



(12) **PATENT**

(11) **345590**

(13) **B1**

NORGE

(19) NO

(51) Int Cl.

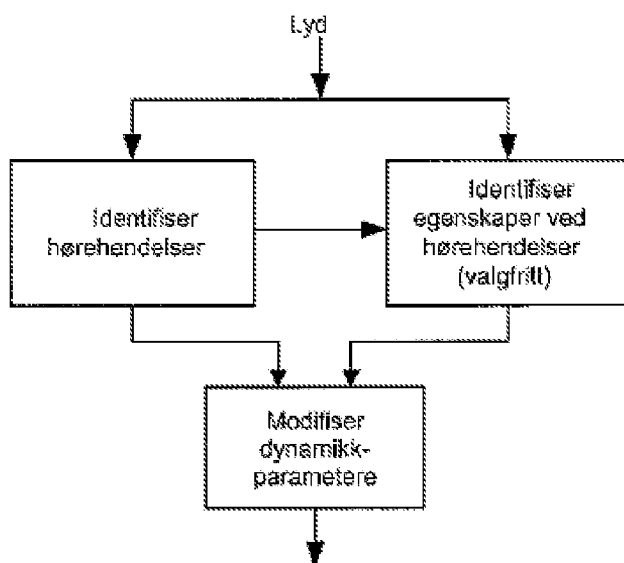
H03G 3/30 (2006.01)

H03G 7/00 (2006.01)

Patentstyret

(21)	Søknadsnr	20191310	(86)	Int.inng.dag og søknadsnr	2007.03.30 PCT/US2007/008313
(22)	Inng.dag	2019.11.01	(85)	Videreføringsdag	2019.11.01
(24)	Løpedag	2007.03.30	(30)	Prioritet	2006.04.27, US, 60/795,808
(41)	Alm.tilgj	2008.11.17			
(45)	Meddelt	2021.05.03			
(62)	Avdelt fra	20190002, 2019.01.07			
(73)	Innehaver	Dolby Laboratories Licensing Corporation, 1275 Market Street, CA94103-1410 SAN FRANCISCO, USA			
(72)	Oppfinner	Brett Graham Crockett, c/o Dolby Laboratories Licensing Corp, 1275 Market Street, CA94103-1410 SAN FRANCISCO, USA Alan Jeffrey Seefeldt, c/o Dolby Laboratories Licensing Corp, 1275 Market Street, CA94103-1410 SAN FRANCISCO, USA			
(74)	Fullmektig	BRYN AARFLOT AS, Stortingsgata 8, 0161 OSLO, Norge			
(54)	Benevnelse	Audioforsterkningsregulering ved bruk av spesifikk lydstyrkebasert hørehendelsesdeteksjon			
(56)	Anførte publikasjoner	US 2004165730 A1, US 2006029239 A1, US 2004122662 A1			
(57)	Sammendrag				

I noen utførelsesformer er en fremgangsmåte for prosessering av et lydsignal i et lydbehandlingsapparat beskrevet. Fremgangsmåten inkluderer mottak av et lydsignal og en parameter, parameteren indikerer et sted for en hørselshendelsesgrense. En lyddel mellom påfølgende hørselshendelsesgrenser utgjør en hørselshendelse. Fremgangsmåten inkluderer videre anvendelse av en modifisering på lydsignalet delvis basert på en forekomst av hørselshendelsesgrensen. Parameteren kan genereres ved å overvåke en karakteristikk av lydsignalet og ved å identifisere en endring i karakteristikken.



Audioforsterkningsregulering ved bruk av spesifikk lydstyrkebasert hørehendelsesdeteksjon

Teknisk område

5 Foreliggende oppfinnelse vedrører fremgangsmåter og en anordning for
audiodynamisk spektrumkontroll der en audiobehandlingsanordning analyserer et
audiosignal og endrer nivået til, forsterkningen av eller det dynamiske spekteret til
lyden, og alle eller noen av parametrene i audioforsterknings- og dynamikk-
prosesseringen blir generert som funksjon av hørehendelser. Oppfinnelsen ved-
10 rører også dataprogrammer for å praktisere slike fremgangsmåter eller styre en
slik anordning.

Foreliggende oppfinnelse vedrører også fremgangsmåter og en anordning
som anvender en spesifikk lydstyrkebasert deteksjon av hørehendelser. Oppfinn-
elsen vedrører også dataprogrammer for å praktisere slike fremgangsmåter eller
15 styre en slik anordning.

Bakgrunn

Dynamikkprosessering av lyd

20 Metodene automatisk forsterkningsregulering (AGC - Automatic Gain
Control) og dynamisk spektrumkontroll (DRC - Dynamic Range Control) er velkj-
ente, og er et vanlig element i mange audiosignalbaner. Generelt måler begge
metodene nivået til et audiosignal på en eller annen måte og forsterknings-
modifiserer så signalet med en mengde som er en funksjon av det målte nivået. I
et lineært, 1:1 dynamikkprosesseringssystem blir ikke den innmatede lyden beh-
25 andlet, og det utmatede audiosignalet svarer ideelt sett til det innmatede audio-
signalet. Videre, i et audiodynamikkprosesseringssystem som automatisk måler
egenskaper ved inngangssignalet og anvender denne målingen for å regulere
utgangssignalet, dersom inngangssignalets nivå stiger med 6 dB og utgangs-
signalet blir behandlet slik at dets nivå bare stiger med 3 dB, blir utgangssignalet
30 komprimert i et forhold på 2:1 med hensyn til inngangssignalet. Den
internasjonale publikasjonen WO 2006/047600 A1 ("Calculating and Adjusting the
Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal", av
Alan Jeffrey Seefeldt) gir en detaljert oversikt over de fem grunnleggende typene
dynamikkprosessering av lyd: komprimering, begrensnings, automatisk forsterk-
35 ningskontroll (AGC), ekspansjon og gating.

US 2004122662 A1 beskriver i et eksempel på utførelse hvordan et lydsignal analysert ved bruk av flere psykoakustiske kriterier identifiserer et område av signalet der tidsskalering og / eller tonehøyde-forskyvningsprosessering kan være hørbar eller minimalt hørbar, og signalet er tidsskala og / eller tonehøyde forskjøvet innenfor omådet. Alternativ er signalet delt inn i 5
auditive hendelser, og signalet blir skalert og / eller tonehøyde forskjøvet i løpet av en auditive hendelse. I et ytterligere alternativ er signalet delt inn i auditive hendelser, og de auditive hendelsene blir analysert ved bruk av et psykoakustisk kriterium for å identifisere de auditive hendelser der tidsskaleringen og / eller 10
tonehøyde-prosesseringen av signalet ville være uhørbart eller minimalt hørbar.

US 2006029239 A1 beskriver en prosess for å kombinere lydkanaler for å produsere en kombinert lydkanal som bruker dynamisk én eller flere av tid, fase og amplitude- eller effektjusteringer til kanalene, for den kombinerte kanalen, eller for både kanalene og den kombinerte kanalen. Én eller flere av justeringene 15
styres i det minste delvis av et mål på hørselshendelser i én eller flere av kanalene og / eller den kombinerte kanalen. Bruksområder inkluderer presentasjon av flerkanalslyd på kinoer og i kjøretøy.

US 2004165730 A1 beskriver oppdeling av et lydsignal i auditive hendelser, som hver har en tendens til å bli oppfattet som separate og distinkte, 20
ved å beregne spektralt innhold i suksessive tidsblokker for lydsignalet (5-1), og beregne forskjellen i spektral innhold mellom suksessive tidsblokker for lydsignalet (5-2), og ved identifisering av en hørselshendelsesgrense som grensen mellom suksessive tidsblokker når forskjellen i spektralt innhold mellom slike suksessive tidsblokker overskrider en terskel (5-3). I et annet aspekt genereres en 25
representasjon med redusert informasjon av et lydsignal ved å dele et lydsignal inn i hørselshendelser, som hver har en tendens til å bli oppfattet som separate og distinkte, og formatere og lagre informasjon relatert til hørselshendelsene (5- 4). Eventuelt kan det videre også tilordnes et kjennetegn til en eller flere av hørselshendelsene (5-5).

30

Hørehendelser og deteksjon av hørehendelser

Inndeling av lyder i enheter eller segmenter som oppfattes som atskilte og distinkte, kalles noen ganger "hørehendelsesanalyse" eller høresceneanalyse (ASA - Auditory Scene Analysis), og segmentene kalles noen ganger "hørehendelser" 35
eller "lydhendelser." En omfattende beskrivelse av høresceneanalyse er gitt av Albert S. Bregman i hans bok *Auditory Scene Analysis -- The Perceptual*

Organization of Sound, Massachusetts Institute of Technology, 1991, fjerde trykk, 2001, Second MIT Press paperback edition). Videre omtaler U.S.-patentet 6,002,776 til Bhadkamkar, m.fl. fra 14. desember 1999 publikasjoner fra tilbake til 1976 som "(oversatt) kjent arbeid knyttet til lydseparasjon ved høre-
5 sceneanalyse". Patentet til Bhadkamkar m.fl. fraråder imidlertid praktisk bruk av hørsceneanalyse, og konkluderer med at "(oversatt) metoder som omfatter hørsceneanalyse, selv om de er interessante sett fra et vitenskaplig ståsted som modeller for lydbehandling hos mennesker, er i dag alt for beregningskrevende og spesialisert til å kunne betraktes som praktisk anvendelige metoder for separasjon
10 av lyd før det er gjort fundamentale fremskritt".

En nyttig måte å identifisere hørehendelser er angitt av Crockett og Crockett m. fl. i forskjellige patentsøknader og artikler listet nedenfor under overskriften "Referanser". I henhold til disse dokumentene blir et audiosignal delt inn i høre-
hendelser, som hver gjerne oppfattes som atskilt og distinkt, ved å detektere end-
15 ringer i spektralsammensetning (amplitude som funksjon av frekvens) som funksjon av tid. Dette kan for eksempel gjøres ved å beregne spektralinnholdet i suksessive tidsblokker av audiosignalet, beregne forskjellen i spektralinnhold mellom suksessive tidsblokker av audiosignalet, og identifisere en hørehendelsesgrense som grensen mellom suksessive tidsblokker når forskjellen i spektralinnhold mel-
20 lom disse suksessive tidsblokkene overstiger en terskel. Alternativt kan endringer i amplitude som funksjon av tiden bli beregnet i stedet for eller i tillegg til endringer i spektralsammensetning som funksjon av tiden.

I sin minst beregningskrevende utførelse deler prosessen inn lyd i tidssegmenter ved å analysere hele frekvensbåndet (lyd med full båndbredde) eller
25 hovedsaklig hele frekvensbåndet (i praksis anvendes ofte båndbegrensende filtrering i endene av spekteret) og vekte de høyeste audiosignalkomponentene tyngst. Denne løsningen utnytter et psykoakustisk fenomen der øret på mindre tidsskalaer (20 millisekunder (ms) og mindre) kan ha en tendens til å fokusere på én enkelt hørehendelse om gangen. Dette innebærer at selv om flere hendelser vil
30 kunne forekomme samtidig, én komponent gjerne er perseptuelt mest fremtredende og kan bli behandlet individuelt som om den var den eneste hendelsen på det aktuelle tidspunktet. Utnyttelse av denne effekten gjør også at hørehendelsesdeteksjonen kan skalere med kompleksiteten til lyden som behandles. Dersom for eksempel det innmatede audiosignalet som behandles er et soloinstrument, vil
35 hørehendelsene som identifiseres trolig være de individuelle notene som spilles. For et innmatet stemmesignal vil likeledes de individuelle talekomponentene, for

eksempel vokalene og konsonantene, mest sannsynlig bli identifisert som individuelle lydelementer. Når lydets kompleksitet øker, så som musikk med trommeslag eller flere instrumenter og stemmer, identifiserer hørehendelsesdeteksjonen til enhver tid det "mest fremtredende" (dvs. det høyeste) lydelementet.

5 På bekostning av større regnekompleksitet kan prosessen også ta hensyn til endringer i spektralsammensetning som funksjon av tid i diskrete delfrekvensbånd (faste eller dynamisk bestemte eller både faste og dynamisk bestemte delbånd) heller enn hele båndbredden. Denne alternative tilnærmingen betrakter flere enn én lydstrøm i forskjellige delfrekvensbånd heller enn å anta at bare én enkelt
10 strøm kan oppfattes på et gitt tidspunkt.

Hørehendelsesdeteksjon kan gjøres ved å dele inn en tidsdomene lyd-
bølgeform i tidsintervaller eller blokker og så konvertere dataene i hver blokk til
frekvensdomenet, med bruk av enten en filterbank eller en tid/frekvens-
transformasjon, for eksempel FFT. Amplituden til spektralinnholdet i hver blokk
15 kan bli normalisert for å fjerne eller redusere innvirkningen av
amplitudeendringer. Hver resulterende frekvensdomenerepresentasjon gir en
angivelse av spektralinnholdet i lyden i den aktuelle blokken. Spektralinnholdet i
suksessive blokker blir sammenliknet, og endringer som er større enn en terskel
brukes som angivelse av tidsmessig begynnelse eller tidsmessig avslutning av en
20 hørehendelse.

