

Nederlands Elektronica- en Radiogenootschap

Correspondentie adres: postbus 39, 2260 AA Leidschendam.
Gironummer 94746 t.n.v. Penningmeester NERG, Leidschendam.

HET GENOOTSCHAP

Het Nederlands Elektronica- en Radiogenootschap (NERG) is een wetenschappelijke vereniging, welke zich ten doel stelt de kennis en het wetenschappelijk onderzoek op het gebied van de elektronica, signaalbewerking, communicatie- en informatietechnologie te bevorderen en de verbreiding en toepassing van die kennis te stimuleren.

(Internet WWW-site:
<http://www.nerg.nl>)

BESTUUR

Prof.dr.ir. W.C. van Etten, voorzitter
ir. W. van der Bijl, vice-voorzitter
dr. M.J.C. van den Homberg,
secretaris
ir. O.B.P. Rikkert de Koe,
penningmeester
prof.dr.ir. A.P.M. Zwamborn,
programma-manager
dr.ir. W.M.C. Dolmans
dr.ir. G.J.M. Janssen
G. van der Schouw,
voorzitter Onderwijscommissie
dr.ir. A.B. Smolders,
tijdschrift-manager

LIDMAATSCHAP

Voor het lidmaatschap wende men zich via het correspondentie-adres tot de secretaris. Het lidmaatschap van het NERG staat open voor hen, die aan een universiteit of hogeschool zijn afgestudeerd en die die door hun kennis en ervaring bij kunnen dragen aan het genootschap. De contributie wordt geheven per kalenderjaar en is inclusief abonnement op het Tijdschrift van het NERG en deelname aan vergaderingen, lezingen en excursies. De jaarlijkse contributie bedraagt voor gewone leden f 75,- en voor

studentleden f 39,-. Bij automatische incasso wordt f 3,- korting verleend. Gevorderde 1e fase studenten en 2e fase studenten aan een technische universiteit komen in aanmerking voor het studentlidmaatschap en kunnen daartoe contact opnemen met de contactpersoon op hun universiteit. In bepaalde gevallen kunnen ook andere leden, na overleg met de penningmeester voor een gereduceerde contributie in aanmerking komen.

HET TIJDSCHRIFT

Het tijdschrift verschijnt gemiddeld vijf maal per jaar. Opgenomen worden artikelen op het gebied van de elektronica en de telecommunicatie. Auteurs, die publicatie van hun onderzoek in het tijdschrift overwegen, wordt verzocht vroegtijdig contact op te nemen met de hoofdredacteur of een lid van de Tijdschriftcommissie.

Toestemming tot overnemen van artikelen of delen daarvan kan uitsluitend worden gegeven door de redactiecommissie. Alle rechten worden voorbehouden.

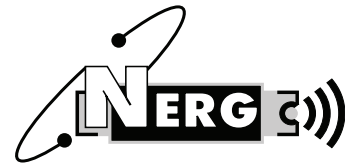
REDACTIECOMMISSIE

Dr. Ir. A.B. Smolders, voorzitter.
ASTRON, Postbus 2, 7990 AA Dwingeloo,
E-mail: smolders@nfra.nl
Ing. A.A. Spanjersberg, Hoofdredacteur, E-mail: aaspan@wxs.nl
Ir. G.W. Kant, ASTRON Dwingeloo,
E-mail: kant@nfra.nl
Ir. H.J. Visser, TNO-FEL, Den Haag,
E-mail: vissser@fel.tno.nl

Deze uitgave van het NERG wordt verzorgd door:



ISSN 03743853



INHOUD

Van de voorzitter	2
<i>Wim van Etten</i>	
Van de redactie	3
“Digitale codering is niet saai”	4
<i>Interview Prof. Dr. Ir. Kees A. Schouhamer Immink</i>	
<i>Door Bart Smolders</i>	
50 jaar Shannon Theorie	7
<i>J. Pieter M. Schalkwijk</i>	
Adaptive Filtering Approach to Multichannel Audio Reproduction Systems	13
<i>John Garas and Piet Sommen</i>	
Digitale signaalbewerking in hoortoestellen	26
<i>Jos Leenen</i>	
Verliesvrij coderen van audiosignalen	38
<i>Fons Bruekers</i>	
Uitreiking Vederprijs	48
Slechthorendheid en soorten hoorapparaten	50
<i>Ing. Peter Termeeer</i>	
Cursus aankondigingen, Conferenties etc	53

Uitgever: ing. D.J. Wassink
Advertenties: M. de Boer
tel: (0570) 648718
E-mail: MdeBoer@kluwer.nl
fax.: (0570) 619904
Opmaak: Henk Visscher, Zutphen

Van de voorzitter

Wim van Etten



Voor u ligt het eerste nummer van de jaargang 2000 van Het Tijdschrift van het NERG.

Het is u ongetwijfeld opgevallen, dat het er anders uitziet dan de vorige jaargang. U zult zich wellicht afvragen: "Alweer anders?" Dat heeft denk ik een verklaring nodig. Twee jaar geleden zijn we begonnen met de vernieuwing van het Tijdschrift, vooral de presentatie is sindsdien veel verbeterd. Dat is tot uiting gekomen in een aantrekkelijker kaftontwerp in kleur, betere papierkwaliteit en betere kwaliteit van het drukwerk. Deze lijn zetten we in principe voort, ook richting steeds verdere verbetering van presentatie en inhoud. Weer een nieuw kaftontwerp vond het bestuur nodig in verband met uniforme presentie van alle NERG-uitingen. Reeds op de vorige ALV werd door enkele leden geklaagd over gebrek daaraan, en terecht. Het bestuur had dat zelf ook al geconstateerd en er was reeds een actie in gang gezet om dit alles de uniformeren en in een nieuw, modern jasje te stoppen. Onlangs heeft u daarvan kennis kunnen nemen. U heeft een mailing ontvangen, die o.a. een ledenlijst bevatte, een nieuwe folder en tevens de pre-

sentatie van ons nieuwe logo. Ook degenen onder u die wel eens de moeite nemen om de NERG-website te bezoeken, hebben geconstateerd, dat ook daar de presentatie in deze lijn is gebracht. Tevens is daar de service uitgebreid; ook daar zullen we nog verder uitbreiding aan geven.

Hiermee is de vernieuwing van de NERG-uitingen in zekere zin en voorlopig voltooid. Ik zeg in zekere zin, omdat dit uitsluitend slaat op de folder, het logo en nu dus de kaft van het Tijdschrift. Wat het binnenwerk betreft, blijft de tijdschriftcommissie steeds zoeken naar nieuwe vormen en wegen om het geheel voor de leden (en ook niet-leden, maar geïnteresseerde lezers) aantrekkelijker te maken. In die zin zal dit voor het tijdschrift nog niet de laatste wijziging zijn. Wel hopen we naar een meer uniforme en zich stabiliserende situatie te kunnen toe groeien.

Het bestuur is de tijdschriftcommissie en de hoofdredacteur zeer erkentelijk voor het vele werk dat zij hebben verricht om tot dit fraaie resultaat te komen. Ik kan u verzekeren dat dit heel moeilijk is voor een commissie die verspreid zit over het hele land en

waarvan de leden allemaal een drukke baan hebben, waar de werkgevers tegenwoordig het nodige vergen van de mensen die zij op de loonlijst hebben staan.

Tot slot wil ik alle leden oproepen om niet alleen passief, maar zo mogelijk vooral ook actief te blijven participeren in de NERG-activiteiten, in dit verband dan nu speciaal ook het tijdschrift. Een oproep dus om artikelen toe te sturen aan de redactie, of nieuws te melden, dat mogelijkerwijs van belang kan zijn voor de leden van het NERG.



Van de redactie



Tijdschrift van het NERG krijgt een nieuwe gezicht

Het beleid van het NERG bestuur is er in het afgelopen jaar op gericht geweest om het NERG nieuw leven in te blazen. Dat is hard nodig om te voorkomen dat het NERG uitsterft. Dit laatste zou men ook letterlijk kunnen opvatten. Naast de themabijeenkomsten is het Tijdschrift het gezicht van het NERG. Het is dan ook belangrijk dat het Tijdschrift jonge ingenieurs aanspreekt. Niet alleen wat betreft inhoud maar ook wat betreft uiterlijk. Uiteraard moet het Tijdschrift de huidige NERG leden ook blijven boeien. Zoals U kunt zien is het uiterlijk en de opmaak van dit nummer duidelijk anders dan voorheen. De nieuwe opmaak is in nauwe samenwerking met een professioneel ontwerp bureau tot stand gekomen. Over smaak valt niet te twisten, maar de nieuwe opmaak oogt duidelijk frisser en levendiger dan die van de afgelopen jaren. Verder hebben is er gestreeft naar een grotere uniformiteit tussen de artikelen en rubrieken.

Er zijn binnen de redactie en binnen het NERG-bestuur vaak discussies over hoe het Tijdschrift zich zou moeten positioneren op de markt. Hierbij zijn er twee uitersten te onderscheiden. Enerzijds zijn er de echte weten-

schappelijke kwaliteitsbladen van IEEE en IEE, zoals de Transactions van de diverse societities. Anderzijds zijn er de meer populaire technische bladen, waaronder de Magazines van de diverse IEEE societities, de KIVI bladen en de vele gratis tijdschriften die iedereen wel eens ontvangt. Het is duidelijk dat het Tijdschrift niet moet proberen zich te meten met de Transactions van IEEE. Daarvoor is Nederland gewoonweg te klein. Verder zal het gemiddelde NERG-lid daar ook niet op zitten te wachten. De redactie is van mening dat het Tijdschrift zou moeten bestaan uit een mix van wetenschappelijke artikelen voor de echte specialisten en artikelen en rubrieken die toegankelijk en interessant zijn voor een veel grotere groep NERG leden. Bedenk hierbij dat het NERG sinds vorig jaar ook HBO ingenieurs uitnodigt om lid te worden. De doelgroep voor het Tijdschrift zal hierdoor breder worden in de komende jaren.

De redactie heeft het komende jaar er voor uit getrokken om de inhoud en de opmaak van het Tijdschrift verder te verbeteren. Hiertoe willen we meer rubrieken gaan invoeren om de leesbaarheid te verbeteren. Een van die nieuwe rubrieken betreft interviews met prominente personen uit de Nederlandse elek-

trotechnische gemeenschap. Hiervan vindt U al een voorbeeld in dit nummer. Graag zouden wij van de lezers willen horen wat hun mening is over dit nummer, zowel wat betreft inhoud en opmaak van de artikelen. Ook al vindt U het helemaal niets, willen we dit graag horen. Tevens zijn wij erg benieuwd naar suggesties om het tijdschrift nog verder te verbeteren, bijvoorbeeld met nieuwe rubrieken. Bij voorkeur ontvangen wij Uw reacties per E-mail met het adres:

Reacties over nummer 1:
Smolders@nfra.nl

U kunt Uw reactie ook opsturen naar het postadres van de redactie.

Het eerste nummer van het jaar 2000 is gewijd aan onderwerpen uit de informatietheorie en signaalprocessing, met bijdrages van Universiteiten en uit het bedrijfsleven. De redactie is bijzonder verheugd met het artikel van Prof. Pieter Schalkwijk van de TUE over 50 jaar Shannon. Het geeft een mooi overzicht van de ontwikkelingen in dit relatief jonge vakgebied. Verder vindt U in dit nummer een interview met een zeer markante persoonlijkheid uit dit vakgebied, Kees Schouhamer-Immink.

“Digitale codering is niet saai”

Interview Prof. Dr. Ir. Kees A. Schouhamer Immink

Door Bart Smolders



Onlangs ontving Prof.dr.ir. Kees Schouhamer Immink de IEEE *Edison Medal* voor zijn grensverleggende werk op het gebied van digitale video, audio en data recording. Als jonge ingenieur was hij in de jaren 70 bij het Natuurkundig Laboratorium (Natlab) van Philips Research in Eindhoven een van de grondleggers van de Compact Disc (CD), en had hij een belangrijke bijdrage in de ontwikkeling van vele andere digitale recording-systemen zoals de CD-ROM, CD-I, MiniDisc, CD-Video, DCC, DVD, enz. Hij heeft bijna 40 octrooien op diverse gebieden zoals codeer-systemen, acoustiek, elektronica, servosystemen en mechanica op zijn naam staan. Een aantal van deze octrooien zijn zo essentieel voor sommige producten dat Philips er nog steeds erg veel geld aan verdient. Een overzichtsartikel van de hand van Kees Schouhamer Immink verscheen recent in het Tijdschrift van het NERG [1]. Rondom de uitreiking van de IEEE *Edison Medal* was er aandacht vanuit de landelijke pers. Het gebeurt niet vaak, zeg maar vrijwel nooit, dat een NERG lid zoveel internationale waardering en aandacht krijgt voor zijn werk. Schouhamer Immink is een bijzondere persoonlijkheid, die zijn mening niet onder stoelen of banken steekt. Reden genoeg dus om

met hem een interview te houden.

De onderscheiding kwam voor Kees Schouhamer Immink als een complete verrassing: “Eind 1998 kreeg ik bericht dat ik uitgekozen was voor de IEEE *Edison Medal*. Dan lees je die fax nog wel even een keer extra door. Je moet je namelijk voorstellen dat deze onderscheiding de oudste IEEE medaille is, die eerder werd gegeven aan beroemdheden zoals bijvoorbeeld Bell, Tesla, Armstrong, de Forrest en Westinghouse. Ik heb mogen aan-

sluiten bij deze stoet van reuzen, die de elektrotechniek hebben gemaakt tot wat het nu is. Dat is toch wel even wennen. Uiteraard ben ik zeer vereerd met de onderscheiding”. De IEEE eerde hem met de Edison Medal “*For a career of creative contributions to the technologies of digital video, audio, and data recording*”. De Edison Medal wordt niet gegeven voor één specifieke uitvinding of geniaal idee, maar voor bijdragen die zich uitstrekken over een zeer lange periode. Schouhamer Immink weet zijn bijdragen goed te relativeren: “Zonder mij zou de digi-



Kees Schouhamer Immink bezig met een van zijn geliefde bezigheden: het optimaliseren van een experimentele opstelling bij Philips Research
(Foto: Karen Mauve).

tale audio en digitale video er niet veel anders uitgezien hebben. Er zitten natuurlijk wel persoonlijke kanten aan, maar het blijft toch teamwork. Dit in tegenstelling tot bijvoorbeeld schilderkunst en muziekcompositie waar veel meer van een enkel individu afhangt."

De carrière van Kees Schouhamer Immink verliep iets anders dan die van de gemiddelde ingenieur. Na zijn HTS opleiding Elektrotechniek in Rotterdam en na het vervullen van de militaire dienstplicht ging hij eind jaren zestig werken bij het Philips Natlab in Eindhoven. Schouhamer Immink: "Het Natlab was in die tijd een ideale werkomgeving voor mij. De werksfeer was goed en inspirerend en je had allerlei interessante regelingen. Hier heb ik dan ook gebruik van gemaakt door de TU Eindhoven opleiding Elektrotechniek te volgen als werkstudent. Door mijn werk had ik weinig tijd om hoorcolleges te volgen. Hoefde ook niet, ik leerde alles uit boeken en dictaten." Hij noemt zichzelf dan ook een echte autodidact. Dat gecombineerd met hard werken resulteerde er in dat Schouhamer Immink zijn studie aan de TU binnen 3 jaar afrondde, wat zeker in die tijd bijzonder kort te noemen was. Pas veel later, in het midden van de jaren tachtig, schreef hij een proefschrift naar aanleiding van een serie artikelen die al eerder van zijn hand verschenen waren.

Na zijn TU-studie begon hij in de groep Optiek aan optische recording. Schouhamer Immink: "In het begin hield ik me bezig met servosystemen voor de optische

video langspeelplaat. Pas jaren later ben ik gaan werken aan de Compact Disc en vele andere producten. Ik kon me voor de volle honderd procent bezig houden met mijn onderzoek." Midden jaren zeventig begon men bij het Natlab met digitale audio recording. Schouhamer Immink kreeg destijds de opdracht om naar de digitale codering te kijken. "Ik kreeg deze opdracht omdat ik een van de weinige elektronici was in de groep Optiek, die met name bestond uit fysici. Ik was meteen gefascineerd. In het begin was ik een eenling in dit vakgebied. Ik heb me het vak dan ook zelf eigen gemaakt.". Hij kan erg enthousiast vertellen over zijn vakgebied. Voor buitenstaanders echter lijkt het vakgebied nogal theoretisch, mogelijk een beetje saai. "Dat is een groot misverstand", volgens Schouhamer Immink. "Ik deed erg veel experimenteel werk, wat erg belangrijk is, wil je de laatste paar dB's extra krijgen uit je systeem. Later, tijdens mijn promotie op de TU Eindhoven bij Prof. Dr. P. Schalkwijk, kwam ik er achter dat het vakgebied op de Universiteit vooral theoretisch benaderd werd. Het zou kunnen dat dit, zeer ten onrechte, het idee opwekt bij sommige studenten dat het vakgebied saai is."

