLITUDI (*10^4)
2
1
0
-1
-2
KOHINA
PUHE
0
1
2
3
4
5
SIGNAALI KOHINAVÄHENNYKSEN JÄLKEEN (KEKSINTÖ)
AIKA (s)

Fig. 6

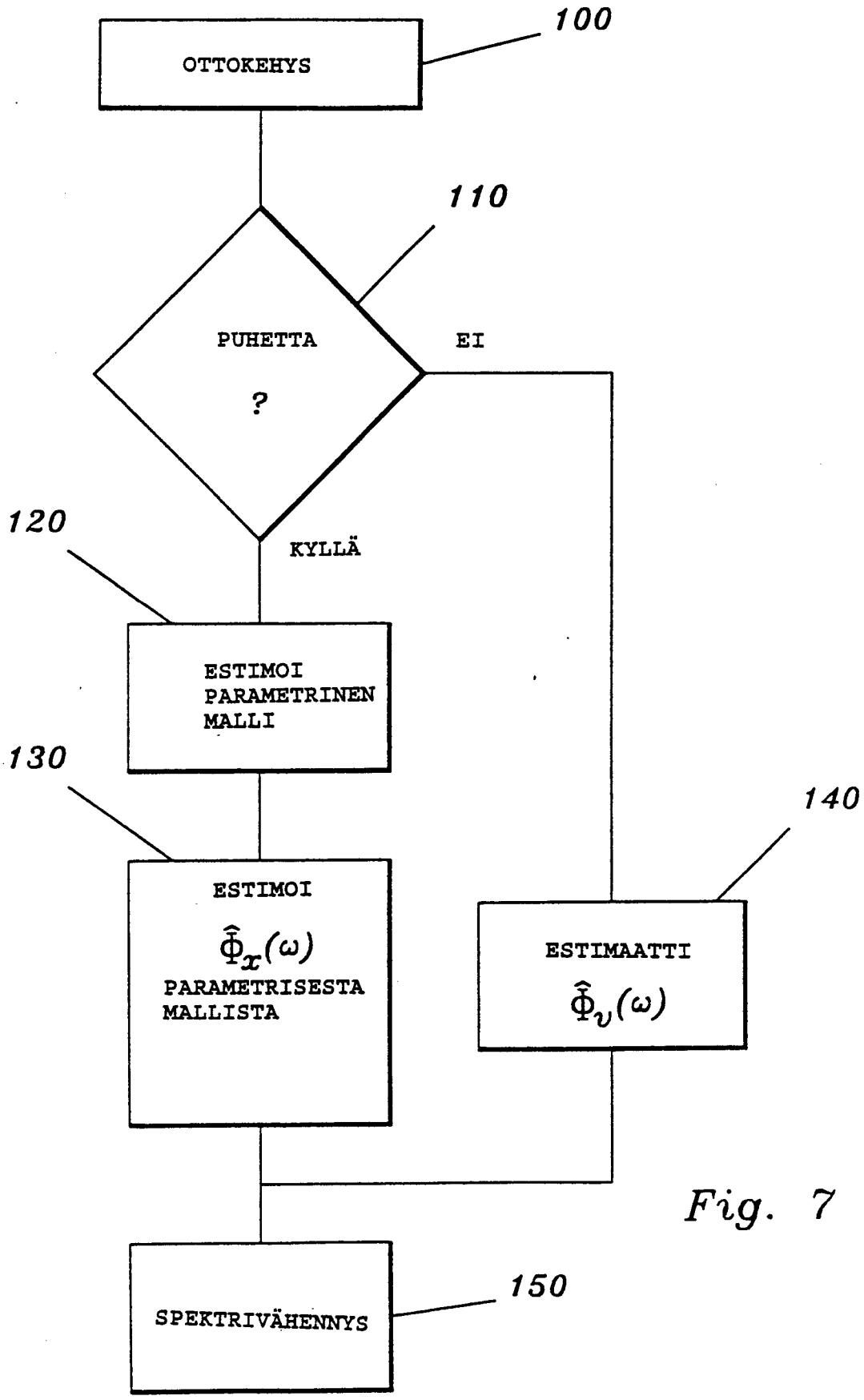
100
OTTOKEHYS
110
PUHETTA
?
EI
KYLLÄ
120
ESTIMOI
PARAMETRINEN
MALLI
130
ESTIMOI
$\hat{\Phi}_x(\omega)$
PARAMETRISESTA
MALLISTA
140
ESTIMAATTI
$\hat{\Phi}_v(\omega)$
150
SPEKTRIVÄHENNYS

Fig. 7

PATENTTI- JA REKISTERIHALLITUS
Patentti- ja innovaatiolinja
PL 1160
00101 Helsinki

TUTKIMUSRAPORTTI

PATENTTIHAKEMUS NRO	LUOKITUS, IPC7
FI973142	G10L 21/02

TUTKITUT PATENTTILUOKAT (luokitusjärjestelmät ja luokkatiedot)
G10L

TUTKIMUKSESSA KÄYTETYT TIETOKANNAT

VIITEJULKAISUT

Kategoria*)	Julkaisun tunnistetiedot ja tiedot sen olennaisista kohdista	Koskee vaatimuksia
A	EP0588526 A 19940323 NOKIA MOBILE PHONES LTD (FI)	
A	WO9418666 A 19940818 CHEETHAM BARRY MICHAEL GEORGE (GB);	

Jatkuu seuraavalla sivulla ☒

*) X Julkaisu, jonka perusteella keksintö ei ole uusi tai ei eroa olennaisesti ennestään tunnetusta tekniikasta.
Y Julkaisu, jonka perusteella keksintö ei eroa olennaisesti ennestään tunnetusta tekniikasta, kun otetaan huomioon tämä ja yksi tai useampi samaan kategoriaan kuuluva julkaisu yhdessä.
A Yleistä tekniikan tasoa edustava julkaisu.

O Tullut julkiseksi esitelmän välityksellä, hyväksikäyttämällä tai muutoin muun kuin kirjoituksen avulla.
P Julkaistu ennen hakemuksen tekemispäivää mutta ei ennen aikaisinta etuoikeuspäivää.
T Julkaistu hakemuksen tekemispäivän tai etuoikeuspäivän jälkeen ja valaisee keksinnön periaatetta tai teoreettista taustaa.
E Aikaisempi suomalainen tai Suomea koskeva patentti- tai hyödyllisyysmallihakemus, joka on tullut julkiseksi hakemuksen tekemispäivänä (etuoikeuspäivänä) tai sen jälkeen.
D Julkaisu, joka on mainittu hakemuksessa.
L Julkaisu, joka kyseenalaistaa etuoikeuden, osoittaa toisen julkaisun julkaisupäivämäärän tai johon viitataan jostakin muusta syystä.

& Samaan patenttiperheeseen kuuluva julkaisu.

Lisätietoja liitteessä ☐

Päiväys	Tutkijainsinööri
16.11.2003	Janne Nummela