Fortrinnsvis blir frekvensdomenedataene normalisert, som beskrevet
nedenfor. I hvilken grad frekvensdomenedataene trenger å bli normalisert gir en
angivelse av amplitude. Dersom en endring i denne normaliseringsgraden over-
stiger en forbestemt terskel, kan således også dette brukes som en angivelse av
25 en hendelsesgrense. En hendelses start- og sluttpunkter som følge av spektral-
endringer og av amplitudeendringer kan bli ELLER-behandlet sammen slik at
hendelsesgrenser som følge av begge typene endring blir identifisert.

Selv om metoder beskrevet i nevnte søknader og artikler av Crockett og
Crockett m.fl. er spesielt nyttige i forbindelse med aspekter ved foreliggende
30 oppfinnelse, kan andre metoder for å identifisere hørehendelser og hendels-
esgrenser anvendes i aspekter ved foreliggende oppfinnelse.

Beslektet teknologi beskrives i BLESSER, BARRY: An Ultraminiature Console
Compression System with Maximum User Flexibility, Journal of Audio Engineering
Society, vol. 20, no. 4, May 1972, New York, sidene 297-302.

35

I et første aspekt tilveiebringer den foreliggende oppfinnelsen en fremgangsmåte for å modifisere en parameter for en audiodynamikkprosessor, omfattende å detektere endringer i spektralegenskaper med hensyn til tid i et audiosignal, identifisere som hørehendelsesgrenser endringer som overstiger en terskel i spektralegenskaper med hensyn til tid i nevnte audiosignal, der et audiosegment mellom konsekutive grenser danner en hørehendelse, og generere et parameter-modifiserende reguleringssignal basert på nevnte identifiserte hørehendelsesgrenser, og modifisere parameteren for audiodynamikkprosessoren som funksjon av reguleringssignalet.

I et andre aspekt tilveiebringer den foreliggende oppfinnelsen et data-program, lagret på et datamaskinlesbart medium, for å bevirke en datamaskin til å utføre en fremgangsmåte som angitt ovenfor.

Foretrukkede utførelsesformer av oppfinnelsen er angitt i kravene 2 – 11.

Tradisjonell, kjent audiodynamikkprosessering omfatter det å multiplisere lyden med et tidsvarierende reguleringssignal som justerer forsterkningen av lyden og gir et ønsket resultat. "Forsterkning" er en skaleringsfaktor som skalerer lydampplituden. Dette reguleringssignalet kan bli generert kontinuerlig eller fra blokker av lyddata, men det blir i alminnelighet avledet fra en form for måling av lyden som behandles, og dets endringsrate bestemmes av glattingsfiltre, noen ganger med faste egenskaper og noen ganger med egenskaper som varierer med lydets dynamikk. For eksempel kan responstider være justerbare i henhold til endringer i lydens absoluttverdi eller effekt. Kjente metoder, så som automatisk forsterkningsregulering (AGC) og dynamisk spektrumkontroll (DRC), vurderer ikke på en psykoakustikk-basert måte tidsintervallene der forsterkningsendringer kan oppfattes som forringelser og når de kan anvendes uten å gi hørbare artefakter. Tradisjonelle audiodynamikkprosesser kan derfor ofte introdusere hørbare artefakter, dvs. at innvirkningen av dynamikkprosesseringen kan introdusere uønskede merkbare endringer av lyden.

Høresceneanalyse identifiserer perseptuelt diskrete hørehendelser, der hver hendelse opptrer mellom to konsekutive hørehendelsesgrenser. De hørbare forringelsene forårsaket av en forsterkningsendring kan reduseres betydelig ved å sikre at forsterkningen innenfor en hørehendelse er tilnærmet konstant og ved å begrense mye av endringen til området rundt en hendelsesgrense. For kompressorer eller ekspandere kan responsen til en økning i lydnivå (ofte kalt attack) være rask, sammenliknbar med eller kortere enn den minste varigheten

av hørehendelser, men responsen til en reduksjon (release eller gjenoppretting) kan være langsommere slik at lyder som burde synes konstante eller å avta gradvis kan bli hørbart forstyrret. Under slike omstendigheter er det veldig gunstig å forsinke forsterkningsgjenoprettelsen til den neste grensen eller å redusere endringsraten til forsterkningen under en hendelse. For anvendelser med automatisk forsterkningsregulering der lydens nivå eller styrke på mellomlang til lang tidsskala blir normalisert og både attack- og releasetider derfor kan være lange sammenliknet med den minste varigheten av en hørehendelse, er det gunstig under hendelser å forsinke endringer av eller redusere endringsrater for forsterkning til neste hendelsesgrense for både økende og avtagende forsterkning.

Det beskrives at et audioprosesseringssystem mottar et audiosignal og analyserer og endrer forsterkningen av og/eller de dynamiske spektrumtrekkene til lyden. Dynamisk spektermodifikasjon av lyden blir ofte styrt av parametere i et dynamikkprosesseringssystem (attack- og releasetid, komprimeringsforhold, etc.) som har betydelig innvirkning på de perseptuelle artefaktene som introduseres av dynamikkprosesseringen. Endringer i signalegenskaper som funksjon av tid i audiosignalet blir detektert og identifisert som hørehendelsesgrenser, slik at et audiosegment mellom konsekutive grenser danner en hørehendelse i audiosignalet. Egenskapene til de aktuelle hørehendelsene kan omfatte egenskaper ved hendelsene så som perseptuell styrke eller varighet. Noen av nevnte én eller flere dynamikkprosesseringsparametere blir generert i hvert fall delvis som reaksjon på hørehendelser og/eller graden av endring i signalegenskaper for nevnte hørehendelsesgrenser.

En hørehendelse er typisk et audiosegment som gjerne oppfattes som atskilt og distinkt. Ett anvendelig mål på signalegenskaper omfatter et mål på spektralinnholdet i lyden, for eksempel som beskrevet i de omtalte dokumentene til Crockett og Crockett m.fl. Alle eller noen av den ene eller de flere audiodynamikkprosesseringsparametrene kan bli generert i hvert fall delvis som reaksjon på tilstedeværelse eller fravær av og egenskaper ved én eller flere hørehendelser. En hørehendelsesgrense kan bli gjenkjent som en endring i signalegenskaper som funksjon av tid som overstiger en terskel. Alternativt kan alle eller noen av den ene eller de flere parametrene bli generert i hvert fall delvis som reaksjon på et kontinuerlig mål for graden av endring i signalegenskaper for nevnte hørehendelsesgrenser. Selv om aspekter ved oppfinnelsen i prinsippet kan realiseres med analoge og/eller digitale signaler, vil utførelser i praksis mest sannsynlig bli realisert med digitale signaler der hvert av audiosignalene er representert av enkeltstående

sampler eller sampler innenfor datablokker. I dette tilfellet kan signalegenskapene være spektralinnholdet i lyd innenfor en blokk, deteksjon av endringer i signalegenskaper som funksjon av tid kan være deteksjon av endringer i spektralinnhold i lyd fra blokk til blokk, og hver av hørehendelsers tidsmessige start- og stoppgrenser kan sammenfalle med en grense for en datablokk. Det skal bemerkes at i det mer tradisjonelle tilfellet med gjennomføring av dynamiske forsterkningsendringer for sample for sample, den beskrevne høresceneanalysen vil kunne utføres på blokkbasis og den resulterende hørehendelsesinformasjonen anvendes for å utføre dynamiske forsterkningsendringer som anvendes sample for sample.

Ved å styre viktige audiodynamikkprosesseringsparametere med bruk av resultatene av høresceneanalysen kan en oppnå en dramatisk reduksjon av hørbare artefakter introdusert av dynamikkprosessering.

Det beskrives to måter å gjennomføre høresceneanalyse. Den første utfører spektralanalyse og identifiserer posisjonen til oppfattbare lydhendelser, som anvendes for å styre de dynamiske forsterkningsparametrene ved å identifisere endringer i spektralinnhold. Den andre måten transformerer lyden til et perseptuell lydstyrkedomene (som kan gi mer psykoakustisk relevant informasjon enn den første måten) og identifiserer posisjonen til hørehendelser, som deretter anvendes for å styre de dynamiske forsterkningsparametrene. Det skal bemerkes at den andre måten krever at lydbehandlingen kjenner absolutte akustiske reproduksjonsnivåer, som kan være umulig i noen utførelser. Det å vise begge høresceneanalyse-metodene muliggjør utførelser av ASA-regulert dynamisk forsterkningsmodifisering som anvender prosesser eller anordninger som kan, men ikke trenger være kalibrert til å ta hensyn til absolutte reproduksjonsnivåer.

Aspekter ved foreliggende oppfinnelse er beskrevet her i et audiodynamikkprosesseringsmiljø som omfatter aspekter ved andre oppfinnelser. Disse andre oppfinnelsene er beskrevet i forskjellige verserende patentsøknader i USA og internasjonale patentsøknader fra Dolby Laboratories Licensing Corporation, eieren av denne søknaden, hvilke søknader er identifisert her.

Foreliggende oppfinnelse tilveiebringer en fremgangsmåte ved behandling av et lydsignal i et lydbehandlingsapparat, hvor fremgangsmåten omfatter:

å motta et lydsignal, lydsignalet omfatter minst én kanal med lydinnhold;
å dele lydsignalet i flere underbåndssignaler med en analyse filterbank,
hver av flerheten av underbåndssignaler omfatter minst ett underbåndseksempel;

å utlede en karakteristikk av lydsignalet, hvor karakteristikken er et effektmål av lydsignalet;

å glatte effektmålingen for å generere et glatt effektmål av lydsignalet;

5 å detektere en plassering av en hørselsbegrensningsgrense ved å overvåke det glatte effektmålet, hvor en lyddel mellom sammenhengende hørselsgrenser utgjør en høringshendelse, omfattende:

å generere en forsterkningsvektor basert på plasseringen av hørselsgrensen,

10 å anvende forsterkningsvektoren på en versjon av flerheten av underbåndssignaler for å generere modifiserte underbåndssignaler; og

syntetisere de modifiserte underbåndssignalene med en syntetiseringsfilterbank for å fremstille et modifisert lydsignal,

hvor deteksjonen videre omfatter å anvende en terskel for det glatte effektmålet for å detektere plasseringen av den hørbare hendelsesgrensen,

15 hvor deteksjonen videre omfatter å sammenligne det glatte effektmålet med et andre glatt effektmål av lydsignalet,

hvor det andre effektmålet er basert på et differansemål, og

hvor lydbehandlingsapparatet er implementert i det minste delvis i maskinvare.

20 Foreliggende oppfinnelse tilveiebringer et lydbehandlingsapparat, hvor apparatet omfatter:

et inngangsgrensesnitt for mottak av et lydsignal, lydsignalet omfatter minst én kanal med lydinnhold;

25 en analyse-filterbank for å dele lydsignalet i en flerhet av underbåndssignaler, hver av flerheten av underbåndssignaler omfatter minst ett underbåndsbåndseksempel; og

en prosessor som:

danner en karakteristikk av lydsignalet, hvor karakteristikken er et effektmål av lydsignalet;

30 glatter ut effektmålet for å generere et glatt effektmål av lydsignalet; hvor lydbehandlingsapparatet videre

detekterer en plassering av en hørselsbegrensningsgrense ved å overvåke det glatte effektmålet, hvor en lyddel mellom sammenhengende hørselsgrenser utgjør en lydhendelse,

35 genererer en forsterkningsvektor basert på plasseringen av hørselsgrensen; og

anvender forsterkningsvektoren på en versjon av flerheten av underbåndssignaler for å fremstille modifiserte underbåndssignaler; og en syntetiserings filterbank som kombinerer de modifiserte underbåndssignalene for å fremstille et modifisert lydsignal,

5 hvor deteksjonen videre omfatter å anvende en terskel for det glatte effektmålet for å detektere plasseringen av den hørbare hendelsesgrensen, hvor deteksjonen videre omfatter å sammenligne det glatte effektmålet med et andre glatt effektmål av lydsignalet,

10 hvor effektmålet eller det andre effektmålet er basert på et differansemål.

Beskrivelse av figurene

Figur 1 er et flytdiagram som viser et eksempel på behandlingstrinn for å gjennomføre høresceneanalyse.

15 Figur 2 viser et eksempel på blokkbehandling, vindusbehandling og gjennomføring av DFT på lyd samtidig med gjennomføring av høresceneanalysen.

Figur 3 er et flytdiagram eller funksjonelt blokkdiagram som viser parallellprosessering der lyd anvendes for å identifisere hørehendelser og for å identifisere hørehendelsenes egenskaper slik at hendelsene og deres egenskaper anvendes for

20 å modifisere dynamikkprosesseringsparametre.