In 1998 nam Kees Schouhamer Immink plotsklaps ontslag bij het Natlab. Dit verbaasde vriend en vijand. Waarom zou iemand na 31 jaar weggaan bij een bedrijf met allerlei mogelijkheden ook buiten het Natlab? Schouhamer Immink: "Het moderne NatLab staat me tegen, voor mij is te veel een ingenieursbureau geworden. Het zal ook wel met de tijdgeest te maken hebben. De verschui-

ving van lange- naar korte-termijn research is een mondiale trend, die ook in de universitaire wereld goed merkbaar is. Ik vraag me wel eens af of uitvindingen die deze eeuw geleid hebben tot de radio, televisie en transistor, nog steeds ergens gedaan zouden kunnen worden. Ik geloof er niks van. Je moet je namelijk voorstellen dat ik altijd de beste ideeën kreeg in periodes dat ik weinig te doen had, zeg maar tegen de verveling aanhing. In zo'n stille periode had ik



Je moet je namelijk voorstellen dat ik altijd de beste ideeën kreeg in periodes dat ik weinig te doen had, zeg maar tegen de verveling aanhing.

tijd om literatuur te lezen en om 'rare' ideeën uit te werken. Tegenwoordig moet de onderzoeker zich houden aan strakke plannings, waardoor die stille periodes ontbreken. Het echte denkwerk komt hiermee natuurlijk in het gedrang." De hoogtijdagen van de 'ongebonden' research behoren tot het verleden. In de nabije toekomst moet een onderzoeker weer, zoals Gauss, aan het hof gaan werken, of, zoals Darwin, een

steenrijke vrouw trouwen om lange-termijn, ongebonden, onderzoek te kunnen doen. "Ik zie de advertenties al voor me in de Volkskrant: 'Wiskundige zoekt mecenas om *Goldbach's Conjecture* te bewijzen'".

Een financieel aantrekkelijke carrière maken als onderzoeker is in Nederland niet mogelijk, volgens Schouhamer Immink. "In de USA hebben ze dat beter bekeken. Een toponderzoeker bij IBM verdient daar gewoon een topsalaris, vergelijkbaar met een topmanager. In Nederland zegt een baas tegen de talentvolle onderzoeker die extra salaris wil, dat de onderzoeker niet zo moet zeuren omdat hij toch leuk werk heeft. Het is merkwaardig, maar je treft vooral bij ingenieurs het

vooordeel aan dat de hoogte van het plezier in je werk omgekeerd evenredig is met de hoogte van het salaris." Hij adviseert startende ingenieurs die een academische loopbaan niet ambiëren, dan ook om niet voor een groot, maar voor een klein bedrijf te gaan werken, of mogelijk zelf een bedrijf te beginnen. "Ik heb het laatste jaar veel contact met diverse startende high-tech bedrijven. Ze zijn opgezet door slimme jonge ingenieurs met pit, die, als ze alleen al het woord prikklok horen, rode vlekken voor hun ogen krijgen.

Schouhamer Immink heeft geen vastomlijnde toekomstplannen. In ieder geval heeft hij voorlopig genoeg om handen. Behalve dat

hij net een nieuw boek heeft afgerond, werkte hij dit najaar aan de Universiteit van Princeton, en was deze winter weer voor een periode van drie maanden verbonden aan de Universiteit van Singapore. Hij is part-time hoogleraar bij het "Instituut voor Experimentele Wiskunde", een onderzoeksinstituut dat gelieerd is aan de Universiteit van Essen, en hij heeft een eigen bedrijf opgestart. Hopelijk is er nog tijd voor een beetje verveling.

Referenties

- [1] K.A.S. Immink, 'Channel Codes for Digital Recorders', *Tijdschrift van het NERG*, vol. 62, nr. 1, pp. 20-29, 1997.

CV

Kees Schouhamer Immink is geboren in 1946 in Rotterdam. In 1967 behaalde hij zijn HTS diploma in de richting Elektrotechniek aan de Hogeschool van Rotterdam. In 1974 studeerde hij af aan de TU Eindhoven, richting Elektrotechniek, en in 1985 promoveerde hij aan dezelfde universiteit bij Prof Piet Schalkwijk. Van 1967 tot 1998 was hij verbonden aan het Philips Natuurkundig Laboratorium. Hij is nu deeltijdhoogleraar aan de Universiteiten van Essen en Singapore, en directeur van *Turing Machines*.

Hij is Fellow van de IEE, IEEE, AES, SMPTE en gewoon lid van de KNAW. Voor zijn bijdragen aan de digitale audio en video-revolutie werd hij meerdere malen onderscheiden; hij ontving onder meer de AES Gold Medal, de Golden Jubilee Award for Technological Innovation van de IEEE Information Theory Society, de IEEE Masaru Ibuka Consumer Electronics Medal, de Poniatoff Gold Medal Award van de Society of Motion Picture and Television Engineers (SMPTE), en de J.J. Thomson Medal van het Institution of Electrical Engineers (IEE). Juni 1999 ontving hij in Londen de IEEE Edison Medal.



50 jaar Shannon Theorie

J. Pieter M. Schalkwijk *

Department of Electrical Engineering, Eindhoven University of Technology

P.O.Box 513, 5600 MB Eindhoven, The Netherlands

E-mail: piet@ei.ele.tue.nl



Allereerst wil ik mijn dank betuigen aan de redactiecommissie van het NERG, die mij gevraagd heeft wat gedachten over de eerste 50 jaar Shannon theorie voor hun tijdschrift op papier te zetten. Het volgende verhaal vertelt dus van mijn persoonlijke kijk op het vakgebied van de informatie- en communicatietheorie. Ik doe dat aan de hand van drie boeken, die de geïnteresseerde lezer meer detail kunnen verschaffen.

1 Een Proefschrift

Ten eerste is daar het proefschrift [1] van Fritz W. Hagemeyer uit 1979 aan de vrije universiteit van Berlijn, getiteld "Entstehung von Informationskonzepten in der Nachrichtentechnik: eine Fallstudie zur Theoriebildung in der Technik in Industrie- und Kriegsforschung." Hagemeyer heeft vele weken in de Verenigde Staten doorgebracht met het interviewen van grote en minder grote sterren bij Bell Labs en MIT. Hagemeyer's proefschrift ¹ heeft betrekking op de periode van 1924-1948. Het beschrijft werk van Nyquist, Bode, Hartley en Shannon en geeft zodoende een goed idee van het klimaat ten tijde van de conceptie van de Shannon theorie. De grootste bijdrage van Shannon is tweeledig. Hij definieerde een informatiemaat, die het mogelijk maakte totaal verschillende communicatiesystemen op een "eerlijke" manier met elkaar te vergelijken. Het doel van al die communicatiesystemen is namelijk het overbrengen van dezelfde door Shannon gedefinieerde "informatie." Shannon's tweede, misschien nog belangrijker, resultaat was dat zelfs over *onbetrouwbare* transmissiekanalen (met ruis) d.m.v. codering toch betrouwbaar gecommuniceerd kan worden! Mijn opmerking betreffende het *belangrijker* zijn van Shannon's tweede resultaat wil ik toelichten aan de hand van een anecdote. Na het behalen van mijn doctor's graad aan de Stanford universiteit werkte ik van

*J.P.M.Schalkwijk is with the Departement of Electrical Engineering, University of Technology, P.O.BOX 513, 5600MB Eindhoven, The Netherlands. E-mail: piet@ei.ele.tue.nl

¹David Slepian van Bell Labs heeft mijn promovendus Erik Meeuwissen op het bestaan van Hagemeyer's proefschrift gewezen

1965 tot 1968 bij GT&E (General Telephone and Electronics) in Boston. In die tijd ging ik regelmatig naar lezingen over informatie- en communicatietheorie op MIT. Een van dergelijke lezingen werd gegeven door een natuurkundige, die beweerde dat Shannon's grootste verdienste het definiëren van een informatiemaat was en dat die maat al bekend was in de statistische mechnica onder de naam "entropie." In de zaal was Robert Price aanwezig. Price is een bekende researcher uit de beginjaren van ons vakgebied en was in die tijd werkzaam bij MIT Lincoln Labs. Om kort te gaan, Price stond op en informeerde onze spreker dat Shannon meer was dan een "measure maker," en dat zijn tweede resultaat betreffende de betrouwbare transmissie over onbetrouwbare kanalen van veel groter belang was! Men zou nog kunnen tegenwerpen dat zonder Shannon's eerste resultaat betreffende de definitie van een informatiemaat er geen tweede resultaat betreffende betrouwbare transmissie zou zijn. Het is echter wel degelijk mogelijk het fenomeen van betrouwbare transmissie aan te tonen zonder van Shannon's informatiemaat gebruik te maken! Er bestaat een eerder (1947) proefschrift [2] van Kotelnikov uit Rusland, waarvoor de schrijver de Lenin prijs ontving! Kotelnikov beschrijft niet lineaire modulatiesystemen d.m.v. een signaallocus in een N - dimensionale signaalruimte. Door Kotelnikov's resultaten door te trekken naar de asymptotische situatie van oneindige dimensie, $N \rightarrow \infty$, stuit men op het fenomeen van de betrouwbare transmissie. Kotelnikov zelf heeft nagelaten deze laatste stap naar signaalruimten van willekeurig grote dimensie te zetten!

2 De Key Papers

Het tweede boek dat ik de geïnteresseerde lezer zou willen aanraden zijn de z.g. "Key Papers in The Development of Information Theory" [3], samengesteld door David Slepian van Bell Labs naar aanleiding van het 25 jarig bestaan van de informatie- en communicatietheorie in 1973. Het grootste deel van de *Key Papers* gaat over wat we nu de klassieke Shannon theorie noemen. De klassieke Shannon theorie omvat de data compactie, de

data compressie en de kanaalcodering.

2.1 Data Compactie

Gegeven een rij van symbolen, waaraan een waarschijnlijkheidsdistributie bepaalde beperkingen oplegt. Een dergelijke rij symbolen kan eenduidig worden gecomprimeerd in een compacte rij zonder dergelijke beperkingen. De *lengte* van de meest compacte representatie geeft de *informatie* van de oorspronkelijke rij van symbolen. Bestaat de meest compacte representatie uit louter 0-len en 1-nen, dan geeft de lengte van deze representatie de informatie in *bits* (binair eenheden) weer. De data compactie houdt zich bezig met algoritmen die, gegeven de oorspronkelijke rij van symbolen, de kortste representatie realiseren en omgekeerd vanuit deze kortste representatie de oorspronkelijke rij weer herstellen!

Bekende compactie algoritmen zijn b.v. Huffman codering, enumeratieve codering en arithmetische codering. Voornoemde algoritmen maken gebruik van de bekend veronderstelde waarschijnlijkheidsbeschrijving van de oorspronkelijke rij symbolen. Is deze niet bekend dan moeten de waarschijnlijkheden, nodig bij het arithmetisch coderen van het huidige symbool, worden afgeleid uit eerdere symbolen van de oorspronkelijke rij². We spreken nu van *universele* data compactie. Een andere universele data compactie methode is het bekende Lempel-Ziv algoritme.

Als afsluiting van deze subsectie over compactie noemen we nog een toepassing van *enumeratief* coderen. Bij enumeratief coderen wordt een woordenboek van rijtjes symbolen aangelegd, welke rijtjes onderworpen zijn aan bepaalde harde (niet probabilistische) beperkingen. Enumeratief coderen kan dan ook gebruikt worden in situaties waar dergelijke harde beperkingen vereist zijn. Als voorbeeld noemen we de (d,k) -codes op de *compact disk*³, waar rijtjes opeenvolgende 0-len niet korter dan d en niet langer dan k mogen zijn!

2.2 Data Compressie

Data compactie levert de kortste representatie van een rij symbolen op, onder de restrictie dat de bijbehorende expansie *exact* weer de oorspronkelijke rij symbolen moet opleveren! Soms, b.v. bij het coderen van bemonsterde en gedigitaliseerde spraak of video, is exacte reproductie niet noodzakelijk. Nuances in audio of video, die niet door oor respectievelijk oog kunnen worden waargenomen, hoeven niet gereproduceerd te worden. Men zegt dat een bepaalde *distortie* in de reproductie van de

²Mijn medewerkers Frans Willems en Tjalling Tjalkens werken aan dit onderwerp

³Zie proefschrift van Kees Schouhamer- Immink

rij symbolen is toegestaan. Het toelaten van distortie opent de mogelijkheid van een kortere representatie van de oorspronkelijke rij symbolen. De *rate-distortie theorie*, ofwel de *data compressie*, houdt zich bezig met algoritmen voor het comprimeren en weer expanderen van een rij symbolen, waarbij de gereconstrueerde rij de oorspronkelijke rij binnen de toegelaten distortie benadert.

Data compressie is b.v. van groot belang bij het digitaal opslaan en verzenden van spraak en van beelden. MPEG-2 is een bekende en veel gebruikte compressie compressiestandaard. Ofschoon rate-distortie theorie voor enkele eenvoudige situaties de ultieme compressie voor gegeven distortie heeft opgeleverd, moet eerlijkheidshalve gezegd worden dat de algoritmen die gebruikt worden bij MPEG-2 codering eerder het werk waren van experts in de signaalverwerking dan van informatietheoretici.

De *universele* data compressie is op dit moment nog een onderwerp van actieve research!

2.3 Kanaalcodering

Shannon (1948) heeft met zijn *random coding argument* het bestaan van codes aangetoond, die betrouwbare transmissie over onbetrouwbare (ruisachtige) kanalen kunnen realiseren! Dit *existentie* bewijs gaf aanleiding tot de geboorte van de coderingstheorie. Met deze geboorte vond een scheiding van de geesten plaats. Sommigen gingen verder op het door Shannon met zijn *random coding* aangegeven pad van de stochastische productie van de codewoorden. Dit leidde tot het ontstaan van de *convolutiecodes*. Anderen meer discreet wiskundig ingestelde onderzoekers zochten naar codewoorden op basis van een algebraïsche structuur. Dit onderzoek leidde tot het nu volwassen gebied van de algebraïsche *blokkodes*.

2.3.1 Convolutiecodes

Shannon's *random* codering suggereert een pseudo random constructie van codewoorden. De rij van binaire digits van het bericht wordt als *seed* gevoed aan een pseudo random nummer generator, die bij een $\text{rate} = \frac{1}{2}$ code voor iedere digit van het bericht twee code-digits produceert. Wonder boven wonder blijkt de allereenvoudigste pseudo random nummer generator prachtige codes te produceren. De binaire digits van het bericht worden een voor een in een schuifregister geschoven⁴. Twee modulo-2 optellers zijn elk met bepaalde schuifregister cellen verbonden en produceren elk een codestroom digit voor iedere ingeschoven digit van het bericht. De aldus

⁴Zie proefschrift van Han Vinck

gevormde convolutiecode kan op twee manieren gedecodeerd worden.

Bij Fano-decodering wordt de code opgevat als een boom, waarbij iedere "0" van het bericht met een vertakking naar beneden en elke "1" met een vertakking naar boven correspondeert. Langs elke tak van de boom staan de twee bijbehorende code-woord digits. Het *Fano* decodeeralgoritme zoekt nu stap voor stap door de boom naar het code-stroompad, dat de meeste overeenkomst met de ontvangen data rij vertoont. Fano decodering heeft geringe complexiteit, maar het nadeel van een variabele decodeervertraging!

Bij Viterbi-decodering wordt de eerder beschreven pseudo random nummer generator als een *finite state machine* beschouwd. Door de toestanden op een bepaald tijdstip verticaal boven elkaar af te beelden en de toestanden een tijdstip (d.w.z. een digit van het bericht) later rechts daarvan, ontstaat een z.g. *trellis diagram*. Een *trellis* verschilt daarin van een *tree*, dat eenmaal van elkaar afwijkende codestroom paden weer samen kunnen komen. Het meest met de ontvangen datastroom overeenkomende van de twee samenkomende codesroompaden wordt de *survivor* genoemd. Door bij het decoderen steeds de survivor te volgen vindt het Viterbi algoritme nu het codestroompad met de meeste overeenkomst door het trellis diagram. Het Viterbi algoritme heeft een grotere complexiteit dan het Fano algoritme, maar de decodeervertraging ligt vast!

2.3.2 Blokcodes

Binaire blokcodes bestaan uit een kleine fractie van alle mogelijke verschillende rijtjes van 0-len en 1-nen van een bepaalde vaste blok lengte, N . Kiezen we deze rijtjes weer *at random*, dan heeft de resulterende binaire blokcode met hoge waarschijnlijkheid een goede afstandstructuur. D.w.z. dat twee willekeurige codewoorden uit de code in veel coördinaten (corresponderende posities) van elkaar zullen verschillen. Bij het decoderen zoeken we het codewoord dat in de meeste coördinaten met het ontvangen datablok overeenstemt. Voor een ongestructureerde (random) blokcode met een groot aantal codewoorden is het stuk voor stuk vergelijken van de codewoorden met het ontvangen blok in de praktijk onuitvoerbaar! Daarom leggen we de codewoorden een algebraïsche structuur op, om zodoende het goede (verzonden) codewoord aan de hand van het ontvangen datablok te kunnen *berekenen*. Helaas levert het opleggen van een algebraïsche structuur aan de codewoorden i.h.a. een code met een slechte afstandstructuur op. Bekende blokcodes zijn de Hamming codes, de Bose-Chaudhuri-Hocquenghem (BCH) codes en de Reed-Solomon (RS) codes.