Figur 4 er et flytdiagram eller funksjonelt blokkdiagram som viser behandling der lyd kun anvendes for å identifisere hørehendelser og hendelsenes egenskaper blir bestemt fra hørehendelsesdeteksjonen slik at hendelsene og deres egenskaper anvendes for å modifisere dynamikkprosesseringsparametrene.

25 Figur 5 er et flytdiagram eller funksjonelt blokkdiagram som viser behandling der lyd kun anvendes for å identifisere hørehendelser og hendelsenes egenskaper blir bestemt fra hørehendelsesdeteksjonen og slik at bare egenskapene til hørehendelsene anvendes for å modifisere dynamikkprosesseringsparametrene.

30 Figur 6 viser et sett av idealiserte lydfilter-egenskapsresponser som tilnærmer kritisk båndning på ERB-skalaen. Den horisontale akse angir frekvens i Hertz og den vertikale akse angir nivå i desibel.

Figur 7 viser konturene for lik lydstyrke i ISO 226. Den horisontale akse angir frekvens i Hertz (logaritmisk skala med grunntall 10) og den vertikale akse angir lydtryknivå i desibel.

35

Figurene 8a-c viser idealiserte inn/ut-egenskaper og innsignal-forsterkningsegenskaper til en lydkompressor med dynamisk spektrum.

Figurene 9a-f viser et eksempel på bruk av hørehendelser for å styre releasetiden i en digital utførelse av en tradisjonell DRC (Dynamic Range Controller) der forsterkningsreguleringen er avledet fra signalets RMS-(Root Mean Square)-effekt.

Figurene 10a-f viser et eksempel på bruk av hørehendelser for å styre releasetiden i en digital utførelse av en tradisjonell DRC der forsterkningsreguleringen er avledet fra RMS-effekten til signalet for et alternativt signal til det anvendt i figur 9.

Figur 11 viser et passende sett av idealiserte AGC- og DRC-kurver for bruk av AGC etterfulgt av DRC i et lydstyrkedomene-basert dynamikkprosesserings-system. Målet med kombinasjonen er å gjøre at all behandlet lyd får omtrent samme oppfattede lydstyrke samtidig som i hvert fall noe av den opprinnelige lydens dynamikk opprettholdes.

Beste måte å realisere oppfinnelsen

Høresceneanalyse (opprinnelig, ikke lydstyrkedomene-basert metode)

Ifølge en utførelsesform av ett aspekt ved foreliggende oppfinnelse kan høresceneanalyse bestå av fire generelle behandlingstrinn som vist i en del av figur 1. Det første trinnet, 1-1 ("Utfør spektralanalyse"), tar et tidsdomene audio-signal, deler det inn i blokker og beregner et spektralprofil eller spektralinnhold for hver av blokkene. Spektralanalyse transformerer audiosignalet til det hurtigskala frekvensdomenet. Dette kan gjøres med bruk av en hvilken som helst filterbank, enten basert på transformasjoner eller banker av båndpassfiltre, og i enten et lineært eller krummet frekvensrom (så som Bark-skalaen eller det kritiske båndet, som bedre tilnærmer egenskapene til det menneskelige øre). I enhver filterbank foreligger det en avveining mellom tid og frekvens. Større tidsoppløsning, og således kortere tidsintervaller, gir en lavere frekvensoppløsning. Større frekvensoppløsning, og således smalere delbånd, leder til lengre tidsintervaller.

Det første trinnet, illustrert konseptuelt i figur 1, beregner spektralinnholdet i suksessive tidssegmenter av audiosignalet. I en praktisk utførelsesform kan ASA-blokkstørrelsen være fra et hvilket som helst antall sampler av det innmatede audiosignalet, selv om 512 sampler gir en god avveining mellom tids- og frekvensoppløsning. I det andre trinnet, 1-2, bestemmes forskjellene i spektralinnhold fra blokk til blokk ("Gjør målinger av forskjeller i spektralprofil"). Det andre

trinnet beregner således forskjellen i spektralinhold mellom suksessive tidssegmenter av audiosignalet. Som beskrevet over antas en god indikator for begynnelsen eller slutten av en oppfattet hørehendelse å være en endring i spektralinhold. I det tredje trinnet, 1-3 ("Bestem posisjonen til hørehendelsesgrensene"),
 5 når spektralforskjellen mellom en spektralprofilblokk og den neste er større enn en terskel, tas blokkgrensen som en hørehendelsesgrense. Audiosegmentet mellom konsekutive grenser danner en hørehendelse. Det tredje trinnet trekker således en hørehendelsesgrense mellom suksessive tidssegmenter når forskjellen i spektralprofilinnhold mellom disse suksessive tidssegmentene overstiger en
 10 terskel, og definerer således hørehendelser. I denne utførelsesformen definerer hørehendelsesgrenser hørehendelser med en lengde som er et heltallig multiplum av spektralprofilblokker med en minimumslengde på én spektralprofilblokk (512 sampler i dette eksempelet). I prinsippet trenger ikke hendelsesgrenser være begrenset på denne måten. Som et alternativ til de praktiske utførelsesformene beskrevet her kan den innmatede blokkstørrelsen varieres, for eksempel slik at den er
 15 tilnærmet lik størrelsen til en hørehendelse.

Etter identifisering av hendelsesgrensene bestemmes nøkkelegenskaper ved hørehendelsene, som vist i trinn 1-4.

Enten overlappende eller ikke-overlappende segmenter av lyden kan bli vindusbehandlet og anvendt for å beregne spektralprofiler av den innmatede lyden. Overlapp resulterer i en finere oppløsning av posisjonen til hørehendelser, og reduserer også sannsynligheten for å overse en hendelse, så som en kortvarig transient. Overlapp gjør imidlertid også beregningene mer kompliserte. Følgelig kan en unngå overlapp. Figur 2 viser en konseptuell representasjon av ikke-
 25 overlappende blokker av N sampler som er vindusbehandlet og transformert til frekvensdomenet i en diskret fouriertransformasjon (DFT). Hver blokk kan bli vindusbehandlet og transformert til frekvensdomenet for eksempel med bruk av en DFT-transformasjon, fortrinnsvis en hurtig fouriertransformasjon (FFT) for å gjøre beregningene raskere.

30 Følgende variabler kan anvendes for å beregne spektralprofilet til den innmatede blokken:

M = antall vindusbehandlede sampler i en blokk som anvendes for å beregne spektralprofil

P = antall overlappende sampler i spektralberegningen

35 Generelt kan variablene over ta et hvilket som helst heltall. Imidlertid vil gjennomføringen være mer effektiv dersom M settes lik en potens av 2 slik at

standard FFT-transformasjoner kan anvendes for spektralprofilberegningene. I en praktisk utførelsesform av høresceneanalyse-prosessen kan de listede parametrene være satt til:

$$\begin{aligned} M &= 512 \text{ sampler (eller 11,6 ms ved 44,1 kHz)} \\ P &= 0 \text{ sampler (ingen overlapp)} \end{aligned}$$

Verdiene angitt over ble bestemt eksperimentelt og ble funnet å generelt identifisere posisjonen og varigheten til hørehendelser med tilstrekkelig nøyaktighet. Imidlertid er det å sette verdien til P til 256 sampler (50% overlapp) heller enn null sampler (ingen overlapp) funnet å være nyttig ved bestemmelse av visse hendelser som er vanskelige å oppdage. Selv om mange forskjellige typer vinduer kan anvendes for å minimere spektralartefakter som følge av vindusbehandling, er vinduet som anvendes i spektralprofilberegningene et M-punkts Hanning, Kaiser-Bessel eller annet passende, fortrinnsvis ikke-rektangulært vindu. De ovenfor angitte verdiene og en Hanning-vindustype ble valgt etter omfattende eksperimentell analyse ettersom de viste seg å gi utmerkede resultater over et stort utvalg av lydmateriale. Behandling med ikke-rektangulære vinduer er foretrukket for behandling av audiosignaler med hovedsaklig lavfrekvent innhold. Rektangulær vindusbehandling skaper spektralartefakter som kan forårsake feilaktig deteksjon av hendelser. Til forskjell fra bestemte koder/dekoder-(kodek)-applikasjoner der en generell overlappings-/tilleggingsprosess må skape et konstant nivå, er det ingen slike begrensninger her, og vinduet kan velges for egenskaper så som dets tids-/frekvensoppløsning og stoppbånd-avvisning.

I trinn 1-1 (figur 1) kan spekteret til hver blokk av M-sampler bli beregnet ved å vindusbehandle dataene med et M-punkts Hanning, Kaiser-Bessel eller annet passende vindu, transformere til frekvensdomenet med bruk av en M-punkts FFT og beregne absoluttverdien til de komplekse FFT-koeffisientene. Resultatdataene blir normalisert slik at den største absoluttverdien settes lik én, og det normaliserte arrayet av M tall blir transformert til log-domenet. Dataene kan også bli normalisert med hensyn til et annet mål, så som midlere absoluttverdi eller midlere effektverdi til dataene. Arrayet trenger ikke bli transformert til log-domenet, men transformasjonen forenkler beregningen av forskjellsmålet i trinn 1-2. Videre sammenfaller log-domenet bedre med egenskapene til menneskets hørselssystem. De resulterende log-domeneverdiene ligger i et område fra minus uendelig til null. I en praktisk utførelsesform kan en nedre grense bli satt for området av verdier; grensen kan være fast, for eksempel -60 dB, eller være frekvensavhengig for å reflektere den dårligere hørbarheten av

stille lyder ved lave og veldig høye frekvenser. (Merk at det ville være mulig å redusere størrelsen til arrayet til $M/2$ ettersom FFT representerer negative så vel som positive frekvenser).

Trinn 1-2 beregner et mål for forskjellen mellom spektrene til konsekutive blokker. For hver blokk blir hver av de M (log-) spektralkoeffisientene fra trinn 1-1 subtrahert fra den motsvarende koeffisienten i den foregående blokken, og absoluttverdien til differansen beregnes (fortegnet ignoreres). Disse M differansene blir så summert til ett tall. Dette forskjellsmålet kan også uttrykkes som en gjennomsnittsforskjell for hver spektralkoeffisient ved å dividere forskjellsmålet med antallet spektralkoeffisienter anvendt i summen (i dette tilfellet M koeffisienter).

Trinn 1-3 identifiserer posisjonen til hørehendelsesgrenser ved å anvende en terskel på gruppen av forskjellsmål fra trinn 1-2 med en terskelverdi. Når et forskjellsmål overstiger en terskel, vurderes endringen i spekteret som tilstrekkelig til å signalisere en ny hendelse, og blokknummeret til endringen blir registrert som en hendelsesgrense. For verdiene av M og P angitt over, og for log-domeneverdier (i trinn 1-1) uttrykt i dB, kan terskelen settes lik 2500 dersom hele FFT-sekvensen (omfattende den speilede delen) blir sammenliknet, eller 1250 dersom halve FFT-sekvensen blir sammenliknet (som angitt over representerer FFT negative så vel som positive frekvenser — for absoluttverdien til FFT-transformasjonen er den ene speilbildet av den andre). Denne verdien ble valgt eksperimentelt og gir en god deteksjon av hørehendelsesgrenser. Denne parameterverdien kan endres for å redusere (øke terskelen) eller øke (redusere terskelen) deteksjonen av hendelser.

Prosessene i figur 1 kan representeres mer generelt av de ekvivalente prosessene i figurene 3, 4 og 5. I figur 3 anvendes et audiosignal parallelt på en "Identifisere hørehendelser"-funksjon, eller trinn 3-1, som deler inn audiosignalet i hørehendelser, som hver gjerne oppfattes som atskilt og distinkt, og på en eventuell "Identifisere egenskaper ved hørehendelser"-funksjon, eller trinn 3-2. Prosessen i figur 1 kan anvendes for å dele inn audiosignalet i hørehendelser og deres identifiserte egenskaper, eller en annen passende prosess kan anvendes. Hørehendelsesinformasjonen, som kan være en identifisering av hørehendelsesgrenser, bestemt av funksjon eller trinn 3-1, blir så anvendt for å modifisere audiodynamikkprosesseringsparametrene (så som attack, release, forhold, etc.), som ønsket, i en "Modifiser dynamikkparametere"-funksjon, eller trinn 3-3. Den valgfri "Identifiser egenskaper"-funksjonen, eller trinn 3-3, mottar også

hørehendelsesinformasjonen. "Identifiser egenskaper"-funksjonen, eller trinn 3-3, kan karakterisere noen av eller alle hørehendelsene ved én eller flere egenskaper. Slike egenskaper kan omfatte en identifisering av det dominerende delbåndet i hørehendelsen, som beskrevet i forbindelse med prosessen i figur 1. Egenskapene kan også omfatte én eller flere lydegenskaper, omfattende for eksempel et mål for effekten til hørehendelsen, et mål for amplituden til hørehendelsen, et mål for spektralflatheten til hørehendelsen og hvorvidt hørehendelsen er hovedsaklig lydløs, eller andre egenskaper som letter modifisering av dynamiske parametere slik at negative hørbare artefakter som følge av behandlingen reduseres eller fjernes. Egenskapene kan også omfatte andre egenskaper, så som hvorvidt hørehendelsene omfatter en transient.

Alternativer til prosessen i figur 3 er vist i figurene 4 og 5. I figur 4 blir ikke det innmatede audiosignalet anvendt direkte i "Identifiser egenskaper"-funksjonen eller trinn 4-3, men det mottar informasjon fra "Identifiser hørehendelser"-funksjonen eller trinn 4-1. Prosessen i figur 1 er et konkret eksempel på en slik prosess. I figur 5 blir funksjonene eller trinnene 5-1, 5-2 og 5-3 utført i serie.

Detaljene i denne praktiske utførelsesformen er ikke avgjørende. Andre måter kan anvendes for å beregne spektralinnholdet i suksessive tidssegmenter av audiosignalet, beregne forskjellen mellom suksessive tidssegmenter og sette hørehendelsesgrenser i de respektive grensene mellom suksessive tidssegmenter når forskjellen i spektralprofilinnhold mellom disse suksessive tidssegmentene overstiger en terskel.

Høresceneanalyse (Ny, lydstyrkedomene-basert metode)

Den internasjonale søknaden under Patentsamarbeidskonvensjonen PCT/US2005/038579, innlevert 25. oktober 2005, publisert som den internasjonale publikasjonen WO 2006/047600 A1 med tittelen "Calculating and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal", av Alan Jeffrey Seefeldt, beskriver blant annet et objektivt mål for oppfattet lydstyrke basert på en psykoakustisk modell.. Som beskrevet i nevnte søknad PCT/US2005/038579 blir det fra et audiosignal beregnet et eksitasjonssignal $E[b,t]$ som tilnærmer fordelingen av energi langs basilmembranen i det indre øret ved et kritisk bånd b under tidsblokk t . Denne eksitasjonen kan beregnes fra en diskret hurtigskala fouriertransformasjon (STDFFT – Short-time Discrete Fourier Transform) av audiosignalet som følger:

$$E[b, t] = \lambda_b E[b, t-1] + (1 - \lambda_b) \sum_k |T[k]|^2 |C_b[k]|^2 |X[k, t]|^2$$

(1)

der $X[k, t]$ representerer STDFT-transformasjonen av $x[n]$ ved tidsblokk t og boks
 5 k . Merk at i likning 1, t representerer tid i diskrete enheter av transformasjons-
 blokker til forskjell fra et kontinuerlig mål, så som sekunder. $T[k]$ representerer
 frekvensresponsen til et filter som simulerer overføring av lyd gjennom det ytre og
 midtre øre, og $C_b[k]$ representerer frekvensresponsen til basilarmembranen på et
 10 sted som svarer til det kritiske båndet b . Figur 6 viser et passende sett av kritiske
 båndfilterresponser der 40 bånd er spredt uniformt langs ekvivalent rektangulær
 båndbredde-(ERB)-skalaen, som definert av Moore og Glasberg. Hver filterform er
 beskrevet av en avrundet eksponentialfunksjon, og båndene er distribuert med en
 spredning på 1 ERB. Endelig kan glattingstidskonstanten λ_b i likning 1 tjenlig vel-
 15 ges proporsjonal med integrasjonstiden til menneskets lydstyrkeoppfatning innen-
 for bånd b .

Ved anvendelse av isolydstyrkekonturer, så som de vist i figur 7, transfor-
 meres eksitasjonen i hvert bånd til et eksitasjonsnivå som vil generere den
 samme oppfattede lydstyrken ved 1kHz. Spesifikk lydstyrke, et mål for perseptuell
 lydstyrke distribuert over frekvens og tid, blir så beregnet fra den transformerte
 20 eksitasjonen, $E_{1kHz}[b, t]$, gjennom en trykk-ikke-linearitet. Én passende funksjon
 for å beregne den spesifikke lydstyrken $N[b, t]$ er gitt ved:

$$N[b, t] = \beta \left(\left(\frac{E_{1kHz}[b, t]}{TQ_{1kHz}} \right)^\alpha - 1 \right)$$

(2)

25

der TQ_{1kHz} er terskelen for lydløs ved 1kHz og konstantene β og α velges for å
 tilpasse til vekst i lydstyrkedata som samlet inn fra lytteeksperimenter. Teoretisk
 kan denne transformasjonen fra eksitasjon til spesifikk lydstyrke representeres av
 funksjonen $\Psi\{ \}$ som er slik at:

30

$$N[b, t] = \Psi\{E[b, t]\}$$

Endelig beregnes den totale lydstyrken, $L[t]$, representert i sone-enheter, ved å summere den spesifikke lydstyrken over bånd:

$$L[t] = \sum_b N[b, t]$$

(3)

Den spesifikke lydstyrken $N[b, t]$ er en spektralrepresentasjon ment for å simulere hvordan et menneske oppfatter lyd som funksjon av frekvens og tid. Den fanger opp variasjoner i følsomhet for forskjellige frekvenser, variasjoner i nivåfølsomhet samt variasjoner i frekvensoppløsning. Følgelig er den en spektralrepresentasjon som er velegnet for deteksjon av hørehendelser. Selv om det gir mer kompliserte beregninger, kan det å sammenlikne differansen for $N[b, t]$ over bånd mellom suksessive tidsblokker i mange tilfeller gi en mer perseptuelt nøyaktig deteksjon av hørehendelser sammenliknet med den direkte bruken av suksessive FFT-spektre beskrevet over.

I nevnte patentsøknad er flere metoder for å modifisere lyden basert på denne psykoakustiske lydstyrkemodellen beskrevet. Blant disse er flere dynamikkprosesseringsalgoritmer, så som AGC og DRC. Disse beskrevne algoritmene kan dra nytte av å bruke hørehendelser til å styre forskjellige tilhørende parametere. Siden spesifikk lydstyrke allerede er beregnet er de lett tilgjengelige for detektering av nevnte hendelser. Detaljer av en foretrukket utførelsesform er beskrevet nedenfor.

25 ***Styring av audiodynamikkprosesseringsparametere med hørehendelser***

To eksempler på utførelser av oppfinnelsen vil nå bli vist. Den første beskriver bruk av hørehendelser til å styre releasetiden i en digital utførelse av en DRC (Dynamic Range Controller) der forsterkningsreguleringen er avledet fra RMS-(Root Mean Square)-effekten i signalet. Den andre utførelsen beskriver bruk av hørehendelser til å styre bestemte aspekter ved en mer avansert kombinasjon av AGC og DRC realisert i forbindelse med den psykoakustiske lydstyrkemodellen beskrevet over. Disse to utførelsene er kun ment å tjene som eksempler på oppfinnelsen, og det må forstås at bruk av hørehendelser til å styre parametere i

en dynamikkprosesseringsalgoritme ikke er begrenset til detaljene beskrevet nedenfor.

Dynamisk spekterkontroll

5 Den beskrevne digitale utførelsen av en DRC deler inn et audiosignal $x[n]$ i vindusbehandlede, halv-overlappende blokker, og for hver blokk beregnes en modifikasjonsforsterkning basert på et mål for signalets lokale effekt og en valgt kompresjonskurve. Forsterkningen blir glattet over blokker og så multiplisert med hver blokk. De modifiserte blokkene blir til slutt overlapp-addert for å generere
10 det modifiserte audiosignalet $y[n]$.

Det skal bemerkes at selv om høresceneanalysen og den digitale realiseringen av DRC som beskrevet her deler opp tidsdomene-audiosignalet i blokker for analyse og behandling, DRC-behandlingen ikke trenger bli utført med bruk av blokksegmentering. For eksempel vil høresceneanalysen kunne utføres med bruk
15 av blokksegmentering og spektralanalyse som beskrevet over, og de resulterende hørehendelsesposisjonene og -egenskapene anvendes for å forsyne styringsinformasjon til en digital utførelse av en tradisjonell DRC som typisk jobber med ett og ett sample. Her anvendes imidlertid den samme blokkstrukturen som ble anvendt for høresceneanalyse for DRC'en for å lette beskrivelsen av kombinasjonen av dem.
20

Videre i beskrivelsen av en blokkbasert DRC-utførelse kan de overlappende blokkene i audiosignalet representeres som:

$$x[n, t] = w[n]x[n + tM / 2] \quad \text{for} \quad 0 < n < M - 1$$

25 (4)

der M er blokk lengden og hopstørrelsen er $M/2$, $w[n]$ er vinduet, n er sampleindeksen internt i blokken og t er blokkindeksen (merk at t her anvendes på samme måte som for STDFT-transformasjonen i likning 1; den representerer for eksempel tid i diskrete blokkenheter heller enn sekunder). Ideelt sett skrår
30 vinduet $w[n]$ til null i begge ender og summerer til én når det halv-overlappes med seg selv; for eksempel oppfyller det mye brukte sinus-vinduet disse kriteriene.

For hver blokk kan en da beregne RMS-effekten for å generere et effektmål $P[t]$ i dB for hver blokk:

35

$$P[t] = 10 * \log_{10} \left(\frac{1}{M} \sum_{n=1}^M x^2[n, t] \right)$$

(5)

Som nevnt tidligere kan en glatte dette effektmålet med et raskt attack og
 5 langsom release før behandling med en kompresjonskurve, men som et alternativ
 kan momentaneffekten $P[t]$ bli behandlet og den resulterende forsterkningen
 glattet. Denne alternative løsningen har den fordelen at en enkel
 kompresjonskurve med skarpe knekkpunkter kan anvendes, men de resulterende
 forsterkningene er likevel glatte mens effekten passerer gjennom knekkpunktet.
 10 Med en kompresjonskurve representert som vist i figur 8c som en funksjon F av
 signalnivå som gir en forsterkning, er blokkforsterkningen $G[t]$ gitt ved:

$$G[t] = F\{P[t]\}$$

(6)

15 Antatt at kompresjonskurven anvender større demping etter hvert som signal-
 nivået øker, vil forsterkningen avta når signalet er i "attackmodus" og øke når
 signalet er i "releasemodus". Følgelig kan en glattet forsterkning $\bar{G}[t]$ beregnes i
 henhold til:

$$20 \quad \bar{G}[t] = \alpha[t] \cdot \bar{G}[t-1] + (1 - \alpha[t])G[t]$$

(7a)

der

$$\alpha[t] = \begin{cases} \alpha_{attack} & G[t] < \bar{G}[t-1] \\ \alpha_{release} & G[t] \geq \bar{G}[t-1] \end{cases} .$$

25 (7b)

og

$$\alpha_{release} \gg \alpha_{attack} \quad (7c)$$

Til slutt blir den glattede forsterkningen $\bar{G}[t]$, som er i dB, anvendt på hver blokk
 30 av signalet, og de modifiserte blokkene blir overlapp-addert for å generere den
 modifiserte lyden:

$$y[n + tM/2] = (10^{\bar{G}[t]/20})x[n, t] + (10^{\bar{G}[t-1]/20})x[n + M/2, t-1] \quad \text{for } 0 < n < M/2$$

(8)

5 Merk at fordi blokkene er multiplisert med et avskrådd vindu, som vist i likning 4, overlapp-adderingen vist over effektivt glatter forsterkningene på tvers av sampler av det behandlede signalet $y[n]$. Følgelig gjennomgår forsterkningsreguleringssignalet glatting i tillegg til den vist i likning 7a. I en mer tradisjonell utførelse av DRC som jobber med enkeltsampler heller enn blokkvis, kan en mer avansert forsterkningsglatting enn det enkle én-polfilteret vist i likning 7a være nødvendig for å hindre hørbar fordreining av det behandlede signalet. Videre introduserer bruken av blokkbasert behandling en naturlig forsinkelse på $M/2$ sampler i systemet, og så lenge reduksjonstiden for α_{attack} er nær denne forsinkelsen trenger ikke signalet $x[n]$ forsinkes ytterligere før anvendelse av forsterkningene for å hindre overshoot.

Figurene 9a til 9c viser resultatet av bruk av den beskrevne DRC-behandlingen på et audiosignal. For denne konkrete utførelsen anvendes en blokk lengde på $M=512$ med en samplingsrate på 44,1 kHz. En kompresjonskurve tilsvarende den vist i figur 8b anvendes:

20 over -20dB i forhold til det digitale fullskalasignalet dempes med et forhold på 5:1, og under -30dB blir signalet styrket med et forhold på 5:1. Forsterkningen blir glattet med en attackkoeffisient α_{attack} svarende til en halvreduksjonstid på 10ms og en releasekoeffisient $\alpha_{release}$ svarende til en halvreduksjonstid på 500ms. Det opprinnelige audiosignalet vist i figur 9a består av seks etterfølgende piano-akkorder, med den endelige akkorden liggende rundt sampelet $1,75 \times 10^5$ og avtagende til lydløshet. I betraktningen av plottet av forsterkningen $\bar{G}[t]$ i figur 9b skal det bemerkes at forsterkningen holder seg nær 0dB mens de seks akkordene blir spilt. Dette er fordi signalenergien i hovedsak holder seg mellom -30dB og -20dB, området innenfor hvilket DRC-kurven ikke krever noen modifikasjon. Etter den siste akkorden faller imidlertid signalenergien under -30dB, og forsterkningen begynner å øke, til slutt til over 15dB, etter hvert som akkorden dør hen. Figur 9c viser det resulterende modifiserte audiosignalet, og en kan se at halen til den endelige akkorden er betydelig styrket. Hørselsmessig gir

denne styrkingen av akkordens naturlige, lett avtagende lyd et ekstremt unaturlig resultat. Det er målet med foreliggende oppfinnelse å hindre problemer av denne typen som er forbundet med en tradisjonell dynamikkprosessor.