Aan de TUE werd het college over algebraïsche blokcodes verzorgd door de professoren Jack van Lint en Henk van Tilborg van de vakgroep discrete wiskunde van de faculteit Wiskunde en Informatica.

2.3.3 Concatenated Codes

In de begin jaren woedde er een controverse tussen de aanhangers van de convolutiecodes aan de ene kant en de algebraïsche blokcoders aan de andere kant. De tegenstelling liep hoog op, getuige het volgende citaat uit een artikel van Jim Massey: "*algebraic codes make for good papers, whereas convolutional codes make good communication systems!*" Later zijn de geschillen bijgelegd, omdat de twee klassen van codes beide hun eigen voordelen hebben. Zo zijn convolutiecodes met Viterbi decodering geschikt voor kanalen met een continue (soft decision) output. De output stroom van de Viterbi decoder vertoont *burst* fouten, die door een overliggende (concatenated coding) algebraïsche RS code in tweede instantie opgeruimd kunnen worden. M.a.w. veel moderne communicatiesystemen gebruiken *concatenated coding*, d.w.z. een combinatie van een convolutie- en een algebraïsche blokcode!

2.3.4 Recente Ontwikkelingen

Aanvankelijk hield de kanaalcodering zich bezig met eenvoudige kanaal modellen, zoals het binair symmetrische kanaal (BSC) en het kanaal met additieve Gaussische witte ruis (AGWN). Tegenwoordig analyseert men complexere kanalen, i.h.b. het magnetische recording kanaal (transmissie in de tijd, van *nu* naar *dan*) en het mobile radio kanaal (transmissie in de ruimte, van *hier* naar *daar*, als we de propagatie-delay tenminste verwaarlozen) met fading en multipath.

Als laatste ontwikkeling in de kanaalcodering noemen wij *turbo decoding*. Terwijl een klassieke decoder de waarde van een te decoderen digit geeft, levert de bij turbo-decoding gebruikte decoder bovendien de a posteriori waarschijnlijkheid van het output digit. In de volgende iteratie gebruikt deze decoder de oude output a posteriori waarschijnlijkheid als een nieuwe input a priori waarschijnlijkheid, enz. Uit simulaties blijkt dat met *turbo decoding* de Shannon capaciteitsgrens heel dicht benaderd kan worden, maar een rigoreuze theorie van turbo-decoding ontbreekt nog!

2.4 Netwerk Informatietheorie

Zoals eerder in Sectie 2 werd opgemerkt, gaat het grootste deel van de *Key Papers* over de klassieke Shannon theorie, die we nu in Sectie 2.3.4 hebben afgerond. Het laatste gedeelte van de *Key Papers* behandelt de *netwerk informatietheorie*, zie

ook hoofdstuk 14 van Tom Cover's [4] boek. Terwijl we in de klassieke Shannon theorie een enkele informatiestroom van zender naar ontvanger bestuderen, beschouwt de netwerk informatietheorie zoals de naam al aangeeft, een netwerk (graph) van simultane stromen van informatie. In analogie van de klassieke, *single-user*, theorie onderscheiden we bij de netwerk, *multi-user*, theorie weer tussen multi-user broncodering en multi-user kanaalcodering. Echter, het beroemde *separation theorem* van de klassieke single-user Shannon theorie, dat de optimaliteit van gescheiden bron- en kanaalcodering aantoont, geldt niet in z'n algemeenheid voor de netwerk informatietheorie. Bij de multi-user kanaalcodering kan een onderscheid gemaakt worden tussen statische- en dynamische coderingssystemen. Om dit onderscheid te verduidelijken, behandelen we na de multi-user broncodering in Sectie 2.4.1 eerst *feedback coding* in Sectie 2.4.2.

2.4.1 Slepian-Wolf Broncodering

Uit Sectie 2.1 over data compactie weten we, dat voor het coderen van een stroom symbolen van bron X , per bronsymbool $H(X)$ code digits nodig zijn. De informatie $H(X)$ per bronsymbool wordt ook wel de *entropie* van de bron X genoemd. Beschouw nu de *multi-user* situatie van twee gecorreleerde geheugenloze bronnen X en Y . De meest compacte representatie van een rij symbolen van bron Y vraagt $H(Y)$ digits per Y -symbool. Slepian en Wolf laten echter zien dat wanneer de corresponderende rij X -symbolen aan de decoder bekend is, er volstaan kan worden met $H(Y|X)$ digits per Y -symbool. Als de bronnen X en Y statistisch afhankelijk zijn, d.w.z. als $P_{XY} \neq P_X P_Y$, dan geldt dat de conditionele entropie $H(Y|X) < H(Y)$. M.a.w. *zonder dat de X -symbolen aan de encoder bekend zijn* kan toch een grotere compactie bij het encoderen van de Y -symbolen bereikt worden, zolang de X -symbolen maar bekend zijn bij het decoderen! Dit is een van de vele verrassende en onverwachte *multi-user* resultaten, die een nieuw hoofdstuk gescheiden van de *single-user* (klassieke) theorie rechtvaardigen. Analoge resultaten, t.w. Wyner-Ziv coding, bestaan in de multi-user data *compressie*. Discussie van multi-user compressie zou echter te ver voeren voor dit overzichtsartikel. Vooral ook omdat de theorie van de data compressie, zoals al opgemerkt in Sectie 2.2, tot nu toe weinig toepassing heeft gezien.

2.4.2 Feedback Coding

Een merkwaardig resultaat van Shannon is het volgende. Als van een *geheugenloos* kanaal de output na iedere transmissie aan de zender bekend wordt gemaakt (via een foutloos feedback kanaal), dan

heeft dit *geen* vergroting van de transmissiecapaciteit in de heenrichting tot gevolg! De *feedback* heeft daartegen een *fundamentele* invloed⁵ op het coderingsproces. Terwijl in klassieke *one-way* communicatie het codewoord onverbreekelijk (statisch) aan het bericht gekoppeld is, is deze koppeling in systeem met *feedback* dynamisch. D.w.z. dat voor een gegeven bericht het codewoord mede (dynamisch) bepaald wordt door de responsie van het kanaal op de voortschrijdende transmissie! Het dynamische karakter van de codering heeft twee belangrijke gevolgen. Ten eerste kan de optimale code m.b.v. basis principes uit de informatietheorie eenduidig worden afgeleid! Om deze rede vond de wiskundige Jacob Wolfowitz de feedback codering veel interessanter dan de *one-way* codering, die toch min of meer ad hoc is. Ten tweede is de complexiteit van de resulterende dynamische feedback decoding wezelijk minder dan die van een vergelijkbaar statisch *one-way* decoderingssysteem!

Shannon heeft dus laten zien dat voor *geheugenloze* kanalen *feedback* de capaciteit niet verhoogt! Voor transmissiekanalen met *geheugen*, d.w.z. afhankelijkheid van het kanaalgedrag tussen opeenvolgende transmissies, is soms wel een capaciteitsverhoging te realiseren [4].

Sinds het 25-jarig bestaan van de Shannon theorie, d.w.z. sinds het IEEE informatietheorie (IT-)symposium van 1973 in Ashkelon (Israel), is er steeds een *Shannon-lecture* op dit symposium, dat iedere 18 maanden wordt gehouden. Een bekende onderzoeker valt de eer te beurt dit Shannon-lecture te mogen geven en het eerste Shannon-lecture in 1973 in Ashkelon werd door Shannon zelf gegeven. Zijn onderwerp was *feedback*, volgens Shannon een van de belangrijkste en minst begrepen aspecten van de netwerk informatietheorie!

2.4.3 Statische Multi-User Kanalen

Eerst behandelen we de multi-user kanalen zonder inherente *feedback*, d.w.z. de statische multi-user kanalen. De bekendste multi-user kanalen in deze categorie zijn het *multiple access channel* (MAC) en het *broadcast channel* (BC).

Bij het MAC hebben we twee of meerdere zenders en een enkele ontvanger. De duale situatie van het BC kent een enkele zender en twee of meerdere ontvangers. De situatie bij het MAC kan vergeleken worden met die van twee kinderen die hun vader over school willen vertellen. Vroeger zei de vader dan: "eerst Jantje en dan Pietje!" In het menselijke verkeer heet die gang van zaken *beleefd* en in de Shannon theorie noemen we dat *time-sharing*! Wat blijkt nu echter. Als Jantje en Pietje door el-

⁵Zie de proefschriften van Jim Tiernan (UCSD), Donald Becker (UCSD), Thijs Veugen (TUE) en Mortaza Shoaie Bargh (TUE)

kaar heen praten kan er in dezelfde tijd meer informatie overgedragen worden! De vader krijgt wel barstende hoofdpijn, hetgeen wijst op een gecompliceerde *decodeer* procedure! De multi-user situatie van het MAC is opgelost en zowel de *bereikbaarheid* van het informatie theoretische *capaciteitsgebied* als het *converse* vragen slechts om een veralgemening van het bewijs voor *one-way* kanalen uit de klassieke Shannon theorie.

De situatie bij het duale geval van het BC kan worden vergeleken met een onderwijzer (de enkele zender), die voor een combinatie 5-de en 6-de klas staat. Het traditionele onderwijs, waarbij de onderwijzer eerst twee uur voor de 5-de klas en vervolgens twee uur voor de 6-de klas staat, correspondeert weer met *time-sharing*. Bij de moderne situatie van een combinatieklas kan de onderwijzer in dezelfde tijd, t.w. vier uur, totaal meer informatie overdragen. Wel krijgt hij knetterende koppijn, hetgeen nu duidt op een gecompliceerde *codeer* procedure! In tegenstelling tot het MAC is het *capaciteitsgebied* van het BC tot op heden niet bekend!

Het *degraded broadcast channel* (DBC) is wel opgelost! De informatie voor de *slechte* ontvanger zit in de locatie van de *cloud centers* van codewoorden. De informatie voor de *goede* ontvanger is als ruis op de cloud centers *gesuperponeerd*. De goede ontvanger decodeert eerst de informatie voor de slechte ontvanger, d.w.z. de identiteit van de *cloud*, die hem in wezen niet interesseert. Daarna decodeert de goede ontvanger zijn eigen informatie, door het gezonden codewoord in de relevante cloud te zoeken. Deze door Cover [4] ontwikkelde manier van coderen in twee stappen wordt *superposition coding* genoemd. Superpositie coding is weer een typisch multi-user resultaat en geen veralgemening van een eerdere codering uit de klassieke Shannon theorie.

2.4.4 Dynamische Multi-User Kanalen

Shannon's artikel over het *two-way channel* (TWC), zie de *Key Papers* [3], wordt algemeen als aanzet van de netwerk informatietheorie gezien. Ofschon dit artikel gewag maakt van het MAC, is de wiskundige theorie van de *dialog*, d.w.z. het TWC, het eigenlijke onderwerp. De situatie van het TWC is te vergelijken met een dialoog tussen twee gelijkwaardige personen, die in de netwerk informatietheorie *terminals* worden genoemd. Terminal 1 produceert een kanaal input X_1 voor het TWC en terminal 2 produceert X_2 . De output Y_1 bij terminal 1 is i.h.a. statistisch afhankelijk van beide inputs X_1 en X_2 , en dat geldt ook voor output Y_2 bij terminal 2. Het generieke TWC is het z.g. *binary multiplying channel* (BMC), waarbij $Y_1 = Y_2 = X_1 X_2$. D.w.z. dat $Y_1 = Y_2 = 1$ als zowel X_1 als ook X_2

gelijk aan 1 is. Als X_1 en/of X_2 gelijk aan 0 is, dan zijn Y_1 en Y_2 beide gelijk aan 0. Merk op dat het BMC *deterministisch* is, d.w.z. dat de outputs Y_1 en Y_2 beide normale (niet stochastische) functies van X_1 en X_2 zijn. Een dergelijk *deterministisch* kanaal is in de klassieke *one-way* theorie niet zo interessant, maar in de netwerk informatietheorie wel omdat de informatiestroom van links naar rechts i.h.a. interfereert met de informatiestroom van rechts naar links. M.a.w. een grotere stroom van links naar rechts zal i.h.a. resulteren in een kleinere van rechts naar links! Merk op dat als terminal 2 z'n input X_2 voortdurend op $X_2 = 1$ houdt er een perfect binair kanaal van X_1 naar Y_2 ontstaat, d.w.z. $Y_2 = X_1$. In dit geval kunnen we $R_1 = 1$ bit per transmissie van links naar rechts sturen. De rate R_2 van rechts naar links is nu natuurlijk $R_2 = 0$ omdat we $X_2 = 1$ vasthouden. Als we alterneren, *time-sharing*, tussen perioden met $X_2 = 1$ en even lange perioden met $X_1 = 1$ dan halen we het time-sharing rate-paar $(R_1, R_2) = (0.5, 0.5)$. Maar net als bij het MAC en bij het BC kunnen we beter dan time-sharing. Met statische codering, d.w.z. met een vast codewoord voor ieder bericht, kan het rate punt $(R_1, R_2) = (0.61695, 0.61695)$, op Shannon's *innerbond*, bereikt worden. Met *dynamische* codering, waarbij de codestroom zowel van het bericht als van de ontvangen data afhangt kan het rate punt $(R_1, R_2) = (0.63056, 0.63056)$ buiten Shannon's *inner bound area* worden bereikt⁶. Het *capaciteitsgebied*, d.w.z. het gebied van bereikbare (R_1, R_2) -paren, van het BMC is vooralsnog onbekend⁷!

3 De Collected Papers

De in 1993 gepubliceerde *Collected Papers* [5] geven een compleet overzicht van Shannon's werk. Hier vinden we ook resultaten betreffende de toepassing van Boole algebra voor het ontwerp van digitale schakelingen. Een tweede onderwerp dat nog niet genoemd is en dat nauw vervlochten is met Shannon's theorie van informatietransmissie is de *cryptologie*.

Een belangrijke grootheid in de cryptologie is de equivocatie $H_N(K|C)$, d.w.z. de conditionele entropie of onzekerheid van de *key*, K , gegeven N letters van de *cipher* tekst, C . Shannon's werk betreft de *absolute security* van crypto-systemen. De moderne *public-key* systemen werken met de *computational security*, d.w.z. dat public-key systemen wel te breken zijn, maar dat de daarvoor benodigde *rekeningtijd* onpractisch groot is!

In het bovenstaande heb ik de meeste aandacht gegeven aan die onderdelen van de Shannon theo-

⁶Met Schalkwijk's methode van vierkantsresolutie

⁷Zie de proefschriften van Andries Hekstra, Ineke van Overveld, Phons Bloemen en Erik Meeuwissen

rie, die mij persoonlijk het meest aanspreken. Min of meer de gebieden waar ik gedurende mijn loopbaan zelf aan gewerkt heb. Misschien komt het bovenstaande overzicht van de Shannon theorie op u over als een afgerond geheel? Met de regelmaat van de klok wordt de Shannon theorie zo om de tien jaar *dood* verklaard. Dan komt er onverwacht een ontdekking, zoals b.v. de *public key cryptography* door Diffie en Hellman in 1976, die de boel weer nieuw leven inblaast. Misschien wilt u van mij weten wat ik voor de toekomst verwacht? In de eerste plaats zullen bekende gebieden verder worden uitgebouwd. Meer werk in audio- en video compressie, meer werk in *speech recognition*. Het *additive white Gaussian noise* (AWGN) kanaal was een goed model voor *deep space* communicatie. Maar de modellen en de codering van b.v. het *magnetic recording*- en het *mobile radio* kanaal, zie ook Sectie 2.3.4, zijn veel ingewikkelder en vergen nog veel onderzoek. Bovenstaande valt echter, oneerbeidig gezegd, onder de categorie *more of the same*, maar moeilijker! Verder zullen veel resultaten uit de netwerk informatietheorie in de naaste toekomst praktische toepassingen zien. Natuurlijk bent u ook benieuwd of er nog geheel nieuwe ontwikkelingen zijn te verwachten. Als we terugkomen op onze eerdere vergelijking van de Shannon theorie met onze menselijke communicatie van elke dag, dan zien we het volgende. Eerst is daar de monoloog, waar A iets aan B vertelt. Dit is het onderwerp van de klassieke Shannon theorie. Dan in de netwerk informatietheorie krijgen we de volgende situaties. A en B vertellen iets aan C, d.w.z. het MAC. A vertelt iets aan B en C, d.w.z. het BC, en tenslotte de situatie waar A en B een dialoog voeren, het TWC. De zaak lijkt dus min of meer rond. Maar, en laten we hiermee sluiten, als informatietheoreticus weet ik als geen ander dat echte innovatie niet te voorspellen is!

CV

J.Pieter M. Schalkwijk, Geboren op 1 nov, 1936.

1959 Diploma electrotechnisch ingenieur (cum laude) aan de TUD

1965 PhD electrical engineering Stanford University

1959-1961 HSA Hengelo (militaire dienst), afdeling keuring digitale radarvuurleidingsapparatuur.