Figurene 10a til 10c viser resultatene av bruk av eksakt samme DRC-system på et annet audiosignal. I dette tilfellet består den første halvdel av et opptempo musikkstykk på et høyt nivå, og så omtrent ved sample 10×10^4 skifter signalet til et andre opptempo musikkstykk, men på et betydelig lavere nivå. Ved å betrakte forsterkningen i figur 6b ser en at signalet dempes med omtrent 10dB under den første halvdel, og så stiger forsterkningen opp til 0dB under den andre halvdel når det saktere stykket spilles. I dette tilfellet oppfører forsterkningen seg som ønsket. En ønsker at det andre stykket styrkes i forhold til det første, og forsterkningen bør øke raskt etter overgangen til det andre stykket for ikke å gi nevneverdige hørbare effekter. En ser en forsterkningsoppførsel som er tilsvarende den til det første signalet beskrevet, men her er oppførselen ønskelig. Derfor ønsker en å korrigere det første tilfellet uten å påvirke det andre. Bruk av hørehendelser for å styre releasetiden til dette DRC-systemet gir en slik løsning.

I det første signalet som ble undersøkt i figur 9 synes økningen i reduksjonen av den siste akkorden unaturlig fordi akkorden og dens reduksjon oppfattes som én enkelt hørehendelse, hvis integritet forventes å være bevart. I det andre tilfellet opptrer imidlertid mange hørehendelser mens forsterkningen øker, noe som innebærer at det er lite endring for hver enkelthendelse. Den totale forsterkningsendringen er derfor ikke så ille. En kan derfor argumentere med at en forsterkningsendring kun bør tillates tidsmessig nær en hørehendelsesgrense. En kunne anvende dette prinsippet på forsterkningen mens den er i enten attack- eller releasemodus, men for de fleste praktiske utførelser av en DRC beveger forsterkningen seg så raskt i attackmodus sammenliknet med menneskets tidsopløsning av hendelsesoppfatning at ingen styring er nødvendig. En kan derfor anvende hendelser for å styre glattingen av DRC-forsterkningen kun i releasemodus.

En passende oppførsel for releasestylingen vil nå bli beskrevet. Kvalitativt, dersom en hendelse detekteres, blir forsterkningen glattet med releasetidskonstanten spesifisert over i likning 7a. Etter hvert som tiden går etter den detekterte hendelsen, og dersom ingen påfølgende hendelser detekteres, øker releasetidskonstanten kontinuerlig slik at den glattede forsterkningen til slutt "fryses". Dersom en annen hendelse detekteres, tilbakestilles glattingstidskonst-

anten til den opprinnelige verdien og prosessen gjentas. For å modulere release-tiden kan en først generere et reguleringssignal basert på de detekterte hendelsesgrensene.

Som beskrevet over kan hendelsesgrenser detekteres ved å se etter endringer i suksessive spektre i audiosignalet. Med denne konkrete metoden kan DFT-transformasjonen av hver overlappende blokk $x[n, t]$ beregnes for å generere STDFT-transformasjonen av audiosignalet $x[n]$:

$$X[k, t] = \sum_{n=0}^{M-1} x[n, t] e^{-j \frac{2\pi kn}{M}}$$

10 (9)

Deretter kan forskjellen mellom de normaliserte log-absoluttverdispektrene til suksessive blokker beregnes i henhold til:

$$15 \quad D[t] = \sum_k |X_{NORM}[k, t] - X_{NORM}[k, t-1]|$$

(10a)

der

$$20 \quad X_{NORM}[k, t] = \log \left(\frac{|X[k, t]|}{\max_k \{|X[k, t]|\}} \right)$$

(10b)

Her anvendes maksimum av $|X[k, t]|$ over bokser k for normalisering, selv om én kunne ha anvendt andre normaliseringsfaktorer; for eksempel gjennomsnittet av $|X[k, t]|$ over bokser. Dersom forskjellen $D[t]$ overstiger en terskel D_{\min} , anses en hendelse å ha forekommet. Videre kan en tildele en styrke til denne hendelsen, som ligger mellom null og én, basert på størrelsen til $D[t]$ sammenliknet med en maksimal terskel D_{\max} . Det resulterende hørehendelse-styrkesignalet $A[t]$ kan beregnes som:

25

$$A[t] = \begin{cases} 0 & D[t] \leq D_{\min} \\ \frac{D[t] - D_{\min}}{D_{\max} - D_{\min}} & D_{\min} < D[t] < D_{\max} \\ 1 & D[t] \geq D_{\max} \end{cases} \quad (11)$$

Ved å tildele en styrke til hørehendelsene som er proporsjonal med graden av spektralendring forbundet med denne hendelsen oppnår en bedre kontroll over dynamikkprosesseringen sammenliknet med en todelt hendelsesbestemmelse. Oppfinnerne har funnet at større forsterkningsendringer er akseptabelt under sterkere hendelser, og signalet i likning 11 muliggjør slik variabel regulering.

Signalet $A[t]$ er et impulssignal der en impuls opptrer ved en hendelsesgrense. For å styre releasetiden kan en videre glatte signalet $A[t]$ slik at det avtar kontinuerlig til null etter deteksjon av en hendelsesgrense. Det glattede hendelsesreguleringssignalet $\bar{A}[t]$ kan beregnes fra $A[t]$ i henhold til:

$$\bar{A}[t] = \begin{cases} A[t] & A[t] > \alpha_{event} \bar{A}[t-1] \\ \alpha_{event} \bar{A}[t-1] & \text{ellers} \end{cases} \quad (12)$$

15

Her styrer α_{event} reduksjonstiden til hendelsesreguleringssignalet. Figurene 9d og 10d viser hendelsesreguleringssignalet $\bar{A}[t]$ for de to motsvarende audio-signalene, med halvreduksjonstiden til det glatteste satt til 250ms. I det første tilfellet ser en at en hendelsesgrense detekteres for hver av de seks piano-akkordene, og at hendelsesreguleringssignalet avtar kontinuerlig mot null etter hver hendelse. For det andre signalet detekteres mange hendelser veldig nær hverandre i tid, og derfor avtar hendelsesreguleringssignalet aldri helt til null.

20

En kan nå bruke hendelsesreguleringssignalet $\bar{A}[t]$ til å variere releasetidskonstanten anvendt for å glatte forsterkningen. Når reguleringssignalet er lik én, er glattingskoeffisienten $\alpha[t]$ fra likning 7a lik $\alpha_{release}$ som før, og når reguleringssignalet er lik null er koeffisienten lik én, slik at en hindrer at den glattede forsterkningen endrer seg. Glattingskoeffisienten interpoleres mellom disse to ekstrempunktene ved hjelp av reguleringssignalet i henhold til:

25

$$\alpha[t] = \begin{cases} \alpha_{attack} & G[t] < \bar{G}[t-1] \\ \bar{A}[t]\alpha_{release} + (1 - \bar{A}[t]) & G[t] \geq \bar{G}[t-1] \end{cases}$$

(13)

5 Ved at glattingskoeffisienten interpoleres kontinuerlig som funksjon av hendelsesreguleringssignalet tilbakestilles relesetiden til en verdi som står i forhold til hendelsesstyrken ved innledning av en hendelse, og øker så glatt til uendelig etter forekomst av en hendelse. Hastigheten til denne økningen bestemmes av koeffisienten α_{event} anvendt for å generere det glattede hendelsesreguleringssignalet.

10 Figurene 9e og 10e viser innvirkningen av å glatte forsterkningen med den hendelsesstyrte koeffisienten fra likning 13 i motsetning til den ikke-hendelsesstyrte koeffisienten fra likning 7b. I det første tilfellet faller hendelsesreguleringssignalet til null etter den siste pianoakkorden, og hindrer dermed forsterkningen i å bevege seg oppover. Som følge av dette utviser ikke den tilhørende modifiserte lyden i figur 9f en unaturlig styrking av reduksjonen av akkorden. I det andre tilfellet går hendelsesreguleringssignalet aldri mot null, og derfor undertrykkes det glattede forsterkningssignalet veldig lite gjennom bruken av hendelsesreguleringen. Kurven til den glattede forsterkningen er nær identisk med den ikke-hendelsesstyrte forsterkningen i figur 10b. Dette er nøyaktig den

20 ønskede effekten.

Lydstyrkebasert AGC og DRC

Som et alternativ til tradisjonelle dynamikkprosesseringsmetoder der signalmodifikasjoner er en direkte funksjon av enkle signalmålinger så som makseffekt eller RMS-effekt, beskriver den internasjonale patentsøknaden PCT/US2005/038579 bruk av den psykoakustikk-baserte lydstyrkemodellen beskrevet over som et rammeverk for å utføre dynamikkprosessering. Flere fordeler er angitt. For det første er målinger og modifikasjoner angitt i sone-enheter, som er et mer nøyaktig mål for lydstyrkeoppfatning enn mer grunnleggende mål

25 så som makseffekt eller RMS-effekt. For det andre kan lyden bli modifisert slik at den oppfattede spektralbalansen til den opprinnelige lyden opprettholdes mens den totale lydstyrken endrer seg. På denne måten blir endringer av den totale lydstyrken mindre perseptuelt fremtredende sammenliknet med en dynamikkprosessor som for eksempel anvender en bredbåndsfosterkning for å

30

modifisere lyden. Endelig er den psykoakustiske modellen naturlig en flerbåndsmodell, og derfor kan systemet enkelt innrettes for å utføre flerbånd dynamikkprosessering for å avhjelpe de velkjente problemene med tverrspektral pumping forbundet med en bredbånd dynamikkprosessor.

5 Selv om det å utføre dynamikkprosessering i dette lydstyrkedomenet allerede gir en rekke fordeler over mer tradisjonell dynamikkprosessering, kan denne metoden forbedres ytterligere gjennom bruk av hørehendelser for å styre forskjellige parametere. Betrakt audiosegmentet som inneholder pianoakkorder som vist i 27a og den tilhørende DRC'en vist i figurene 10b og c. En kunne utføre 10 en tilsvarende DRC i lydstyrkedomenet, og i dette tilfellet, når lydstyrken til den siste pianoakkordens reduksjon økes, ville økningen være mindre fremtredende fordi spektralbalansen til den avtagende noten ble opprettholdt når økningen ble anvendt. En bedre løsning er imidlertid å ikke øke reduksjonen i det hele tatt, og derfor kan en tjenlig anvende samme prinsipp for å styre attack- og releasetider 15 med hørehendelser i lydstyrkedomenet som beskrevet over for en tradisjonell DRC.

Det lydstyrkedomene-baserte dynamikkprosesseringssystemet som nå skal beskrives består av AGC etterfulgt av DRC. Målet med denne kombinasjonen er å gjøre at all behandlet lyd får omtrent samme oppfattede lydstyrke samtidig som i 20 hvert fall noe av den opprinnelige lydens dynamikk bevarer. Figur 11 viser et passende sett av AGC- og DRC-kurver for denne søknaden. Merk at innmatingen til og utmatingen fra begge kurvene er representert i sone-enheter siden behandlingen blir utført i lydstyrkedomenet. AGC-kurve jobber for å bringe den utmatede lyden nærmere et målnivå, og gjør som tidligere nevnt dette med forholdsvis 25 langsomme tidskonstanter. En kan tenke på AGC som å gjøre lydstyrken på en langsom tidsskala lik den ønskede, men på en hurtig tidsskala kan lydstyrken fluktuere betydelig rundt dette målet. En kan derfor anvende mer hurtigvirkende DRC for å begrense disse fluktusjonene til et område som anses som akseptabelt for den aktuelle anvendelsen. Figur 11 viser en slik DRC-kurve der AGC-målet 30 faller innenfor "nullbåndet" til DRC, den delen av kurven som ikke krever noen modifikasjon. Med denne kombinasjon av kurver bringer AGC lydstyrken på den langsomme tidsskalaen innenfor nullbåndet til DRC-kurven, slik at kun et minimum av hurtigvirkende DRC-modifikasjoner er nødvendig. Dersom lydstyrken på den hurtige tidsskalaen fortsatt fluktuerer utenfor nullbåndet, tjener DRC da til å flytte 35 lydstyrken mot dette nullbåndet. Som en generell kommentar til slutt kan en

anvende langsomtvirkende AGC slik at alle båndene i lydstyrkemodellen gjennomgår samme mengde lydstyrkemedifikasjon og dermed opprettholder den oppfattede spektralbalansen, og en kan anvende hurtigvirkende DRC på en måte som gjør at lydstyrkemedifikasjonen kan variere over bånd for å undertrykke
 5 tverrspektral pumping som ellers vil kunne resultere fra hurtigvirkende bånd-uavhengig lydstyrkemedifikasjon.