1961-1963 RVO/TNO Den Haag, onderzoek aan digitale radioverbindingen voor de marine.

1963 NLR Amsterdam, digitale meetapparatuur voor windtunnels.

1963-1965 Research assistent at Stanford University.

1965-1968 General Telephone - & Electronics Boston, digital transmission.

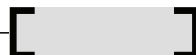
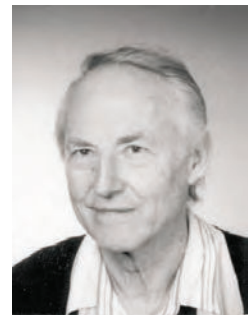
1968-1972 Assistent professor Information and Communication Theory at the Univ. of Cal. at San Diego (UCSD).

1972-1998 Professor in Informatie- en Communicatietheorie aan de TUE.

Fellow IEEE, Erelid NERG, past member Board of Governors IEEE-IT Society, oud bestuurslid NERG en ontving in 1967 (samen met T.Kailath) de IEEE-IT Best Paper Award.

Referenties

- [1] F.-W. Hagemeyer. *Die Entstehung von Informationskonzepten in der Nachrichtentechnik: Eine Fallstudie zur Theoriebildung in der Technik in Industrie- und Kriegsforchung*. PhD thesis, Freien Universität Berlin, Fachbereichs Philosophie und Sozialwissenschaften, November 1979.
- [2] V.A. Kotelnikov. *The Theory of Optimum Noise Immunity*. McGraw-Hill, New York, 1959. (Doctoral dissertation presented in January 1947 before the accademic council of the Molotov Energy Institute in Moscow.)
- [3] *Key Papers in The Development of Information Theory*. Edited by David Slepian. IEEE Press, New York, 1973.
- [4] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, 1991.
- [5] *Claude Elwood Shannon Collected Papers*. Edited by Neal Sloane and Aaron Wyner. IEEE Press, New York. 1993.



Adaptive Filtering Approach to Multi-channel Audio Reproduction Systems

John Garas* and Piet Sommen
Eindhoven University of Technology
E-Hoog 3.34, P.O. Box 513
5600 MB Eindhoven, the Netherlands
Email: P.C.W.Sommen@tue.nl



1 Foreword

One of the research areas of the Signal Processing group at the Eindhoven University of Technology is in *adaptive array signal processing*. It is concerned with the application of adaptive signal processing techniques to the outputs of a spatially distributed array of sensors. The power of these techniques lies in the possibility of achieving improvements in directivity and measured Signal-to-Noise Ratio (SNR) as compared to that attainable with a single input sensor. The spatial dimensions introduced by processing a distributed array of sensors allow directional discrimination of signals against interference. Sensor arrays are used in many applications including radar, sonar, seismology and tomography [5]. Most of these applications use the fact that the used signals are band limited and that the sources are so far away that at the sensors they can be treated as a flat wavefront. In many (certainly 'in-the-house') audio applications these narrowband and far field assumptions certainly do not hold. Mainly due to the vast increase of available computational power, the commercial applications of array processing advanced audio systems have become feasible in recent years.

Recently two PhD thesis [4, 11] have been successfully completed in the Signal Processing group in which adaptive array processing techniques were applied to audio applications. The work of [11] was focused around a part of the transparent audio communication project. The term 'transparent communication' refers to audio communication which is free from recording and transmission artifacts, such as reverberation, noise and acoustic echoes. Also, it is desirable to separate speech signals of multiple speakers who are speaking simultaneously.

The current paper, based on the results of the PhD thesis work outlined in [4], discusses theoretical and implementation issues concerning adaptive

multichannel audio reproduction systems.

2 Introduction

Filters that are used in loudspeaker displays must implement both binaural synthesis and cross-talk cancellation functions. The cross-talk cancellation subsystem is the inverse of an often non-minimum phase and ill-conditioned matrix of electro-acoustic transfer functions. An exact solution to the cross-talk canceller is difficult to calculate, and in some cases, it may not be possible at certain frequencies. Alternative to exact direct inversion, a statistical least mean square solution may be obtained for the matrix inverse. The advantage of such an approach is that it can be made adaptive, so that the filters can be designed *in-situ*. Such an adaptive solution also offers the possibility of tracking and correcting the filters when changes in the electro-acoustic system occur.

An adaptive loudspeaker display that creates a single sound image at the ears of one listener requires two loudspeakers and two microphones placed inside the listener's ears. More microphones are needed as the number of listeners increases, and more filters are needed as the number of virtual sound images increases. Instead of treating any of the above special cases, a generalised model is introduced in Section 3. In addition to generalising the number of sound sources, loudspeakers and microphones used in the reproduction system, it will be shown in Section 4 that the model is capable of describing a wide class of applications including synthesis of virtual sound sources, cross-talk cancellation, and active noise control.

The optimum least mean square solution for the control filters in the generalised model is derived in Section 5. Although important from the numerical analysis point of view, real-time implementation of the system would require approaching the optimum solution using an iterative algorithm.

*John Garas is currently with TNO-TPD Delft



The Multiple Error Filtered- X Least Mean Square (MEFX) algorithm, is such an algorithm and is introduced in Section 6.

Several implementation details are discussed by considering the implementation of the MEFX algorithm in active noise control, virtual sound source synthesis, and cross-talk cancellation applications in Sections 7, 8, and 9, respectively.

In reverberant acoustic environments, the impulse response between two points may last for several hundreds of milliseconds. At the standard audio compact disc sampling frequency of 44.1 kHz, thousands of FIR filter coefficients are needed to properly model and store such an impulse response. Processing many of these transfer functions requires a huge amount of computational power. Reducing the system complexity is, therefore, essential for real-time implementation. Efficient implementations using the Adjoint LMS and its Block Frequency Domain version are discussed in Section 10.

3 A Generalised Model

A generalised block diagram of a multichannel audio reproduction system is shown in Fig. 1. A set of L reproduction loudspeakers $\{S_1, S_2, \dots, S_L\}$ are used to play K pre-recorded audio signals defined at the sample index n by

$$\underline{\mathbf{x}}(n) = [x_1(n) \quad x_2(n) \quad \dots \quad x_K(n)]^T$$

The sound field generated due to these loudspeakers is required to be controlled at the proximity of a set of M microphones (receivers) $\{R_1, R_2, \dots, R_M\}$ to a set of desired values defined as

$$\underline{\mathbf{d}}(n) = [d_1(n) \quad d_2(n) \quad \dots \quad d_M(n)]^T$$

Sound waves emitted from the loudspeakers are filtered through the $[M \times L]$ matrix of electro-acoustic transfer functions $\mathbf{C}(\omega)$ before reaching the microphones' positions. The matrix $\mathbf{C}(\omega)$ contains the Fourier transforms of the impulse responses $\{c_{ml} : m = 1, 2, \dots, M, l = 1, 2, \dots, L\}$ evaluated at a frequency ω and defined as

$$\mathbf{C}(\omega) = \begin{bmatrix} C_{11}(\omega) & \dots & C_{1L}(\omega) \\ C_{21}(\omega) & \dots & C_{2L}(\omega) \\ \vdots & \ddots & \vdots \\ C_{M1}(\omega) & \dots & C_{ML}(\omega) \end{bmatrix}$$

where $C_{ml}(\omega)$ is the electro-acoustic transfer function between the m^{th} microphone R_m and the l^{th} reproduction loudspeaker S_l calculated at ω . The above mentioned control task is achieved by introducing an $[L \times K]$ matrix of (adaptive) digital filters $\mathbf{W}(\omega)$ in the reproduction chain defined at a

frequency ω as

$$\mathbf{W}(\omega) = \begin{bmatrix} W_{11}(\omega) & \dots & W_{1K}(\omega) \\ W_{21}(\omega) & \dots & W_{2K}(\omega) \\ \vdots & \ddots & \vdots \\ W_{L1}(\omega) & \dots & W_{LK}(\omega) \end{bmatrix}$$

where $W_{lk}(\omega)$ is the filter driving the l^{th} reproduction loudspeaker S_l and has $x_k(n)$ as its input. The filters $\mathbf{W}(\omega)$ are designed to produce the control signals $\underline{\mathbf{y}}(n)$ to drive the L reproduction loudspeakers such that the resulting sound field at the microphones $\hat{\underline{\mathbf{d}}}(n)$ is as close as possible to the desired sound field $\underline{\mathbf{d}}(n)$. The vectors $\underline{\mathbf{y}}(n)$ and $\hat{\underline{\mathbf{d}}}(n)$ are defined as

$$\begin{aligned} \underline{\mathbf{y}}(n) &= [y_1(n) \quad y_2(n) \quad \dots \quad y_L(n)]^T \\ \hat{\underline{\mathbf{d}}}(n) &= [\hat{d}_1(n) \quad \hat{d}_2(n) \quad \dots \quad \hat{d}_M(n)]^T \end{aligned} \quad (1)$$

The set of desired sound fields given by the vector $\underline{\mathbf{d}}(n)$ is usually well correlated with the set of K input signals $\underline{\mathbf{x}}(n)$. Therefore, $\underline{\mathbf{d}}(n)$ may be considered as the result of filtering $\underline{\mathbf{x}}(n)$ through an $[M \times K]$ matrix of transfer functions $\mathbf{H}(\omega)$ (not shown in Fig. 1). This filtering operation may be expressed in the frequency domain as

$$\underline{\mathbf{D}}(\omega) = \mathbf{H}(\omega) \underline{\mathbf{X}}(\omega)$$

where $\underline{\mathbf{D}}(\omega)$ and $\underline{\mathbf{X}}(\omega)$ are the DFTs of the time history of $\underline{\mathbf{d}}(n)$ and $\underline{\mathbf{x}}(n)$, respectively. The matrix $\mathbf{H}(\omega)$ contains the DFT of the impulse responses $\{h_{mk} : m = 1, 2, \dots, M, k = 1, 2, \dots, K\}$ between the microphones and the audio signals evaluated at a frequency ω and is defined as

$$\mathbf{H}(\omega) = \begin{bmatrix} H_{11}(\omega) & \dots & H_{1K}(\omega) \\ H_{21}(\omega) & \dots & H_{2K}(\omega) \\ \vdots & \ddots & \vdots \\ H_{M1}(\omega) & \dots & H_{MK}(\omega) \end{bmatrix}$$

The key point in the above mentioned model is to design the matrix $\mathbf{W}(\omega)$ of $[L \times K]$ filters to achieve the control task. The design of those filters is discussed later. Examples illustrating the use of the generalised model in different audio reproduction applications are discussed in the next section.

4 Example Applications

In the generalised model introduced above, different audio reproduction applications may be described by defining different desired sets of sound fields $\underline{\mathbf{d}}(n)$ at the microphones. Consequently, it is the matrix $\mathbf{H}(\omega)$ that defines the nature of the application. Examples of the applications that may be described by the model are *cross-talk cancellation*,

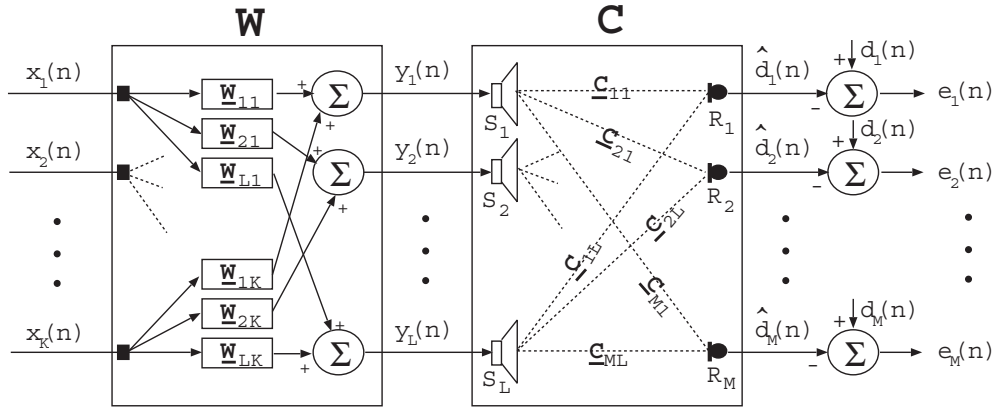


Figure 1: Generalised block diagram of multichannel audio reproduction system.

virtual source synthesis, concert hall simulation, correction of the responses of the reproduction loudspeakers, and active noise control.

In cross-talk cancellation (see Section 9), it is desired to exactly reproduce the input signals $\underline{x}(n)$ at the microphones. In this case, $\underline{d}(n) = \underline{x}(n)$, and $\mathbf{H}(\omega) = \mathbf{I}$, the $[K \times K]$ identity matrix. Applying this for all frequencies, the transfer function between the m^{th} microphone and the k^{th} input signal has all its elements equal to unity, $\{\mathbf{H}_{mk}(\omega) = \underline{1} : m = k = 1, 2, \dots, K\}$, which corresponds to a unit impulse response in the time domain $\underline{h}_{mk}(n) = \delta(n)$ ¹. Alternatively, K virtual sound images may be generated at the ears of $M/2$ (M even) listeners if $\underline{d}(n)$ is the result of filtering the input signals $\underline{x}(n)$ through the appropriate matrix of the so-called Head Related Transfer Functions (HRTFs).² The generalised model may also be used to simulate a concert hall by setting $\mathbf{H}(\omega)$ to be a matrix of transfer functions measured in a concert hall. Another application that may be described by the model is correcting the responses of the reproduction loudspeakers to obtain better sound quality. This may be achieved by letting $\mathbf{W}(\omega)$ be an inverse model of the loudspeaker's transfer functions. The most popular application of the generalised model is in active noise cancellation [3]. In such an application, a set of K (undesired rather than desired) sound disturbances $\underline{d}(n)$ are to be silenced at the microphones. This is achieved by filtering the input signals through $\mathbf{W}(\omega)$ to generate sound waves, $\hat{\underline{d}}(n)$, that are equal in amplitude but

¹Note that in practise $\underline{h}_{mk}(n) = \delta(n)$ results in a non-causal set of filters $\mathbf{W}(\omega)$. Causal filters produce a delayed version of the inputs $\hat{\underline{d}}(n) = [x_1(n - \Delta_1) \ x_2(n - \Delta_2) \ \dots \ x_K(n - \Delta_K)]$ at the microphones. In this case, the matrix \mathbf{H} contains delayed unit impulse responses $\{\mathbf{h}_{mk}(n) = \delta(n - \Delta_k) : m = k = 1, 2, \dots, K\}$.

²HRTFs are referred to as the pair of acoustic transfer functions from the source to each of the listener's eardrums embedding all physical parameters associated with the localisation cues that occur in natural listening situations.

opposite in phase to the disturbances at the microphone's positions, such that the net microphone's outputs $\underline{e}(n)$ are minimised, where $\underline{e}(n)$ is given by

$$\underline{e}(n) = [e_1(n) \ e_2(n) \ \dots \ e_M(n)]^T \cdot \gamma \quad (2)$$

The only difference between the above mentioned applications when using the generalised model is the desired response $\underline{d}(n)$ at the microphones. Therefore, a generalised solution for $\mathbf{W}(\omega)$ that is valid for the whole class of applications may be obtained by solving the system equations for an arbitrary desired response. The solution for a specific application is then obtained by substituting its specific desired response into the general solution.

5 The Optimum Least Mean Square Solution

The optimum Least Mean Square (LMS) solution for the system shown in Fig. 1 is obtained by minimising a performance index function $\xi(n)$, usually taken as the sum of the mean squared error signals

$$\begin{aligned} \xi(n) &= \sum_{m=1}^M E\{e_m^2(n)\} \\ &= E\{\underline{e}^T(n) \underline{e}(n)\} \end{aligned} \quad (3)$$

where $E\{\cdot\}$ denotes the mathematical expectation. In the following discussion, all filters $\{\underline{w}_{lk} : l = 1, 2, \dots, L, k = 1, 2, \dots, K\}$ are assumed to be Finite Impulse Response (FIR) digital filters, each of length N_w and defined as

$$\underline{w}_{lk} = [w_{lk,0} \ \dots \ w_{lk,N_w-1}]^T$$

The signal driving the l^{th} reproduction loudspeaker $y_l(n)$ is the sum of the outputs of the filters $\{\underline{w}_{lk} : k = 1, 2, \dots, K\}$, which may be expressed as

$$y_l(n) = \underline{x}_1^T(n) \underline{w}_{l1} + \dots + \underline{x}_K^T(n) \underline{w}_{lK} \cdot \gamma \quad (4)$$

where the time history of the k^{th} input signal $\underline{x}_k(n)$ is defined as

$$\underline{x}_k(n) = [x_k(n) \cdots x_k(n - N_w + 1)]^T$$

Defining the $[KN_w \times 1]$ composite input signal vector $\underline{x}(n)$ and the $[KN_w \times 1]$ composite weight vector \underline{w}_l as

$$\begin{aligned} \underline{x}(n) &= [\underline{x}_1(n) \cdots \underline{x}_K(n)]^T \\ \underline{w}_l &= [\underline{w}_{l1} \cdots \underline{w}_{lK}]^T, \end{aligned} \quad (5)$$

equation (4) can be written as

$$y_l(n) = \underline{x}^T(n) \underline{w}_l \quad (6)$$

From (1) and (6), the $[L \times 1]$ vector of control signals $\underline{y}(n)$ driving the reproduction loudspeakers is expressed as

$$\begin{bmatrix} y_1(n) \\ \vdots \\ y_L(n) \end{bmatrix} = \begin{bmatrix} \underline{x}(n) & \cdots & \underline{\mathbf{0}} \\ \vdots & \ddots & \vdots \\ \underline{\mathbf{0}} & \cdots & \underline{x}(n) \end{bmatrix}^T \begin{bmatrix} \underline{w}_1 \\ \vdots \\ \underline{w}_L \end{bmatrix}$$

Further, defining the $[KLN_w \times 1]$ composite vector $\underline{w} = [\underline{w}_1 \cdots \underline{w}_L]^T$, and the $[KLN_w \times L]$ composite matrix $\mathbf{x}(n)$ of input signals given by the first factor in the right hand side of this equation, it can be written compactly as

$$\underline{y}(n) = \mathbf{x}^T(n) \underline{w}$$

The system response vector $\hat{\underline{d}}(n)$ results from filtering the input to the loudspeakers $\underline{y}(n)$ through the matrix of electro-acoustic transfer functions $\mathbf{C}(\omega)$. The resulting component at the m^{th} microphone is, therefore, given by

$$\hat{d}_m(n) = \underline{c}_{m1} * \underline{y}_1(n) + \cdots + \underline{c}_{mL} * \underline{y}_L(n),$$

where all electro-acoustic impulse responses $\{\underline{c}_{ml} : m = 1, 2, \dots, M, l = 1, 2, \dots, L\}$ are assumed to be FIR filters of length N_c , defined as

$$\underline{c}_{ml} = [c_{ml,0} \cdots c_{ml,N_c-1}]^T$$

Defining the time history of the signal driving the l^{th} loudspeaker $y_l(n)$ as

$$\underline{y}_l(n) = [y_l(n) \cdots y_l(n - N_c + 1)]^T$$

and the $[M \times LN_c]$ composite matrix of acoustic impulse responses \mathbf{c} as

$$\mathbf{c} = \begin{bmatrix} \underline{c}_{11}^T & \underline{c}_{12}^T & \cdots & \underline{c}_{1L}^T \\ \underline{c}_{21}^T & \underline{c}_{22}^T & \cdots & \underline{c}_{2L}^T \\ \vdots & \vdots & \ddots & \vdots \\ \underline{c}_{M1}^T & \underline{c}_{M2}^T & \cdots & \underline{c}_{ML}^T \end{bmatrix},$$

the $[M \times 1]$ vector $\hat{\underline{d}}(n)$ of sound waves at the microphones can be expressed as

$$\hat{\underline{d}}(n) = \mathbf{c} * \underline{y}(n) = \mathbf{c} * [\mathbf{x}^T(n) \underline{w}] = \mathbf{x}_f(n) \underline{w},$$

where the $[M \times KLN_w]$ matrix $\mathbf{x}_f(n)$ is given by

$$\mathbf{x}_f(n) = \begin{bmatrix} \underline{x}_{f111}(n) & \cdots & \underline{x}_{f1LK}(n) \\ \underline{x}_{f211}(n) & \cdots & \underline{x}_{f2LK}(n) \\ \vdots & \ddots & \vdots \\ \underline{x}_{fM11}(n) & \cdots & \underline{x}_{fMLK}(n) \end{bmatrix} \quad (7)$$

and the vector $\underline{x}_{fmlk}(n) = \underline{c}_{ml} * \underline{x}_k(n)$ contains the last N_w samples of the result of filtering the k^{th} input signal through the electro-acoustic impulse response between the m^{th} microphone and the l^{th} loudspeaker. The error vector $\underline{e}(n)$ in Fig. 1 can then be expressed as

$$\underline{e}(n) = \hat{\underline{d}}(n) - \mathbf{x}_f(n) \underline{w} \quad (8)$$

Substituting in the performance index (3) gives

$$\begin{aligned} \xi(n) &= E\{\underline{d}^T(n) \hat{\underline{d}}(n) - 2 \underline{w}^T \mathbf{x}_f^T(n) \hat{\underline{d}}(n) + \\ &\quad + \underline{w}^T \mathbf{x}_f^T(n) \mathbf{x}_f(n) \underline{w}\} \end{aligned} \quad (9)$$

The optimum LMS solution $\underline{w}_{opt}(n)$ for the composite weight vector is the vector that minimises the performance index $\xi(n)$ given by (9). The optimum solution is, therefore, obtained by setting the first derivative (gradient) of $\xi(n)$ with respect to the composite weight vector \underline{w} to zero. The gradient is readily obtained from (9) as

$$\begin{aligned} \underline{\nabla}(n) &= \frac{\partial \xi(n)}{\partial \underline{w}} \\ &= 2 E\{\mathbf{x}_f^T(n) \mathbf{x}_f(n) \underline{w} - \mathbf{x}_f^T(n) \hat{\underline{d}}(n)\} \end{aligned} \quad (10)$$

The optimum LMS solution $\underline{w}_{opt}(n)$ is then obtained by setting $\underline{\nabla}(n)$ to zero resulting in:

$$E\{(\mathbf{x}_f^T(n) \mathbf{x}_f(n))^{-1}\} E\{\mathbf{x}_f^T(n) \hat{\underline{d}}(n)\}, \gamma \quad (11)$$

and the corresponding minimum value ξ_{min} is given by

$$E\{\hat{\underline{d}}^T(n) \hat{\underline{d}}(n)\} - E\{\underline{w}_{opt}^T(n) \mathbf{x}_f^T(n) \hat{\underline{d}}(n)\}.$$

From the first equation it follows that the optimum weight vector $\underline{w}_{opt}(n)$ exists only if the matrix $(\mathbf{x}_f^T(n) \mathbf{x}_f(n))$ is non-singular. Since the matrix $\mathbf{x}_f(n)$ is the convolution between the input signals and the electro-acoustic transfer functions as given by (7), not only $\underline{x}(n)$ but also $\mathbf{C}(\omega)$ influence the solution. Depending on the dimensions of $\mathbf{C}(\omega)$, three cases may be recognised:

1. The number of loudspeakers equals the number of microphones ($L = M$):

In this case, $\mathbf{C}(\omega)$ is a square matrix. Consider

$K = 1$ for simplicity and using frequency domain representations, the system of equations

$$\underline{\mathbf{E}}(\omega) = \underline{\mathbf{D}}(\omega) - \mathbf{C}(\omega)\underline{\mathbf{Y}}(\omega) \quad (12)$$

is fully determined. Provided $\mathbf{C}(\omega)$ is non-singular, a unique solution for the control vector $\underline{\mathbf{Y}}(\omega)$ exists, which drives the error vector exactly to $\underline{\mathbf{0}}$, namely $\underline{\mathbf{Y}}_{opt}(\omega) = \mathbf{C}^{-1}(\omega)\underline{\mathbf{D}}(\omega)$.

2. The number of loudspeakers is less than the number of microphones ($L < M$):

The matrix $\mathbf{C}(\omega)$ has more rows than columns and the system of equations (12) is over determined. There are more equations to solve than there are unknowns. In this case, provided that $(\mathbf{x}_f^T(n) \mathbf{x}_f(n))$ is positive definite, a unique global solution exists and is given by (11).

3. The number of loudspeakers is greater than the number of microphones ($L > M$):

There are less equations to solve than there are unknowns. The system of equations (12) is under determined and there exists an infinite number of solutions corresponding to infinite local minima on the error surface. In this case, extra constraints must be imposed to select one of those local solutions. A practical constraint may be limiting the power of the signals $\underline{\mathbf{y}}(n)$, driving the loudspeakers to avoid nonlinear distortion that may occur due to overloading of the loudspeakers. Equivalently, the same result may be achieved by imposing the constraint on the values of the coefficients of $\underline{\mathbf{w}}$. The performance index in this case becomes

$$\xi(n) = E\{\underline{\mathbf{e}}^T(n) \Gamma_e \underline{\mathbf{e}}(n) + \underline{\mathbf{w}}^T \Gamma_w \underline{\mathbf{w}}\},$$

where Γ_e and Γ_w are (often diagonal) weighting matrices. The optimum weight vector $\underline{\mathbf{w}}_{opt}(n)$ in this case is given by [6]:

$$E\{(\mathbf{x}_f^T(n) \Gamma_e \mathbf{x}_f(n) + \Gamma_w)^{-1}\} E\{\mathbf{x}_f^T(n) \Gamma_e \underline{\mathbf{d}}(n)\} \quad (13)$$

and the corresponding minimum value of ξ_{min} is given by

$$E\{\underline{\mathbf{d}}^T(n) \Gamma_e \underline{\mathbf{d}}(n)\} - E\{\underline{\mathbf{w}}_{opt}^T(n) \Gamma_e \mathbf{x}_f^T(n) \underline{\mathbf{d}}(n)\}$$

The weighting Γ_w has also proven to improve the stability of the solution in the fully determined and over determined cases [6]. Alternatively, the constraint may be imposed on the number of filter taps N_w , which enables an exact solution rather than a least mean square one [7].

6 The Multiple Error Filtered-X LMS Algorithm

An alternative approach to obtaining the optimum LMS solution, other than direct calculation using

(11) or (13), is to use an iterative algorithm such as the Multiple Error Least Mean Square (MELMS) algorithm [3, 8]. Such an adaptive approach is preferred over the direct calculation of $\underline{\mathbf{w}}_{opt}(n)$, since it offers *in-situ* design of the filters. It also enables a convenient method to readjust the filters whenever a change occurs in the electro-acoustic transfer functions. The MELMS algorithm employs the steepest descent approach to search for the minimum of the performance index (3). This is achieved by successively updating the filters' coefficients by an amount proportional to the negative of the gradient $\underline{\nabla}(n)$,

$$\underline{\mathbf{w}}(n+1) = \underline{\mathbf{w}}(n) + \mu(-\underline{\nabla}(n)), \gamma \quad (14)$$

where μ is the step size that controls the convergence speed and the final misadjustment [16]. An approximation often used in such iterative LMS algorithms is to update the vector $\underline{\mathbf{w}}$ using the instantaneous value of the gradient $\underline{\nabla}(n)$ instead of its expected value $\underline{\nabla}(n) = E\{\underline{\nabla}(n)\}$ [16], leading to the well-known LMS algorithm. Using (8) in (10), the gradient can be written as $\underline{\nabla}(n) = 2E\{-\mathbf{x}_f^T(n) \underline{\mathbf{e}}(n)\}$. The update equation for the MELMS algorithm is then given by replacing $\underline{\nabla}(n)$ in (14) by its instantaneous value,

$$\underline{\mathbf{w}}(n+1) = \underline{\mathbf{w}}(n) + 2\mu \mathbf{x}_f^T(n) \underline{\mathbf{e}}(n) \quad (15)$$

This update algorithm is often referred to as the Multiple Error Filtered-X (MEFX) algorithm. Implementation of (15) requires calculating the matrix $\mathbf{x}_f(n)$ given by (7), which implies measuring all the electro-acoustical transfer functions $\underline{\mathbf{c}}_{ml}$ and filtering each input signal through all ML transfer functions to construct the KLM elements of $\mathbf{x}_f(n)$. This is shown in Fig. 2, where the measured matrix

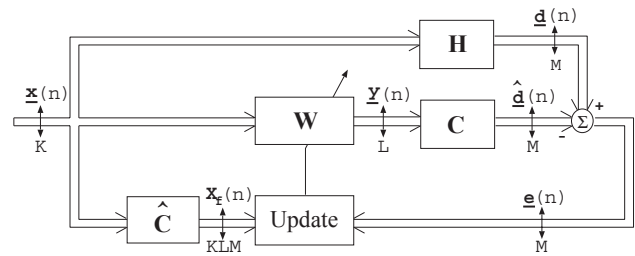


Figure 2: Multiple Error Filtered-X LMS (MEFX) algorithm.

of electro-acoustic transfer functions is represented by the block $\hat{\mathbf{C}}$ to distinguish it from the physical one represented by the block \mathbf{C} . The MEFX algorithm is known to be robust to estimation errors. It shows stable convergence properties as long as the phase error in any of the measured transfer functions at any frequency is less than $\pm 90^\circ$ [3, 12, 16],

while amplitude errors result in less accurate solutions [10]. Similarly, the noisy steepest descent method may be used to iteratively approach the optimum solution in the case of weighted performance index. In this case

$$\underline{\nabla}(n) = 2E\{-\mathbf{x}_f^T(n)\Gamma_e\mathbf{e}(n) + \Gamma_w\mathbf{w}\}$$

which leads to the following update equation

$$\mathbf{w}(n+1) = (\mathbf{I} - \mu\Gamma_w)\mathbf{w}(n) + 2\mu\mathbf{x}_f^T(n)\Gamma_e\mathbf{e}(n)$$

For $\Gamma_w = \text{diag}\{\gamma_1, \gamma_2, \dots, \gamma_{KLN_w}\}$, each filter weight w_i is independently weighted by the weighting factor γ_i . In this case, when $\Gamma_e = \mathbf{I}$, and letting $\Gamma = \mathbf{I} - \mu\Gamma_w$, the update equation becomes the multiple error leaky LMS algorithm

$$\mathbf{w}(n+1) = \Gamma\mathbf{w}(n) + 2\mu\mathbf{x}_f^T(n)\mathbf{e}(n).$$

7 Adaptive Active Noise Cancellation

The detailed implementation of the MEFX algorithm is further explained by unpacking the composite vector $\mathbf{w}(n)$ in (15) into its individual filters $\{\mathbf{w}_{lk}(n) : l = 1, 2, \dots, L, k = 1, 2, \dots, K\}$ giving

$$\mathbf{w}_{lk}(n+1) = \mathbf{w}_{lk}(n) + 2\mu \sum_{m=1}^M e_m(n)\mathbf{x}_{f_{lmk}}(n) \quad (16)$$

where $\mathbf{x}_{f_{lmk}}(n)$ is the result of filtering the k^{th} input through the measured electro-acoustic impulse response $\hat{\mathbf{C}}_{ml}$ between the m^{th} microphone and the l^{th} loudspeaker. The detailed implementation of (16) is visually illustrated in Fig. 3 for a $[K \times L \times M = 2 \times 2 \times 2]$ active noise control system. In this system, it is desired to reduce the sound field due to the primary sources P_1 and P_2 at two receiver points (microphones) R_1 and R_2 using an anti-sound field generated by the secondary loudspeakers S_1 and S_2 . Using frequency domain representations, the inputs to the secondary sources $Y_1(\omega)$ and $Y_2(\omega)$ are controlled by four adaptive filters $\{\mathbf{W}_{lk}(\omega) : l = 1, 2, k = 1, 2\}$ to achieve the above cancellation task. The desired response is the sound field generated by the primary sources P_1 and P_2 at the microphones' positions: $\mathbf{D}(\omega) = [D_1(\omega) D_2(\omega)] = \mathbf{H}(\omega)\mathbf{X}(\omega)$, where $\mathbf{H}(\omega)$ is the matrix of electro-acoustic transfer functions between the two microphones and the primary sources and $\mathbf{X}(\omega) = [X_1(\omega) X_2(\omega)]$. The microphones' outputs are, therefore, the *sum* of the sound fields due to the primary and the secondary sources $\mathbf{E}(\omega) = [E_1(\omega) E_2(\omega)] = \mathbf{D}(\omega) + \hat{\mathbf{D}}(\omega)$, where $\hat{\mathbf{D}}(\omega) = [\hat{D}_1(\omega) \hat{D}_2(\omega)] = \mathbf{C}(\omega)\mathbf{Y}(\omega)$ is the (unmeasurable) sound field due to the secondary sources alone at the microphones.

At each time sample, the update algorithm adjusts the coefficients of the adaptive filters to minimise the error signals measured by the microphones using the MEFX algorithm (16)³. Four update blocks are needed, one for each adaptive filter. These are indicated by the MEFX boxes in Fig. 3. According to (16), each of the update blocks requires two filtered input signals and in total $KLM = 8$ filtered input signals are needed. The filtered input signals are calculated by filtering each of the two input signals through the four measured impulse responses $\{\hat{\mathbf{C}}_{ml}(\omega) : m = 1, 2, l = 1, 2\}$ that are estimates of the physical electro-acoustic transfer functions between the secondary sources and the microphones represented in Fig. 3 by the matrix \mathbf{C} .