Hørehendelser kan anvendes for å styre attack og release i både AGC og DRC. For AGC er både attack- og releasetiden stor sammenliknet med tidsoppløsningen av hendelsesoppfatningen, og derfor kan hendelsesregulering tjenlig
 10 anvendes i begge tilfeller. For DRC er attacktiden forholdsvis kort, og derfor kan hendelsesregulering kun være nødvendig for release som i den tradisjonelle DRC beskrevet over.

Som beskrevet over kan en anvende det spesifikke lydstyrkespekteret for den anvendte lydstyrkemodellen for hendelsesdeteksjon. Et differansesignal $D[t]$,
 15 tilsvarende det i likningene 10a og b, kan beregnes fra den spesifikke lydstyrken $N[b, t]$, definert i likning 2, som følger:

$$D[t] = \sum_b |N_{NORM}[b, t] - N_{NORM}[b, t - 1]|$$

(14a)

20

der

$$N_{NORM}[b, t] = \frac{N[b, t]}{\max_b \{N[b, t]\}}$$

(14b)

25

Her anvendes maksimum av $|N[b, t]|$ over frekvensbånd b for normalisering, selv om en vil kunne anvende andre normaliseringsfaktorer; for eksempel gjennomsnittet av $|N[b, t]|$ over frekvensbånd. Dersom differansen $D[t]$ overstiger en terskel D_{min} , anses en hendelse å ha forekommet. Differansesignalet kan da bli
 30 behandlet på samme måte vist i likningene 11 og 12 for å generere et glatt hendelsesreguleringssignal $\bar{A}[t]$ som anvendes for å styre attack- og releasetidene.

AGC-kurven vist i figur 11 kan representeres som en funksjon som tar som innmating et mål for lydstyrke og genererer en ønsket utmatet lydstyrke:

$$L_o = F_{AGC} \{L_i\}$$

5 (15a)

DRC-kurven kan representeres på tilsvarende måte:

$$L_o = F_{DRC} \{L_i\}$$

10 (15b)

For AGC er den innmatede lydstyrken et mål for lydets lydstyrke på en langsom tidsskala. En kan beregne et slikt mål ved å glatte den momentane lydstyrken $L[t]$, definert i likning 3, ved anvendelse av forholdsvis lange tidskonstanter (i størrelsesorden flere sekunder). Det har vært vist at mennesket i bedømmelsen av et audiosegments lydstyrke på en langsom tidsskala vektlegger de høyere delene mer enn de saktere, og en kan anvende en raskere attack enn release i glattingen for å simulere denne effekten. Med innlemmelsen av hendelsesregulering for både attack og release kan lydstyrken på den langsomme tidsskalaen som anvendes for å bestemme AGC-modifikasjonen derfor beregnes i henhold til:

15
20

$$L_{AGC}[t] = \alpha_{AGC}[t]L_{AGC}[t-1] + (1 - \alpha_{AGC}[t])L[t]$$

25 (16a)

der

$$\alpha_{AGC}[t] = \begin{cases} \bar{A}[t]\alpha_{AGCattach} + (1 - \bar{A}[t]) & L[t] > L_{AGC}[t-1] \\ \bar{A}[t]\alpha_{AGCrelease} + (1 - \bar{A}[t]) & L[t] \leq L_{AGC}[t-1] \end{cases} .$$

(16b)

30 I tillegg kan en beregne et tilhørende spesifikt lydstyrkespekter på en langsom tidsskala som senere vil bli anvendt for flerbånd DRC:

$$N_{AGC}[b,t] = \alpha_{AGC}[t]N_{AGC}[b,t-1] + (1 - \alpha_{AGC}[t])N[b,t]$$

(16c)

I praksis kan en velge glattingskoeffisientene slik at attacktiden er omtrent halvparten av releasetiden. Gitt lydstyrkemålet på den langsomme tidsskalaen kan en da beregne lydstyrkemedifikasjonsskaleringen for AGC som forholdet mellom utmatet lydstyrke og innmatet lydstyrke:

$$S_{AGC}[t] = \frac{F_{AGC}\{L_{AGC}[t]\}}{L_{AGC}[t]}$$

(17)

DRC-modifikasjonen kan nå beregnes fra lydstyrken etter bruk av AGC-skaleringen. Heller enn å glatte et mål for lydstyrken før bruk av DRC-kurven kan en alternativt anvende DRC-kurven på den momentane lydstyrken og deretter glatte den resulterende modifikasjonen. Dette er tilsvarende metoden beskrevet over for å glatte forsterkningen i den tradisjonelle DRC. Videre kan DRC anvendes med flere bånd, hvilket innebærer at DRC-modifikasjonen er en funksjon av den spesifikke lydstyrken $N[b,t]$ i hvert bånd b heller enn den totale lydstyrken $L[t]$. For å opprettholde den gjennomsnittlige spektralbalansen til den opprinnelige lyden kan en imidlertid anvende DRC på hvert bånd slik at de resulterende modifikasjonene har samme gjennomsnittseffekt som ville fremkomme ved bruk av DRC på den totale lydstyrken. Dette kan oppnås ved å skalere hvert bånd med forholdet mellom den totale lydstyrken på en langsom tidsskala (etter bruk av AGC-skalering) og den spesifikke lydstyrken på en langsom tidsskala, og ved å anvende denne verdien som parameter til DRC-funksjonen. Resultatet blir da reskalert med den inverse av nevnte forhold for å generere den utmatede spesifikke lydstyrken. Følgelig kan DRC-skaleringen i hvert bånd beregnes i henhold til:

$$S_{DRC}[b,t] = \frac{N_{AGC}[b,t]}{S_{AGC}[t]L_{AGC}[t]} F_{DRC} \left\{ \frac{S_{AGC}[t]L_{AGC}[t]}{N_{AGC}[t]} N[b,t] \right\}$$

(18)

AGC- og DRC-modifikasjonene kan så kombineres for å danne en total lydstyrkeskalering for hvert bånd:

$$S_{TOT}[b,t] = S_{AGC}[t]S_{DRC}[b,t] \quad (19)$$

5

Denne totale skaleringen kan så glattes over tid uavhengig for hvert bånd med rask attack og langsom release og med hendelsesregulering anvendt kun for release. Ideelt sett utføres glatting på logaritmen av skaleringen tilsvarende som forsterkningene i tradisjonell DRC som glattes i sin desibelrepresentasjon, selv om dette ikke er avgjørende. For å sikre at den glattede totale skaleringen beveger seg synkronisert med den spesifikke lydstyrken i hvert bånd, kan attack- og releasemodus bestemmes gjennom samtidig glatting av spesifikk lydstyrke:

10

$$\bar{S}_{TOT}[b,t] = \exp(\alpha_{TOT}[b,t] \log(\bar{S}_{TOT}[b,t-1]) + (1 - \alpha_{TOT}[b,t]) \log(S_{TOT}[b,t])) \quad (20a)$$

15

$$\bar{N}[b,t] = \alpha_{TOT}[b,t] \bar{N}[b,t-1] + (1 - \alpha_{TOT}[b,t]) N[b,t] \quad (20b)$$

der

20

$$\alpha_{TOT}[b,t] = \begin{cases} \alpha_{TOTattack} & N[b,t] > \bar{N}[b,t-1] \\ \bar{A}[t] \alpha_{TOTrelease} + (1 - \bar{A}[t]) & N[b,t] \leq \bar{N}[b,t-1] \end{cases} \quad (20c)$$

Endelig kan en beregne en ønsket spesifikk lydstyrke basert på den glattede skaleringen anvendt på den opprinnelige spesifikke lydstyrken

25

$$\hat{N}[b,t] = \bar{S}_{TOT}[b,t] N[b,t] \quad (21)$$

og så løse for forsterkningene $G[b,t]$ som når de anvendes på den opprinnelige eksitasjonen resulterer i en spesifikk lydstyrke lik den ønskede:

30

$$\hat{N}[b,t] = \Psi\{G^2[b,t]E[b,t]\} \quad (22)$$

Forsterkningene kan bli anvendt på hvert bånd i filterbanken som anvendes for å beregne eksitasjonen, og den modifiserte lyden kan så bli generert ved å invertere filterbanken for å generere et modifisert tidsdomene audiosignal.

5

Styring av ytterligere parametere

Selv om beskrivelsen over har fokusert på styring av attack- og releaseparametere for AGC og DRC gjennom høresceneanalyse av lyden som behandles, kan også andre viktige parametere tjenlig styres ved hjelp av ASA-resultatene. For eksempel kan hendelsesreguleringssignalet $\bar{A}[t]$ fra likning 12
 10 anvendes for å variere verdien til DRC-forholdsparameteren som anvendes for dynamisk å justere forsterkningen av lyden. Forholdsparameteren, tilsvarende attack- og releasetidsparametrene, bidrar betydelig til de perseptuelle artefaktene som introduseres av dynamiske forsterkningsreguleringer.

15

Realisering

Oppfinnelsen kan realiseres i maskinvare eller programvare, eller en kombinasjon av begge (*f.eks.* en programmerbar portmatrise). Dersom ikke annet er spesifisert, er ikke algoritmene innlemmet som del av oppfinnelsen uløselig knyttet til noen som helst spesifikk datamaskin eller annen anordning. Spesielt
 20 kan forskjellige generelle maskiner anvendes med programmer skrevet i henhold til det som er vist her, eller det kan være mer hensiktsmessig å konstruere mer spesialiserte anordninger (*f.eks.* integrert kretser) for å utføre de nødvendige fremgangsmåte-trinnene. Følgelig kan oppfinnelsen realiseres i ett eller flere dataprogrammer som kjører på ett eller flere programmerbare datasystemer som
 25 hvert omfatter minst én prosessor, minst ett datalagringsystem (omfattende volatile og ikke-volatile minner og/eller lagringselementer), minst én innmatingsanordning eller -port og minst én utmatingsanordning eller -port. Programkode anvendes for å mate inn data for å utføre funksjonene beskrevet her og generere utdata. Utdataene mates til én eller flere utmatingsanordninger, på
 30 en kjent måte.

35

Hvert slikt program kan implementeres i et hvilket som helst ønsket datamaskinspråk (omfattende maskinkode, assemblerkode eller et høynivå prosedyreorientert, logisk eller objekt-orientert programmeringsspråk) for å kommunisere med et datasystem. Språket kan være et compilert eller interpretert språk.

Hvert slikt dataprogram blir fortrinnsvis lagret på eller lastet ned til et lagringsmedium eller en lagringsanordning (*f.eks.* solid-state minne eller medier, eller magnetiske eller optiske medier) som kan leses av en generell eller spesialisert programmerbar datamaskin, for å konfigurere og kjøre datamaskinen når lagringsmediene eller -anordningen leses av datasystemet for å utføre fremgangsmåtene beskrevet her. Systemet ifølge oppfinnelsen kan også betraktes som realisert som et datamaskinlesbart lagringsmedium som inneholder et dataprogram, der lagringsmediet innrettet på denne måten bevirker et datasystem til å kjøre på en spesifikk og forhåndsdefinert måte for å utføre funksjonene beskrevet her.

Et antall utførelsesformer av oppfinnelsen er beskrevet. Ikke desto mindre vil det forstås at en kan foreta forskjellige modifikasjoner uten å fjerne seg fra oppfinnelsens ramme. For eksempel kan noen av trinnene beskrevet her være rekkefølgeuavhengige, og kan således bli utført i en annen rekkefølge enn den som er beskrevet.

Det må forstås at andre variasjoner og modifikasjoner av oppfinnelsen og dens forskjellige aspekter vil sees av fagmannen, og at oppfinnelsen ikke er begrenset av de konkrete utførelsesformene beskrevet. Det forutsettes derfor at foreliggende oppfinnelse dekker enhver modifikasjon, variasjon eller ekvivalent som faller innenfor rammen til de grunnleggende underliggende prinsippene som er beskrevet og som det kreves beskyttelse for ifølge patentkravene.

Referanser

De følgende patenter, patentsøknader og publikasjoner beskriver ytterligere kjent teknikk.

Audiodynamikkprosessering

Audio Engineer's Reference Book, redigert av Michael Talbot-Smith, 2. utgave. Limiters and Compressors, Alan Tutton, 2-1492-165. Focal Press, Reed Educational and Professional Publishing Ltd., 1999.

Deteksjon og bruk av hørehendelser

U.S.-patentsøknaden 10/474,387, "High Quality Time-Scaling and Pitch-Scaling of Audio Signals" av Brett Graham Crockett, publisert 24. juni 2004 som US 2004/0122662 A1.

U.S.-patentsøknaden 10/478,398, "Method for Time Aligning Audio Signals Using Characterizations Based on Auditory Events", av Brett G. Crockett m.fl., publisert 29. juli 2004 som US 2004/0148159 A1.