After successful convergence, the sound waves generated by S_1 and S_2 at R_1 and R_2 equal in magnitude and opposite in phase to those generated by P_1 and P_2 , and reduction in the net sound field at R_1 and R_2 results. This may be expressed mathematically as $\mathbf{H}(\omega)\mathbf{X}(\omega) = -\mathbf{C}(\omega)\mathbf{W}(\omega)\mathbf{X}(\omega)$, and the solution to the matrix of control filters is given by

$$\mathbf{W}(\omega) = -\mathbf{C}^{-1}(\omega)\mathbf{H}(\omega)$$

8 Adaptive Virtual Sound Sources

The system in Fig. 3 is exactly (except for a $-$ sign) what is needed to generate two virtual sound images at the positions of the primary sources P_1 and P_2 at the ears of one listener when the two microphones are positioned inside the listener's ear canals. By controlling the sound at $M = 2B$ microphones, the same sound images are perceived by B listeners, and by increasing the number of inputs, more images are created. This suggests a procedure for implementing the loudspeaker display system. With probe microphones inserted in the listener's ear canals, the set of adaptive filters are adjusted to cancel uncorrelated white noise signals from physical loudspeakers placed where virtual sound images are required⁴. After successful conversion, the physical primary sources are disconnected and the coefficients of the filters are multiplied by -1 to compensate for the sign difference mentioned above. Monophonic signals filtered through the previously obtained filters will move the auditory event to the positions where the primary sources have been. This procedure has been successfully used in [1, 14] and proved

³Since (16) was derived assuming that the error is formed by the difference between rather than the sum of $\mathbf{D}(\omega)$ and $\hat{\mathbf{D}}(\omega)$, the $+$ sign in (16) must be changed to $-$ sign.

⁴The adaptation process requires first measuring all electro-acoustic transfer functions $\hat{\mathbf{C}}_{ml}$. On-line estimation of these transfer functions is discussed in detail in [4]

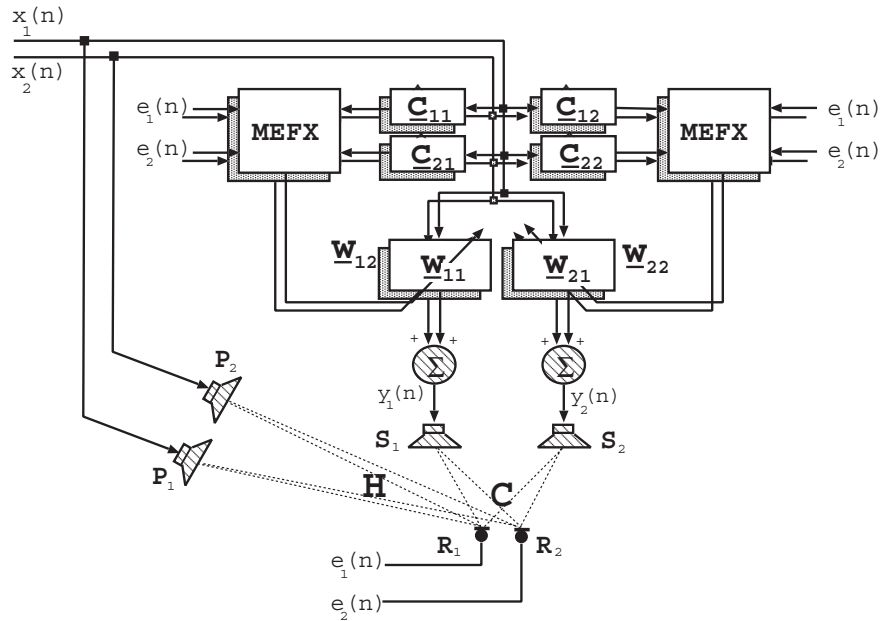


Figure 3: MEFX algorithm for a $[K \times L \times M = 2 \times 2 \times 2]$ system in an active noise cancellation setting.

to have the following advantages compared to the other methods:

- The listener's own HRTFs are used to design the filters and therefore correct spectral cues are maintained.
- The solution also includes the room impulse response and therefore it maintains the distance cues and the environmental context. This solves the In-Head-Localisation (IHL) problem and eliminates the need for artificial reverberation.
- Direct inversion of the electro-acoustic transfer function matrix $C(\omega)$ that is required for the cross-talk cancellation is avoided by iteratively searching the optimum solution in the least mean square sense.
- The binaural synthesis (convolution with the HRTF matrix $H(\omega)$) and the cross-talk cancellation (inversion of $C(\omega)$) are combined. This is both more efficient and numerically more stable.
- The problem is mapped from the difficult domain of virtual sound image synthesis to the well-developed one of multichannel active noise control (ANC). This not only allows employing the techniques used in ANC systems, but also facilitates describing the system performance in terms of the sound attenuation achieved at the listener's eardrums. The larger this attenuation, the more realistic the virtual source is perceived.
- Active noise control systems are real-time systems. Therefore, the above procedure allows real-time design and implementation of the system's filters. This differs from the commonly used methods of measuring and storing a set of HRTFs, calculating the cross-talk cancellation filters, and real-time

filtering of the audio signals through those previously calculated fixed filters.

In spite of the above mentioned advantages, adaptive filters approach to virtual source synthesis suffers the following drawbacks:

- Including the room impulse response in the filters limits the virtual space to the same measurement environment.
- The listener is asked to insert a pair of microphones inside his/her ears, which may be objectionable. However, this is the only approach to obtaining individualised HRTFs.
- White noise or chirp signals that are spectrally rich must be used in the identification and adaptation stages.
- The electro-acoustic transfer functions $C(\omega)$ and $H(\omega)$ are very complex functions of frequency and space coordinates. The impulse response of a room may also last for several hundreds of milliseconds. At a sampling frequency of 44.1 kHz, thousands of FIR coefficients are required to properly model these transfer functions. The adaptive filters are, therefore, of high order, which makes system implementation in real-time a great challenge. Efficient implementation of the filtering and adaptation operations are, therefore, essential (see Section 10).
- At high frequencies, the acoustic wavelength is very small. Therefore, the solution obtained by the adaptive process is valid only in a very small area in space. Therefore, the listener's head must be fixed during the adaptation and filtering operations.

9 Adaptive Cross-Talk Cancellation

The cross-talk cancellation may also be realised using adaptive filters. The system shown in Fig. 3 can be used for this purpose after a few modifications. In cross-talk cancellation, it is desired to reproduce $x_1(n)$ at R_1 and $x_2(n)$ at R_2 . The desired response is, therefore, given by $\underline{d}(n) = \underline{x}(n)$, and the adaptive filters are required to model the inverse of the matrix $\mathbf{C}(\omega)$. For the inverse to be realisable using FIR filters, delayed input signals are used to calculate the desired responses,

$$\begin{bmatrix} d_1(n) \\ d_2(n) \end{bmatrix} = \begin{bmatrix} x_1(n - \Delta_1) \\ x_2(n - \Delta_2) \end{bmatrix},$$

where Δ_1 and Δ_2 are delays that are assumed to be longer than that introduced by the electro-acoustic transfer functions comprising $\mathbf{C}(\omega)$. Since the microphones' outputs in this application are due to S_1 and S_2 only, they correspond to the vector $\hat{\underline{d}}(n)$ in Fig. 1. The error signals, to be minimised by the adaptive algorithm, are then constructed by *electrically* subtracting $\hat{\underline{d}}(n)$ from $\underline{d}(n)$

$$\begin{bmatrix} e_1(n) \\ e_2(n) \end{bmatrix} = \begin{bmatrix} x_1(n - \Delta_1) - \hat{d}_1(n) \\ x_2(n - \Delta_2) - \hat{d}_2(n) \end{bmatrix}.$$

Provided the FIR filters \mathbf{W} are of sufficiently high order to accommodate the inverse solution, the mean square error is minimised by the adaptive algorithm. After conversion, the sound pressure at the microphones R_1 and R_2 are the best least squares estimations of $x_1(n - \Delta_1)$ and $x_2(n - \Delta_2)$, respectively.

10 Efficient Implementations

In reverberant acoustic environments, the impulse response between two points may last for several hundreds of milliseconds. At the standard audio compact disc sampling frequency of 44.1 kHz, thousands of FIR coefficients are needed to properly model and store such an impulse response. Adaptive multichannel audio reproduction systems require estimating, filtering through, and updating the coefficients of many of these large filters. This implies a huge amount of computational power which makes real-time implementation on reasonable hardware resources a difficult task. Reducing the system complexity is, therefore, essential for real-time implementation. In this section, two approaches are introduced to decrease the number of required calculations. The adjoint LMS algorithm discussed in Section 10.1 uses filtered error signals rather than filtered input signals in the update process. This leads to a huge computational saving as the system dimensions increase. Significant saving may further be achieved

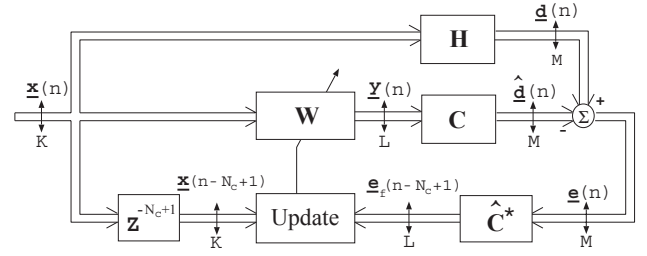


Figure 4: Adjoint Least Mean Square Algorithm.

by implementing the convolution and correlation operations in the frequency domain using block processing techniques. This leads to the Block Frequency Domain Adaptive Filters (BFDAP) discussed in Section 10.2.

10.1 The Adjoint LMS Algorithm

Setting the electro-acoustic transfer function matrix $\mathbf{C}(\omega)$ in Fig. 2 at each frequency to be the identity matrix \mathbf{I} results in a regular adaptive filter system [16]. In this case, $\hat{\underline{d}}(n) = \underline{y}(n)$ and the adaptive filters' outputs are directly observable from the error vector $\underline{e}(n) = \underline{d}(n) - \underline{y}(n)$. The steepest descent update given by (15) becomes

$$\underline{w}(n+1) = \underline{w}(n) + 2\mu \mathbf{x}^T(n) \underline{e}(n), \gamma \quad (17)$$

where $\mathbf{x}(n)$ is defined in (5). The adaptive process in (17) basically aims at minimising the cross correlation between the input signals $\mathbf{x}(n)$ and the error signals $\underline{e}(n)$. When this cross correlation reaches its minimum value, the filters are as close as possible to their optimal solutions. For correct adaptation, it is essential that the cross correlation is performed between correctly aligned history samples of $\underline{e}(n)$ and $\mathbf{x}(n)$.

Introducing the matrix $\mathbf{C}(\omega)$ into the system results in $\hat{\underline{d}}(n)$ being a filtered version of $\underline{y}(n)$. This has two severe consequences on the adaptation process. The first is that the cross correlation is performed between two misaligned time sequences, since $\underline{y}(n)$ is delayed while propagating through the matrix of acoustical transfer function $\mathbf{C}(\omega)$, which leads to an unstable adaptive algorithm. The second is the spectral deformation caused by $\mathbf{C}(\omega)$. The effect of the latter may be explained by the extreme case of each of $\{\mathbf{C}_{ml}(\omega_0) : l = 1, 2, \dots, L\}$ has a zero response at the same frequency ω_0 . In this case, $e_m(n)$ contains no information about the filters' outputs at ω_0 , resulting in an unobservable system at that frequency.

A stable adaptive algorithm with $\mathbf{C}(\omega) \neq \mathbf{I}$ may only be obtained if the above mentioned filtering effects are compensated. The MEFX algorithm discussed in Section 6 solves this problem by filter-

ing $\underline{x}(n)$ through estimates of $\mathbf{C}(\omega)$ prior to using them as inputs to the adaptation process as shown in Fig. 2. This effectively delays the inputs to be correctly aligned in time with the error signals, and at the same time introduces the required spectral correction. Since the dimensions of matrices may not match and matrices do not commute, the MEFX effectively filters every input signal $\underline{x}_k(n)$ with every electro-acoustic response \underline{c}_{ml} to construct the KLM elements of the matrix $\mathbf{x}_f(n)$. This requires in total KLM convolution operations only for constructing $\mathbf{x}_f(n)$. Assuming time domain implementation, the total number of multiplications required per iteration for calculating $\mathbf{x}_f(n)$ is $KLMN_c$. Updating the weights requires $KLN_w(M+1)$ multiplications, and calculating the filters' outputs requires KLN_w multiplications. The total number of multiplications required to implement the MEFX algorithm at each iteration is, therefore,

$$\Psi_{MEFX} = KL[(N_w + N_c)M + 2N_w] \quad (18)$$

which increases rapidly with increasing system dimensions. A much more efficient algorithm may be obtained by advancing the error signals $\underline{e}(n)$ in time to be aligned with $\mathbf{x}(n)$. This leads to the Adjoint Least Mean Square (ALMS) algorithm [15], which has its update equation given by (with $n' = n - N_c + 1$):

$$\underline{\mathbf{w}}(n+1) = \underline{\mathbf{w}}(n) + 2\mu \mathbf{x}^T(n') \underline{\mathbf{e}}_f(n') \quad (19)$$

where $\underline{\mathbf{e}}_f = [\underline{e}_{f_1} \quad \underline{e}_{f_2} \quad \dots \quad \underline{e}_{f_L}]^T$ is the result of filtering the error signals through the *adjoint* (time mirror) of the transfer functions comprising $\mathbf{C}(\omega)$. This non-causal operation can only be performed in real-time after a delay of $N_c - 1$ samples. The reason for using delayed input and filtered error signals is given in (19). The time mirror operation is equivalent to calculating the complex conjugate of the frequency response as shown in Fig. 4. Expressing the adjoint of \underline{c}_{ml} as

$$\check{\underline{c}}_{ml} = [\underline{c}_{ml, N_c-1} \quad \dots \quad \underline{c}_{ml, 0}]^T$$

the filtered error signal vector \underline{e}_{f_i} may be written as (with $n' = n - N_c + 1$)

$$\underline{e}_{f_i}(n') = \check{\underline{c}}_{1l} \underline{e}_1^T(n) + \dots + \check{\underline{c}}_{Ml} \underline{e}_M^T(n)$$

The ALMS algorithm reduces the number of multiplications required for updating the adaptive weights to $2KLN_w$. Adding LMN_c multiplications for calculating $\underline{e}_f(n)$, and KLN_w multiplications for calculating the filters' outputs sums to

$$\Psi_{ALMS} = KL \left[\frac{N_c}{K} M + 3N_w \right] \quad (20)$$

Comparing (20) with (18) shows that for a $[K \times L \times M = 1 \times 1 \times 1]$ system, the ALMS and MEFX have

the same complexity of $(3N_w + N_c)$ multiplications. For multichannel systems, however, the ALMS algorithm is much more efficient than the MEFX.

A detailed implementation of the ALMS algorithm may be seen from the update equation of an individual filter $\underline{\mathbf{w}}_{lk}(n)$, which is given by (with $n' = n - N_c + 1$)

$$\underline{\mathbf{w}}_{lk}(n+1) = \underline{\mathbf{w}}_{lk}(n) + 2\mu e_{f_l}(n') \underline{\mathbf{x}}_k(n')$$

This update is shown in Fig. 5 for a $[K \times L \times M = 2 \times 2 \times 2]$ system in an active noise control setting similar to that in Fig. 3. A comparison of Fig. 5 and Fig. 3 reveals the structural simplicity offered by the ALMS over the MEFX, in addition to the computational saving mentioned above. Two sources of computational saving may be recognised in this example. The first is the reduction of the number of convolutions required to calculate the filtered signals (4 in case of ALMS against 8 in MEFX). The second is the simplified ALMS update that always uses a single error signal and a single input signal compared to 2 error and 2 input signals in the MEFX case.

10.2 Frequency Domain Implementations

Implementation of the MEFX (Fig. 2) or ALMS (Fig. 4) requires performing three main tasks:

1. Calculating the adaptive filters' outputs $\underline{\mathbf{y}}(n)$.
2. Calculating the filtered inputs $\mathbf{x}_f(n)$ (or errors $\underline{\mathbf{e}}_f(n)$).
3. Updating the coefficients of the adaptive filters $\underline{\mathbf{W}}$.

The first two tasks are convolution operations between $\underline{\mathbf{x}}(n)$ and FIR filters comprising $\underline{\mathbf{W}}(\omega)$ and $\mathbf{C}(\omega)$, respectively. The third task implies calculating the instantaneous gradient that is essentially a crosscorrelation between the input and error signals. Since all FIR filters are of high order, implementing the convolution and correlation in the frequency domain results in considerable computational saving [9]. Since the input signals are infinitely long, real-time implementation of the convolution or correlation in the frequency domain must be performed on successive overlapping blocks of data. Two known and frequently used methods to construct and process such overlapping blocks of data are the overlap-add and overlap-save methods [9]. Applying any of those block processing techniques to the LMS adaptive process leads to the Block Frequency Domain Adaptive Filter (BFDAF) [2, 13].

Besides the speed gained by carrying out the convolution and correlation operations in the frequency domain, BFDAF offers the possibility of approximate decorrelation of the adaptive process from the input signals' statistics. This is done by normalising each frequency bin by the input power

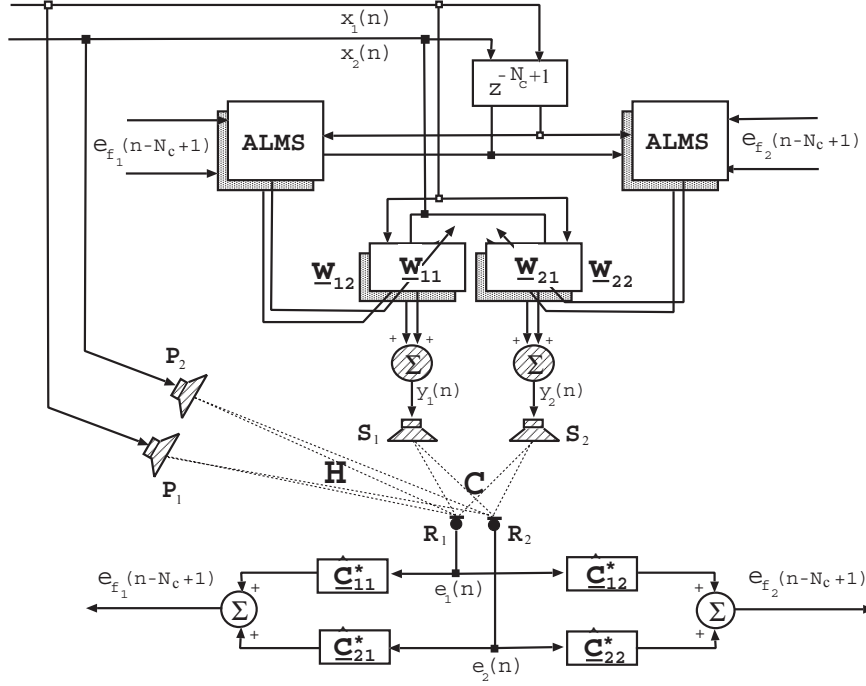


Figure 5: ALMS algorithm for a $[K \times L \times M = 2 \times 2 \times 2]$ system in an active noise cancelling setting (refer with Fig. 3)

in that bin [13]. Unlike the systems discussed previously which update the coefficients of the adaptive filters every sample, BFDFAF performs such update every $L \geq 1$ samples, which leads to lower complexity. Due to block processing, reduction in the step size upper bound and processing delay of L samples are also introduced. In this section, implementations of 3D sound systems using BFDFAF are considered. Since the adjoint LMS algorithm is more efficient for multichannel systems as shown in Section 10.1, only BFDFAF implementation of that algorithm is considered below. For clarity, the BFDFAF implementation of the ALMS algorithm is illustrated for a single channel active noise control system, since the extension to multichannel systems is straight-forward as given by (19). The block diagram of the ALMS algorithm implemented using overlap-save BFDFAF for $[K \times L \times M = 1 \times 1 \times 1]$ ANC system is shown in Fig. 6. In this block diagram, vectors are represented by double parallel lines while scalars are represented by single lines. Since we are dealing with a single channel case, vectors are used here to represent frequency samples of the signals rather than multiple signals at the same frequency.

The overlap-save method requires sectioning the stream of input time samples $x(n)$ into overlapping blocks, each of length N_B . Each block contains L new samples and $N_w + N_c - 2$ samples from the previous block. The sectioning and overlapping operations are represented in Fig. 6 by the *serial-*

to-parallel and *overlap* blocks. For each L samples, a complete block of time samples $\underline{\mathbf{x}}_B(nL)$ is constructed and processing of this block of data may be commenced. For correct real-time operation, the processing of each block must be completed before the next block is constructed, i.e. in a maximum of L samples. To perform convolution and correlation in the frequency domain, each block $\underline{\mathbf{x}}_B(nL)$ is transformed to the frequency domain using a length $N_B = N_w + N_c + L - 2$ Fast Fourier Transform (FFT) algorithm, $\underline{\mathbf{X}}(nL) = \mathbb{F}_{N_B} \underline{\mathbf{x}}_B(nL)$. The length of the block N_B is chosen such that the circular correlation performed by element-wise multiplication in the frequency domain results in N_w correct time domain correlation coefficients as will be shown shortly. The processing of each block can be divided into three main tasks, as mentioned above:

1. Calculation of the filter's output $y(n - L + 1)$: The convolution between the transformed input block $\underline{\mathbf{X}}(nL)$ and the frequency domain adaptive filter weights $\underline{\mathbf{W}}(nL)$ is carried out by element-wise multiplication. Defining the matrix $\underline{\mathbf{X}}(nL) = \text{diag}\{\underline{\mathbf{X}}(nL)\}$, this convolution may be represented by

$$\underline{\mathbf{Y}}(nL) = \underline{\mathbf{X}}(nL) \underline{\mathbf{W}}(nL)$$

The vector $\underline{\mathbf{Y}}(nL)$ is then transformed back into the time domain using an Inverse Fast Fourier Transform (IFFT) algorithm of length N_B . The last L samples of this time domain vector represent the

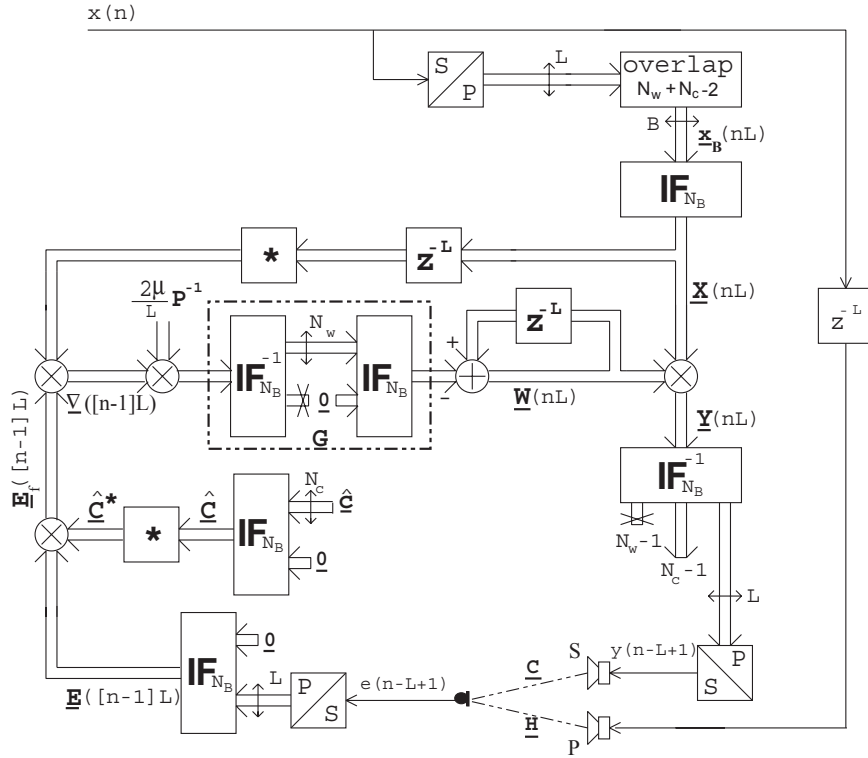


Figure 6: Block Frequency Domain implementation of the ALMS algorithm for a single channel $[K \times L \times M = 1 \times 1 \times 1]$ system.

required linear convolution. During processing the next block, these L samples are sent, one sample at each clock cycle, to drive the loudspeaker S . This is represented in Fig. 6 by the *parallel-to-serial* block. The first $N_w - 1$ samples represent cyclic convolution results and must be discarded. An extra $N_c - 1$ samples represent correct linear convolution results but are not used since only L samples are needed to be sent out at every block. This extra $N_c - 1$ samples are due to the choice of a large block length N_B to accommodate successive convolution and correlation as described below.

2. Calculation of the filtered error $\underline{\mathbf{E}}_f([n-1]L)$:

The microphone signal $e(n-L+1)$ is buffered into a length L vector. This vector is augmented with $N_B - L$ leading zeros and transformed into the frequency domain using an FFT of length N_B , resulting in the block error vector $\underline{\mathbf{E}}([n-1]L)$. A previously measured impulse response $\hat{\mathbf{c}}$ of length N_c between loudspeaker S and the microphone is padded with $N_B - N_c$ trailing zeros and transformed to the frequency domain using an FFT of length N_B , producing the vector $\hat{\mathbf{C}}$. The adjoint (time mirror) operation required for the ALMS is performed by calculating the complex conjugate of $\hat{\mathbf{C}}$. The filtered error vector $\underline{\mathbf{E}}_f([n-1]L)$ is then calculated by element-wise multiplication of $\hat{\mathbf{C}}^*$ and

$\underline{\mathbf{E}}([n-1]L)$. The length of the result of this convolution is $L + N_c - 1$, and since N_B is chosen larger than this length, all samples of $\underline{\mathbf{E}}_f([n-1]L)$ represent correct linear convolution results. Defining the matrix $\hat{\mathbf{C}} = \text{diag}\{\hat{\mathbf{C}}\}$, this convolution may be expressed mathematically as

$$\underline{\mathbf{E}}_f([n-1]L) = \hat{\mathbf{C}}^* \underline{\mathbf{E}}([n-1]L)$$

Since the microphone signal $e(n-L+1)$ is due to the output samples $y(n-L+1)$ from the previous block, the error signal is already delayed by L samples. Provided $L \geq N_c$, this delay will suffice to implement the delay required by the ALMS as mentioned in Section 10.1. To further ensure that the optimum filter solution is causal, the delay in the path through the loudspeaker P must be longer than that through the loudspeaker S . Since the algorithm introduces a delay of L samples, this delay must be compensated. Assuming that the acoustic transfer function $\underline{\mathbf{H}}$ possesses longer delay than $\underline{\mathbf{C}}$, this compensation may be done by delaying the signal to loudspeaker P by L samples. This is represented in Fig. 6 by the box z^{-L} . An elegant alternative is to move loudspeaker P an equivalent distance away from the microphone to save DSP memory.

3. Updating the filter weights:

Transforming the update equation of the ALMS

(19) to the frequency domain results in the following block frequency domain update equation ⁵ (with $\underline{\nabla}([n-1]L) = \mathbf{X}^*([n-1]L) \underline{\mathbf{E}}_f([n-1]L)$)

$$\begin{aligned} \underline{\mathbf{W}}(nL) &= \underline{\mathbf{W}}([n-1]L) - \\ &\quad - \frac{2\mu}{L} \mathbf{G} \mathbf{P}^{-1} \underline{\nabla}([n-1]L), \end{aligned} \quad (21)$$

where it is assumed that $L \geq N_c$, the delay required by the ALMS algorithm. The diagonal matrix \mathbf{P} represents an estimate of the input signal power at each frequency bin and performs the decorrelation mentioned above. The inverse of this power matrix is scaled by a step size of $2\mu/L$ and the result is used as the new step size. This frequency-dependent step size vector is used to scale the estimated gradient vector. The instantaneous block mean square gradient estimate vector given by the cross correlation $\underline{\nabla}([n-1]L) = \mathbf{X}^*([n-1]L) \underline{\mathbf{E}}_f([n-1]L)$ is obtained by first delaying $\underline{\mathbf{X}}(nL)$, calculating its complex conjugate, and element-wise multiplication of the result by the block filtered error as shown in Fig. 6. Since $\underline{\mathbf{X}}([n-1]L)$ is not padded with zeros, this element-wise multiplication corresponds in the time domain to N_w linear correlation coefficients, while the last $N_B - N_w$ samples are the cyclic correlation coefficients and must be discarded. However, the correct correlation coefficients must be extracted in the time domain, which is done using the $N_B \times N_B$ constraining window \mathbf{G} . This window transforms the estimated gradient to the time domain, replaces the last $N_B - N_w$ samples by zeros, and transforms the result back to the frequency domain. Finally, the weighted and constrained gradient is used to update the weight vector $\underline{\mathbf{W}}([n-1]L)$ according to (21).

Finally, it is worth mentioning that the frequency domain implementation of a single channel system using the adjoint LMS may lead to more computational saving over that using the filtered- x algorithm if N_w is much larger than N_c . This second source of computational saving stems from the fact that the filtered- x algorithm calculates \mathbf{x}_f by filtering the infinitely long input signal through the finite impulse response $\underline{\mathbf{C}}$. Implementing this convolution in the frequency domain requires the use of an overlapping method, and possibly two extra FFTs to separate the correct convolution samples in the time domain. On the other hand, the frequency domain adjoint LMS algorithm calculates $\underline{\mathbf{E}}_f$ by filtering (the already in block form) error signal through $\underline{\mathbf{C}}^*$ and no overlapping is needed. BFDAF implementations of the single channel adjoint LMS and several implementations of the single channel filtered- x have already been presented

⁵Since signals are summed up at the microphones in Fig. 6 rather than subtracted, the + sign in (19) has been changed to a - sign in (21).

in the previous works [1, 14]. A complexity comparison between these implementations shows that for a single channel BFDAF, the filtered- x is more efficient than the adjoint LMS when $N_w < 2N_c$, while the adjoint LMS is much more efficient when $N_w > 2N_c$.

11 Conclusions

One of the research areas of the Signal Processing group at the Eindhoven University of Technology is *adaptive array signal processing*. It has been shown that these adaptive array signal processing techniques can also be applied to audio applications as well. The current paper, based on the results of the PhD work [4], discusses theoretical and implementation issues concerning adaptive multichannel audio reproduction systems.

First a general model was derived that is capable of describing a wide class of multichannel audio reproduction system applications including synthesis of virtual sound sources, cross-talk cancellation and active noise control. The optimum least squares solution was given and from this an adaptive solution was derived resulting in the Multiple Error Filtered-X Least Mean Square (MEFX) algorithm. Efficient implementations have been discussed.

References

- [1] R.M. Aarts, P.C.W. Sommen, A.W.M. Mathijssen, and J. Garas, *Efficient Block Frequency Domain Filtered-x applied to Phantom Sound Source Generation*, AES 104th convention, Amsterdam, May 1998, Preprint No. 4650.
- [2] G.A. Clark, S.R. Parker, and S.K. Mitra, *A Unified Approach to Time- and Frequency- Domain Realization of FIR Adaptive Digital Filters*, IEEE Trans. on Acoust., Speech and Sig. Proc., Vol. ASSP-31, No. 5, Oct. 1983, pp. 1073-1083.
- [3] S.J. Elliott, I.M. Stothers, and P.A. Nelson, *A Multiple Error LMS Algorithm and Its Application to the Active Control of Sound and Vibration*, IEEE Trans. on Acoust., Speech and Sig. Proc., Vol. ASSP-35, No. 10, Oct. 1987, pp. 1423-1434.
- [4] J. Garas, *Adaptive 3D Sound Systems*, PhD dissertation, Eindhoven University of Technology, The Netherlands, September 1999, ISBN 90-386-1640-6.
- [5] D. H. Johnson and D. E. Dugeon *Array Signal Processing: Concepts and Techniques* Prentice Hall Canada, 1993, ISBN 0-13-048513-6.
- [6] O. Kirkeby, P.A. Nelson, H. Hamada, and F. Orduna-Bustamante, *Fast Convolution of*

Multichannel Systems Using Regularization, IEEE Trans. on Speech and Audio Proc., Vol. 6, No. 2, March 1998, pp. 189-194.

- [7] M. Miyoshi and Y. Kaneda, *Inverse Filtering of Room Acoustics*, IEEE Trans. on Acoust., Speech and Sig. Proc., Vol. ASSP-36, No. 2, Feb. 1988, pp. 145-152.
- [8] P.A. Nelson, F. Orduna-Bustamante, and H. Hamada, *Multichannel Signal Processing Techniques in the Reproduction of Sound*, J. Audio Eng. Soc., Vol. 44, No. 11, Nov. 1996, pp. 973-989.
- [9] A.V. Oppenheim and R.W. Schaffer, *Discrete-Time Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1989, ISBN 0-13-216-771-9.
- [10] N. Saito and T. Sone, *Influence of Modeling Error on Noise Reduction Performance of Active Noise Control Systems using Filtered-X LMS Algorithm*, J. Acoust. Soc. Jpn. (E), Vol. 17, No. 4, 1996, pp. 195-202.
- [11] D.W.E. Schobben, *Efficient Adaptive Multichannel Concepts in Acoustics: Blind Signal Separation and Echo Cancellation*, PhD dissertation, Eindhoven University of Technology, The Netherlands, September 1999, ISBN 90-386-1630-9.

CV

Piet Sommen received in 1981 the Ingenieur degree in electrical engineering from Delft University of Technology, The Netherlands and his Ph.D. in 1992 from Eindhoven University of Technology, The Netherlands.



From 1981 until 1989 he joined Philips Research Laboratories, Eindhoven. Here he was first engaged in research on CAD for circuit design, while later he worked in a radio and data transmission group where his main interest was on adaptive signal processing. Since 1989 he joins the faculty of electrical engineering from Eindhoven University of Technology, where he is currently an associate professor.