U.S.-patentsøknaden 10/478,538, "Segmenting Audio Signals Into Auditory Events", av Brett G. Crockett, publisert 26. august 2004 som US 2004/0165730

A1. Aspekter ved foreliggende oppfinnelse tilveiebringer en måte å detektere hørehendelser i tillegg til de beskrevet i nevnte søknad av Crockett.

U.S.-patentsøknaden 10/478,397, "Comparing Audio Using Characterizations Based on Auditory Events", av Brett G. Crockett m.fl., publisert 2. september 2004 som US 2004/0172240 A1.

Den internasjonale søknaden, under Patentsamarbeidkonvensjonen, PCT/US 05/24630, innlevert 13. juli 2005 med tittelen "Method for Combining Audio Signals Using Auditory Scene Analysis", av Michael John Smithers, publisert 9. mars 2006 som WO 2006/026161.

Den internasjonale søknaden, under Patentsamarbeidkonvensjonen, PCT/US 2004/016964, innlevert 27. mai 2004 med tittelen "Method, Apparatus and Computer Program for Calculating and Adjusting the Perceived Loudness of an Audio Signal" av Alan Jeffrey Seefeldt m. fl., publisert 23. desember 2004 som WO 2004/111994 A2.

Den internasjonale søknaden, under Patentsamarbeidkonvensjonen, PCT/US2005/038579, innlevert 25. oktober 2005 med tittelen "Calculating and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal", av Alan Jeffrey Seefeldt og publisert med det isjonale publikasjonsnummeret. Etersom denne søknaden ikke enda er publisert er en kopi av denne vedlagt her, og danner en del av denne søknaden.

"A method for Characterizing and Identifying Audio Based on Auditory Scene Analysis", av Brett Crockett og Michael Smithers, Audio Engineering Society konferanseartikkel 6416, 118. konferanse i Barcelona, 28-31. mai 2005.

"High Quality Multichannel Time Scaling and Pitch-Shifting using Auditory Scene Analysis", av Brett Crockett, Audio Engineering Society, konferanseartikkel 5948, New York, oktober 2003.

"A new Objective Measure of Perceived Loudness" av Alan Seefeldt m.fl., Audio Engineering Society, konferanseartikkel 6236, San Francisco, 28. oktober 2004.

Handbook for Sound Engineers, The New Audio Cyclopedia, redigert av Glen M. Ballou, 2. utgave. Dynamics, 850-851. Focal Press, et trykk av Butterworth-Heinemann, 1998.

Audio Engineer's Reference Book, redigert av Michael Talbot-Smith, 2. utgave, Seksjon 2.9 ("Limiters and Compressors" av Alan Tutton), s. 2.149-2.165, Focal Press, Reed Educational and Professional Publishing Ltd., 1999.

PATENTKRAV

1. Fremgangsmåte ved behandling av et lydsignal i et lydbehandlingsapparat, hvor fremgangsmåten omfatter:

å motta et lydsignal, lydsignalet omfatter minst én kanal med lydinnhold;

å dele lydsignalet i flere underbåndssignaler med en analyse filterbank (1-1), hvor av flerheten av underbåndssignaler omfatter minst ett underbåndseksempel;

å utlede en karakteristik av lydsignalet, hvor karakteristikken er et effektmål av lydsignalet (1-2);

å glatte effektmålingen for å generere et glattet effektmål av lydsignalet; å detektere en plassering av en hørselshendelsesgrense ved å overvåke det glattede effektmålet (1-3), hvor en lyddel mellom sammenhengende hørselshendelsesgrenser utgjør en hørselshendelse (1-4),

å generere en forsterkningsvektor basert på plasseringen av hørselshendelsen (1-4);

å anvende forsterkningsvektoren på en versjon av flerheten av underbåndssignaler for å generere modifiserte underbåndssignaler; og

å syntetisere de modifiserte underbåndssignalene med en syntese-filterbank for å produsere et modifisert lydsignal,

hvor deteksjonen videre inkluderer å anvende en terskel på det glattede effektmålet for å detektere plasseringen av hørselshendelsesgrensen, og

hvor lydbehandlingsapparatet i det minste delvis er implementert med maskinvare.

2. Fremgangsmåte ifølge krav 1, hvor karakteristikken videre inkluderer lydstyrke.

3. Fremgangsmåte ifølge krav 1, hvor karakteristikken videre inkluderer opplevd lydstyrke.

4. Fremgangsmåte ifølge krav 1, hvor karakteristikken videre inkluderer fase.

5. Fremgangsmåte ifølge krav 1, hvor karakteristikken videre inkluderer en plutselig endring i signaleffekten.

6. Fremgangsmåte ifølge krav 1, hvor lydsignalet inkluderer to eller flere kanaler med lydinnhold og hørselshendelsesgrensen blir identifisert ved å undersøke endringer i karakteristikken mellom de to eller flere kanaler.

7. Fremgangsmåte ifølge krav 6, hvor karakteristikken inkluderer interkanal-faseforskjell.

8. Fremgangsmåte ifølge krav 6, hvor karakteristikken inkluderer interkanals korrelasjon.

9. Fremgangsmåte ifølge krav 1, hvor grensen for hørselshendelser sammenfaller med en begynnelse eller slutt på en datablokk i lydsignalet.

10. Fremgangsmåte ifølge krav 1, videre omfattende å justere hørselshendelsesgrensen til å sammenfalle med en grense av en datablokk i lydsignalet.

11. Ikke-forbigående datamaskinlesbart lagringsmedium kodet med et dataprogram som forårsaker at en datamaskin utfører fremgangsmåten ifølge krav 1.

12. Lydbehandlingsapparat, der apparatet omfatter:

et inngangsgrensesnitt for mottak av et lydsignal, hvor lydsignalet omfatter minst en kanal med lydinnhold;

en analysefilterbank for å dele lydsignalet i et antall underbåndssignaler (1-1), hvor hvert av flerheten av underbåndssignaler inkluderer minst ett underbåndseksempel;

en prosessor som:

utleder en karakteristikk av lydsignalet, hvor karakteristikken er et effektmål av lydsignalet (1-2);

glatter effektmålet for å generere et glattet effektmål av lydsignalet;

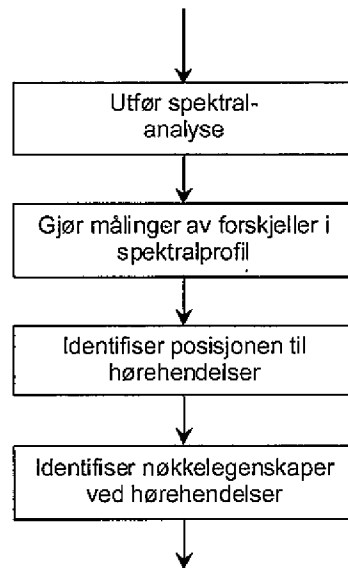
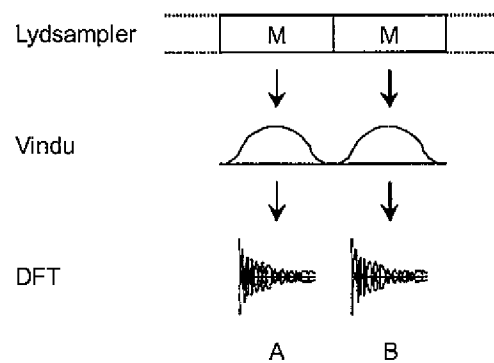
detekterer en plassering av en hørselshendelsesgrense ved å overvåke det glattede effektmålet (1-3), hvor en lyddel mellom påfølgende hørselshendelsesgrenser utgjør en hørselshendelse,

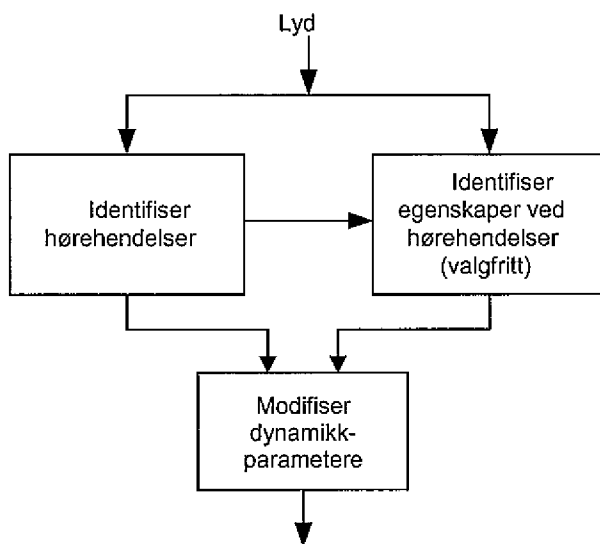
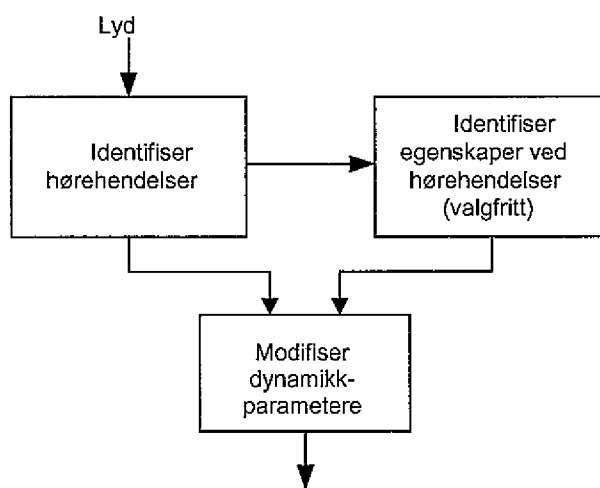
genererer en forsterkningsvektor basert på plasseringen av hørselshendelsesgrensen (1-4), og

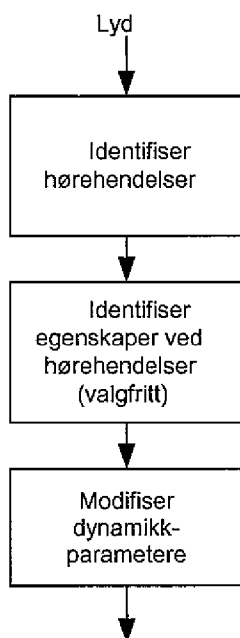
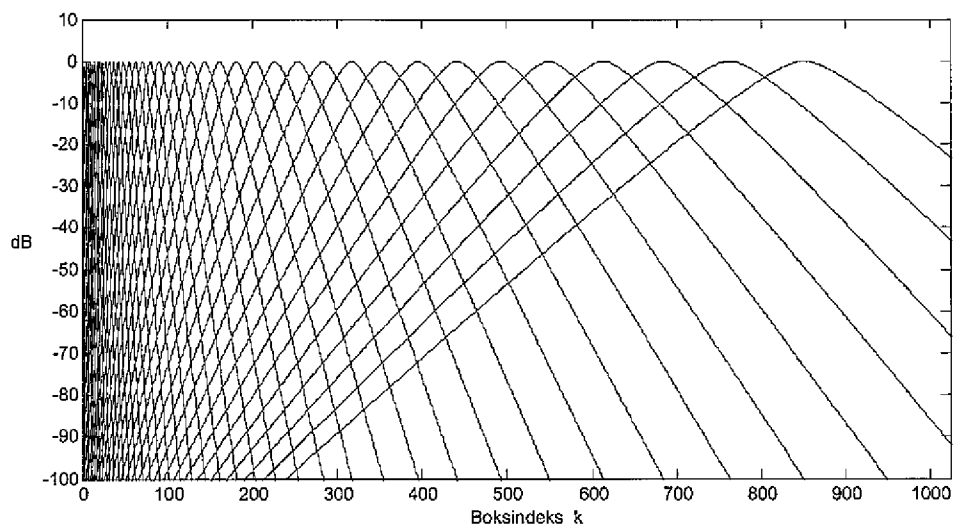
bruker forsterkningsvektoren på en versjon av flerheten av underbåndssignaler for å produsere modifiserte underbåndssignaler; og

en syntese-filterbank som kombinerer de modifiserte underbåndssignalene for å produsere et modifisert lydsignal, og

hvor deteksjonen videre inkluderer å anvende en terskel på det glattede effektmålet for å detektere plasseringen av hørselshendelsesgrensen.

**FIG. 1****FIG. 2**

**FIG. 3****FIG. 4**

**FIG. 5****FIG. 6**

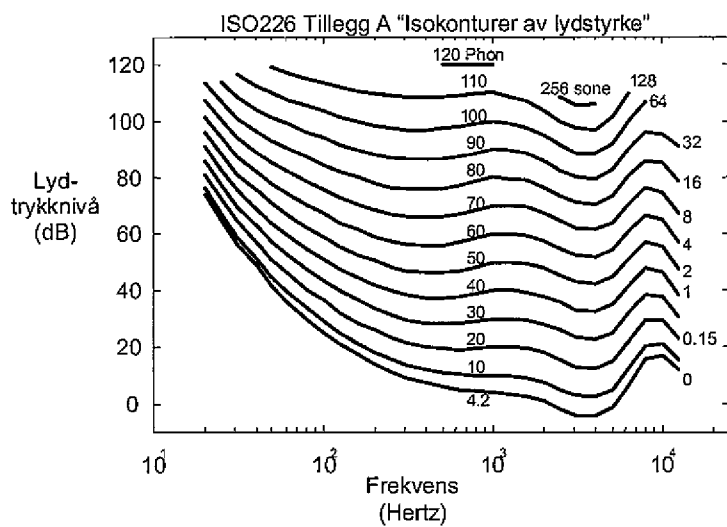


FIG. 7

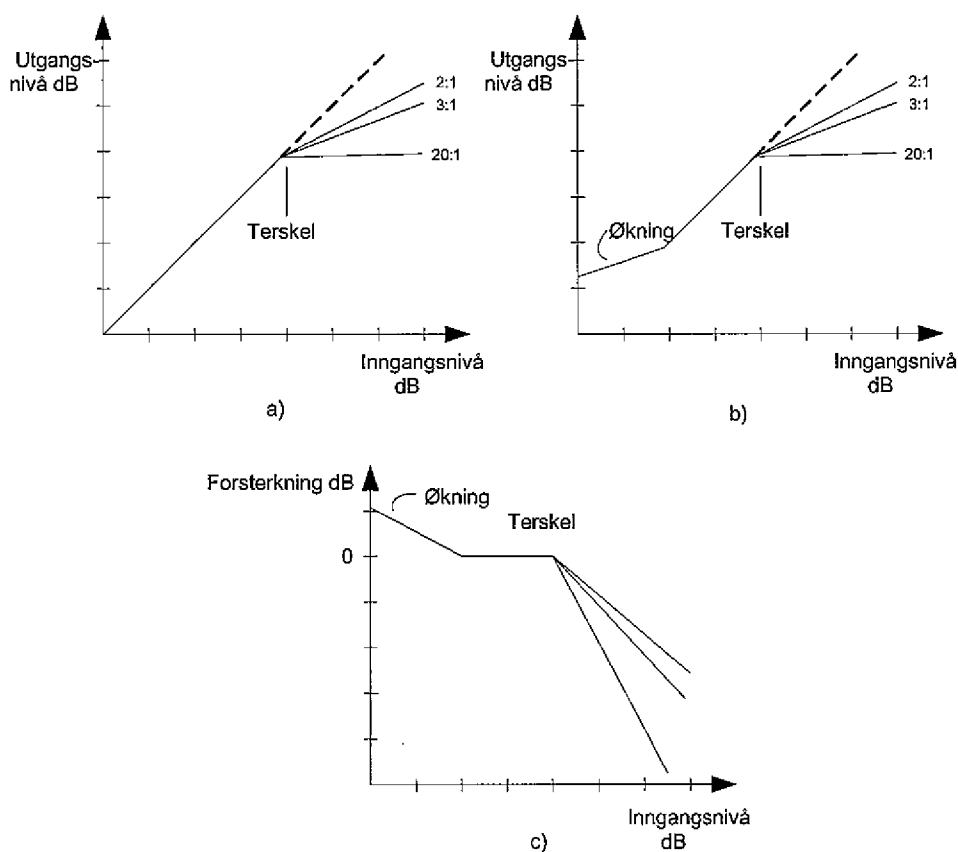


FIG. 8

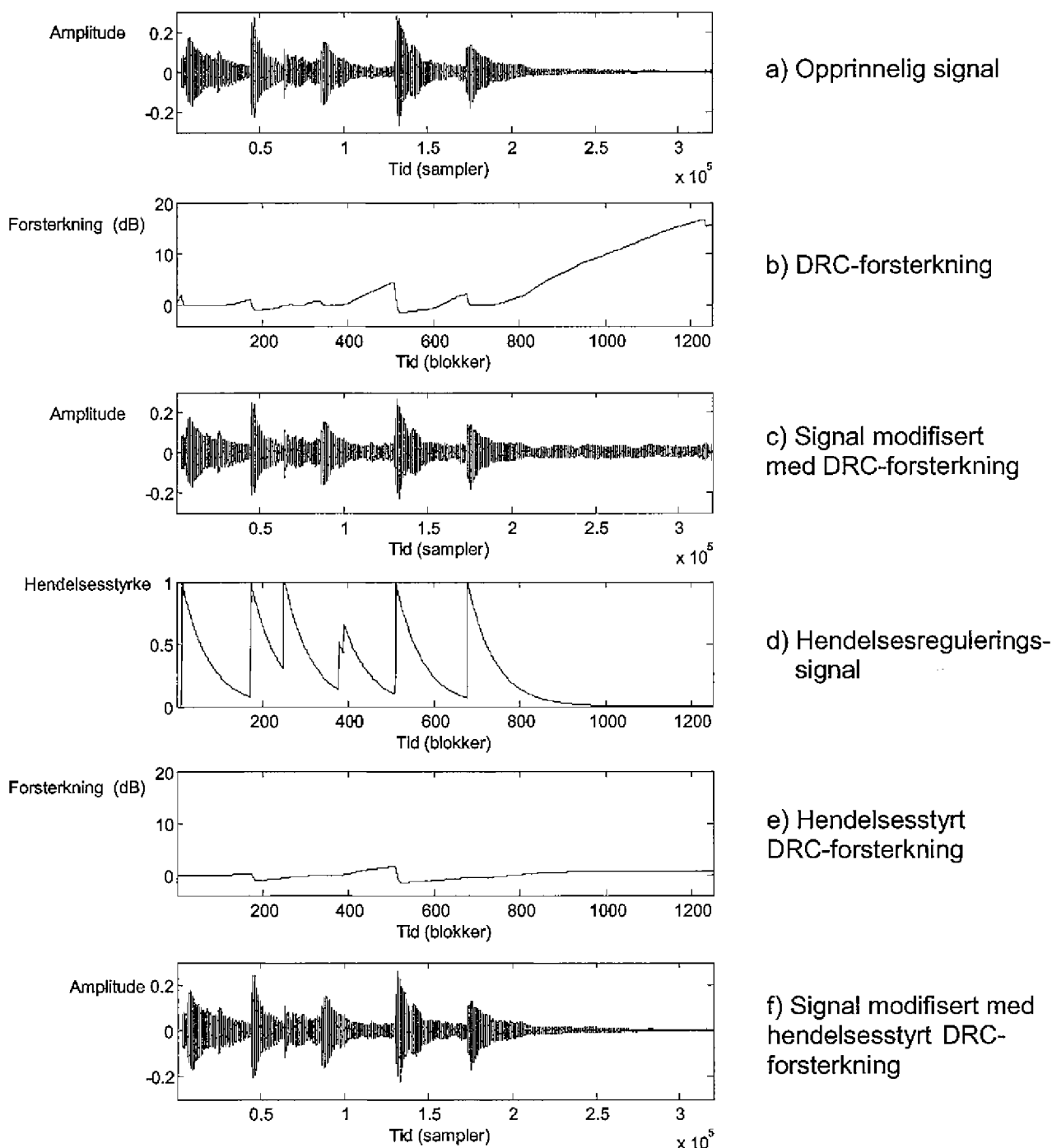


FIG. 9

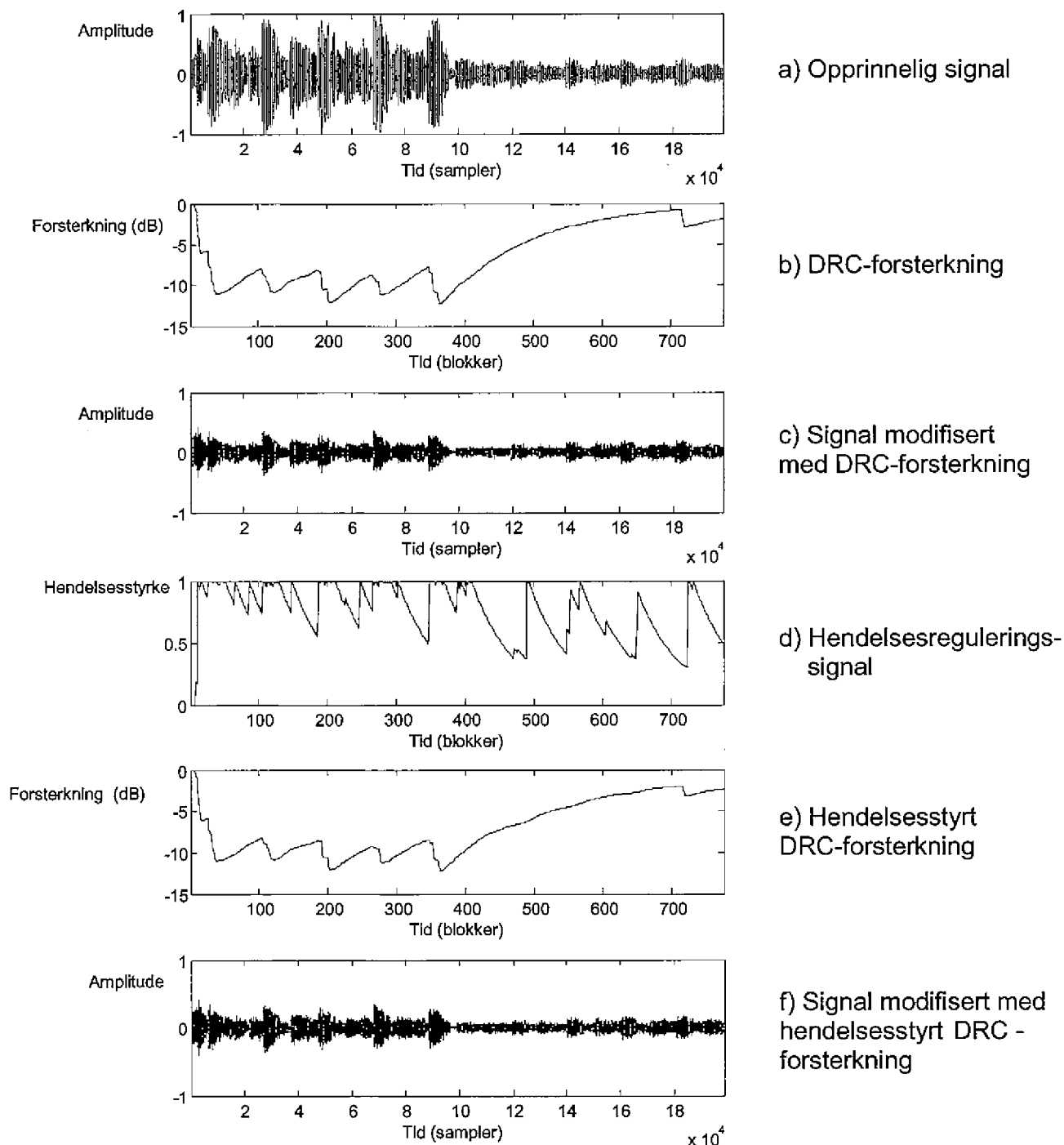
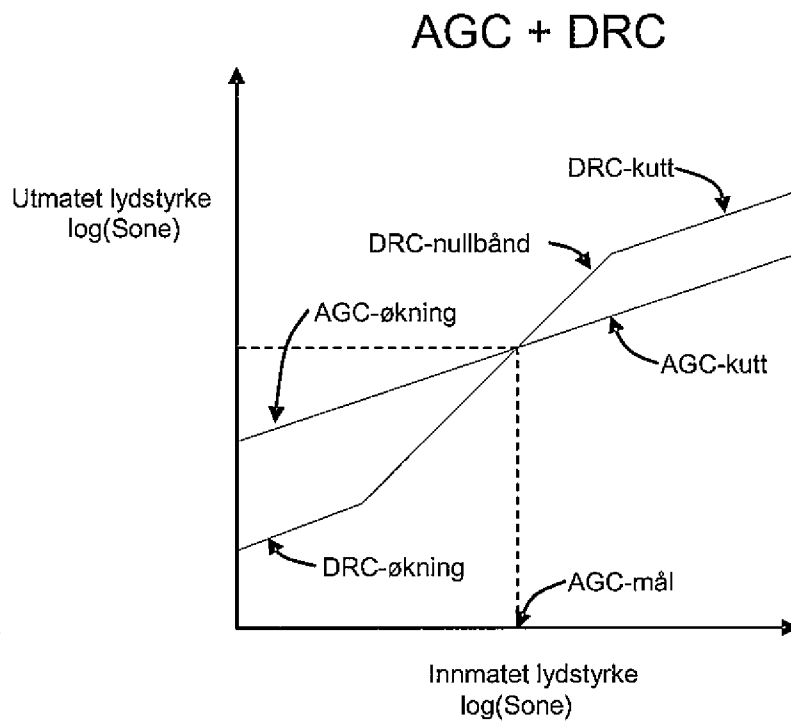


FIG. 10

**FIG. 11**