From January 1993 until June 1995 he was a member of the faculty board as a research dean. Besides three internal university courses Dr. Sommen is also involved with three external courses, all dealing with different new subjects in signal processing (adaptive filters, adaptive arrays, filterbanks, noise shaping, etc.). His main field of

- [12] S.D. Snyder and C.H. Hansen, *The Influence of Transducer Transfer Functions and Acoustic Time Delays on the Implementation of the LMS Algorithm in Active Noise Control Systems*, J. of Sound and Vibration, Vol. 141, No. 3, 1990, pp. 409-424.
- [13] P.C.W. Sommen, *Adaptive Filtering Methods*, PhD dissertation, Eindhoven University of Technology, The Netherlands, June 1992, ISBN 90-9005143-0.
- [14] P.C.W. Sommen, R.M. Aarts, A.W.M. Mathijssen, J. Garas, and H. He, *Efficient Frequency Domain Filtered-x Realization of Phantom Sources*, Proc. CCSP-97, Mierlo, The Netherlands.
- [15] E.A. Wan, *Adjoint LMS: An Efficient Alternative to the Filtered-X LMS and Multiple Error LMS Algorithm*, Proc. ICASSP-96, pp. 1842-1845.
- [16] B. Widrow and S.D. Stearns, *Adaptive Signal Processing*, Prentice-Hall Inc., Englewood Cliffs, NJ, 1985, ISBN 0-13-004029-0

interest is signal processing, with the emphasis on research and applications of adaptive array signal processing.

Dr. Sommen is member of the ProRISC board, he is vice-president of the IEEE Benelux Signal Processing Chapter and he is officer of the Administrative Committee of EURASIP.

CV

John Garas received his B.Sc. degree in Electronics and Communication from Ain Shams University, Cairo, Egypt in 1989.



He joined the Electronics Research Institute in Cairo in 1991 as a research assistant in the field of Solar Energy Systems.

In May 1994, he received his M.Sc. degree in Electronic Engineering from Eindhoven University of Technology, The Netherlands.

He further pursued his studies towards his Ph.D. degree at the Electronic Signal Processing Systems Group, Faculty of Electrical Engineering, Eindhoven University of Technology. His Ph.D. thesis concerned real-time generation of robust phantom audio sources using adaptive systems.

Currently he is with TNO Institute of Applied Physics, Active Noise and Vibration Control, Acoustic Division.

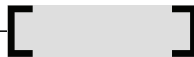
Digitale signaalbewerking in hoortoestellen

Jos Leenen

Beltone Netherlands b.v. / Philips Hearing Instruments

Afdeling R&D, Eindhoven

(E-mail: LeenenJ@Beltone.com)



1. Doelstelling

Dit artikel beoogt een overzicht te geven van de electronica die gebruikt wordt in moderne digitale hoortoestellen en de diverse bewerkingen die daarmee worden uitgevoerd op het inkomende geluidssignaal, alsmede de ontwikkeling daarvan en trends daarin.

Daarnaast wordt een beschrijving gegeven van het nieuwste DSP-IC voor hoortoestellen van Beltone / Philips dat in het najaar van 1999 op de markt kwam in de 'SpaceLine' hoortoestellenlijn.

2. Historisch overzicht

De oudste hulpmiddelen die men bij het horen gebruikte, naast de hand achter het oor, waren hoornvormige toeters. Mits deze niet lekkend op de gehoorgang aansluiten valt hiermee nog een verbazingwekkend redelijk resultaat te bereiken, wat veroorzaakt wordt door een combinatie van eigenschappen van een hoorn die voor de meest voorkomende vormen van slechthorendheid gunstig uitpakken. De versterking, op basis van resonantie (overigens zeker niet toereikend bij zwaardere hoorverliezen) is bandbegrensd, waarbij de belangrijkste frequenties voor het verstaan van spraak, zo'n 1 tot 3 kHz, worden bevoordeeld [1]. Bovendien is er sprake van een redelijke richtwerking in dit frequentiegebied, wat in een lawaaige omgeving, die voor slechthorenden extra moeilijk is, van groot belang is. Ook vandaag de dag horen filtering van het frequentiespectrum en richtinggevoeligheid tot de belangrijkste eigenschappen van hoortoestellen.

De ontwikkeling van de electronica werd natuurlijk ook in dit vakgebied al snel allesoverheersend. Bij Philips begonnen meer dan 60 jaar geleden 'knutselende radio-ingenieurs' voor hun slechthorende ouders met microfoons, versterkers en hoofdtelefoons 'hoortoestellen' te maken die ongetwijfeld niet erg handzaam zijn geweest. Pas rond de Tweede Wereldoorlog, toen er voor militaire walkie-talkies miniatuur electronenbuizen ontwikkeld waren, begon er in korte tijd een hoortoestellen-industrie te ontstaan. Het eerste commerciële hoortoestel van Beltone, Chicago, kwam in 1940 op de markt, het eerste van Philips in april 1949 [1].

De komst van de transistor (ontdekt 1948, in Nederland commercieel toegepast vanaf ongeveer 1953 [2]) werd door deze nieuwe industrie uiteraard als een geschenk uit de hemel ervaren. Het grootste voordeel was aanvankelijk de veel kleinere afmeting van de benodigde batterijen, maar toen de transducers ook kleiner werden speelde al snel de integratie-dichtheid van de electronica de hoofdrol. Met de komst van de geïntegreerde schakelingen kwamen de ontwikkelingen in een nog hogere versnelling. Een van de allereerste door Philips commercieel toegepaste IC's (de OM200 in 1966) werd onmiddellijk in hoortoestellen gebruikt. Dit IC bevatte 3 transistoren en 2 weerstanden in een zeer kleine behuizing waaruit 4 draden staken.

Al snel ging de ontwikkeling van in de zak gedragen toestellen met oortelefoontje, via achter het oor, naar in het oor gedragen toestellen. Dat in Europa de achter het oor gedragen toestellen tot vandaag de norm bleven ligt meer aan de manier van verstrekking en de landelijke vergoedingstelsels. Een in het oor gedragen toestel moet met

relatief veel moeite in een individueel gemaakte schaalvormige behuizing worden ingebouwd, terwijl bij een achter het oor gedragen toestel vanwege de standaardbehuizing fabrieksmatig geproduceerd en getest kan worden hetgeen de totale kosten en het aantal 'moeilijke handelingen' aanmerkelijk reduceert.

In 1983 was de IC-technologie zover gevorderd dat er voor het eerst digitale signaal processing (DSP) in consumenten producten toegepast kon worden (in het FIR filter vóór de DAC in Philips' CD100 compact disc speler). Voor toepassing van DSP in hoortoestellen was er toen nog een lange weg te gaan. Het benodigde elektrisch vermogen voor een compleet digitaal audio-systeem bestaande uit een AD-converter (ADC), een DSP met voldoende geheugen en een DA-converter (DAC) was nog tienduizenden malen te hoog voor een hoortoestel en ook de benodigde afmetingen zouden compleet onacceptabel geweest zijn, nog geheel afgezien van de batterijen.

Maar de lange weg werd met grote snelheid afgelegd. Al rond 1985 kwamen de eerste DSP-hoortoestellen in de USA op de markt, o.a. de 'Phoenix' van Nicolet in 1987. Deze bestond uit een achter het oor gedragen 'ear-piece' dat de transducers bevatte en dat met een kabeltje verbonden was met een in de zak gedragen kastje met de processor, andere electronica en de batterij. 'We gaan weer terug naar het zak-apparaat' werd er gegruip. De prijs van deze systemen was meer dan 10.000 gulden en geen enkel project uit die tijd bleek commercieel succesvol, maar ze spraken technisch wel enorm tot de verbeelding, ook buiten de hoortoestellen-wereld. Dankzij de grote ontwikkelsnelheid van de IC-technologie was iedereen er vast van overtuigd dat er ooit dit soort systemen mogelijk zouden worden.

'Digitale hoortoestellen' werd tijdelijk een wat vervuilde term omdat toestellen met een analoge audio-pad maar met (digitaal) programmeerbare parameters ook zo genoemd gingen worden (en ietwat denigrerend ook wel 'electronische schroevendraaiers').

De 'tweede digitale golf' werd eind 1992 ingeluid door het Deense Danavox met een niet programmeerbare DSP (zelfs met een totaal hard-wired algoritme) speciaal gemaakt voor (actieve) onderdrukking van rondzingen. De afmeting en het stroomverbruik (40 mm², 2,8 mA) lieten slechts gebruik in grote high power toestellen toe.

Eind 1995 kwam het eveneens Deense Widex met een niet programmeerbare DSP met 3 frequentiebanden, (passieve) rondzingonderdrukking en een zekere vorm van ruisonderdrukking. Een paar maanden later kwam de derde Deense firma, Oticon, met een zeven kanaals niet programmeerbaar systeem.

Onder een niet programmeerbare DSP wordt hier verstaan dat alleen de parameters van het algoritme te veranderen zijn, niet de bewerkingen zelf of de topologie (het blokschema) ervan.

Begin 1997 kwam Philips als eerste met een wel programmeerbare DSP. Met dit systeem kon de eindgebruiker met een simpele druk op de knop van algoritme veranderen, bijvoorbeeld omdat in een andere omgeving een andere processing meer wenselijk is. Dit systeem werd ook ingezet om aan andere bedrijven, op OEM-basis, een DSP hoortoestel-chip te leveren, inclusief algoritme volgens specificatie.

Een commercieel nadeel van dit systeem was dat het permanente geheugen (de EEPROM) vanwege de beperkte ruimte in een afstandsbediening moest worden ondergebracht. Eind 1999 kwam de opvolger van dit systeem op de markt waarbij de EEPROM wel in het hoortoestel was ondergebracht (als tweede IC). Bovendien was de rekenkracht van de DSP verdubbeld, de totale batterijstroom gehalveerd, de complete microprocessor, incl. geheugens, uit de afstandsbediening op de chip gezet, en dat alles terwijl tegelijkertijd het IC-oppervlak drie maal kleiner geworden was. Dit is een duidelijke illustratie van de snelheid van de ontwikkelingen in de huidige IC-technologie en applicaties daarvan.

Momenteel zijn er in de relatief kleine hoortoestellenwereld meer dan 10 (!) verschillende, in eigen beheer ontworpen, DSP systemen op de markt. Elke firma probeert zich met zijn eigen processor en zijn eigen algoritmes te profileren.

Bevatten de eerste toestellen 3 transistoren, moderne digitale hoortoestellen hebben (in knap 30 jaar) de 1 miljoen transistoren reeds overschreden en het einde van deze ontwikkeling is nog lang niet in zicht! Ondanks deze ware explosie van complexiteit vormt de electronica zeker niet het grootste onderdeel van een digitaal hoortoestel. De transducers, de batterij en de behuizing zelf vergen meer mm³ en al is ook hier de miniaturisering ver gevorderd [3], de ontwikkelingen gaan er veel

minder snel. Daardoor is er momenteel geen dwingende reden meer om de electronica nog veel kleiner te maken. De nog steeds toenemende integratie-dichtheid van nieuwe IC processen (wet van Moore, zie fig. 1) zal dus in de hoortoestellen industrie gebruikt worden om steeds betere digitale signaalverwerking en nieuwe features te implementeren. De grote vraag zal verschuiven van 'hoe' naar 'wat' er gemaakt moet worden. De audiologische ontwikkelingen zijn intussen dus minstens zo belangrijk als de elektronische geworden.

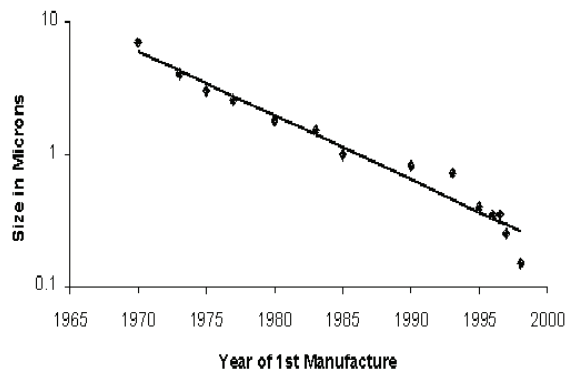


Fig.1 Ontwikkeling in CMOS IC-procestechnologie van de kleinst mogelijke structuur.

Door de introductie van DSP maakt de hoortoestellenindustrie momenteel een uiterst innovatieve fase door. Een gevolg is echter wel dat de IC-ontwikkelkosten momenteel ook exploderen, vooral sinds de intrede in het submicron tijdperk rond 1990. Alleen die bedrijven die vooraan kunnen blijven lopen in toepassing van de modernste IC-technologie zullen in de toekomst steeds de meest geavanceerde audiologische concepten kunnen aanbieden. In de relatief kleine en versnipperde bedrijfstak geeft deze ontwikkeling al enige tijd aanleiding tot een fusie-golf waarbij naar verwachting hooguit 5 grotere spelers zullen overblijven, die deze markt geheel zullen gaan domineren.

De vooruitzichten voor de hoortoestellenmarkt als geheel zijn uitermate gunstig, waarbij de vergrijzing en de algemene economische vooruitgang voorlopig de overheersende trends zijn. Dat de toestellen zelf steeds beter presteren, steeds kleiner worden en dat mede daardoor de stigmatisering van de dragers ervan verder afneemt zal zeker ook helpen.

3. Trends in signaalbewerking in hoortoestellen

Inleiding

Algemeen gesteld is het doel van een hoortoestel om het hoorvermogen van een slechthorende zoveel mogelijk te verbeteren en het horen weer zo aangenaam mogelijk te maken. Hierin kunnen soms tegenstrijdigheden optreden bijvoorbeeld omdat voor het verstaan van spraak een wat 'scherpe', niet maximaal comfortabele klank vaak gunstiger is, of omdat men het horen van zachte geluiden zoals tikkende klokken geheel ontwend is geraakt.

Het ultieme doel: herstel van het 'normale' gehoor, is meestal onbereikbaar, zodat het in de praktijk zal gaan om de beste benadering daarvan. Alleen voor problemen in het middenoor (rond de keten van gehoorbeentjes) is vaak een nagenoeg perfect herstel m.b.v. een hoortoestel mogelijk, dit omdat het middenoor een lineaire karakteristiek heeft. Het overgrote deel van de slechthorenden heeft echter (ook) een probleem met het binnenoor (slakkenhuis, cochlea) waarvan de werking uiterst complex, sterk niet-lineair en nog lang niet geheel begrepen is. In de meeste gevallen is de capaciteit van transmissie en bewerking van informatie in de cochlea, dus van middenoor naar hoorzenuw, blijvend aangetast en waarschijnlijk niet meer volledig met 'pre-processing' in de vorm van een hoortoestel, hoe complex ook, te compenseren. Desalniettemin kan met een goed aan het hoorverlies aangepast (gefit) toestel in de meeste gevallen een grote verbetering in het sociaal functioneren van de slechthorende bereikt worden.

Feit is ook dat met DSP deze 'pre-processing' vele malen complexer en meer flexibel aanpasbaar generaliseerd kan worden dan met analoge technieken ooit mogelijk zou zijn, zodat met DSP de 'restmogelijkheden' van het oor in principe veel beter benut kunnen worden.

Basis-algoritme

Een vaak toegepast basis-algoritme in DSP's is de multiband AGC. Hierbij worden eerst met een filterbank een aantal frequentiebanden gemaakt, die elk hun eigen automatische versterkingsregeling (AGC) en gain-instelling krijgen. Na optelling en eventueel nog volumeregeling en begrenzing van het uitgangssignaal gaan deze signalen naar de speaker (in hoortoestellen de receiver genoemd).

Extra features die wel worden toegepast zijn automatische volumeregeling, rondzingeronderdrukking, multi-microfoon technieken ter verhoging van de richtinggevoeligheid en andere methoden voor onderdrukking van ongewenste signalen. Op deze technieken wordt nu verder kort ingegaan.

Versterking en filtering

Het zal duidelijk zijn dat de eerste functie van een hoortoestel het voldoende versterken van het inkomend geluid is, in de praktijk in de range van 15 tot 80 dB. Maar niet alle frequenties moeten evenveel versterkt worden. In de meeste gevallen is het hoorverlies bij hogere frequenties, zo boven 1 tot 3 kHz, veel groter dan bij lagere frequenties. De meeste hoortoestellen hebben dan ook een sterk tot zeer sterk hoogdoorlaat karakter. In de paragraaf 'ruisonderdrukking, fixed filter' worden hierbij nog enkele overwegingen genoemd.

Alleen bij de zeer zware hoorverliezen gaat het vaak om de lage frequenties, namelijk indien alleen daar nog gehoorresten aanwezig zijn. De toestellen moeten dan net een laagdoorlaat karakter hebben om het oor niet met nutteloze of zelfs schadelijke geluidsenergie te belasten en ook om het risico van rondzingen te verkleinen.

In het algemeen heeft het geen zin om frequenties hoger dan zo'n 8 kHz te willen compenseren. Voor de meeste slechthorenden is het hoorverlies hier al zo groot dat er teveel energie aan het oor toegevoerd zou moeten worden en verder zijn deze frequenties voor de belangrijkste functie van een hoortoestel: het verstaan van spraak, vrij onbelangrijk. Bovendien kunnen de gebruikte receivers deze frequenties niet met voldoende hoog rendement weergeven.

Met meer complexe filters, die digitaal in een kleiner volume en veel flexibeler te realiseren zijn dan analoog, is het toonaudiogram, dat sterk persoonsafhankelijk is en grillig kan verlopen, steeds nauwkeuriger te compenseren. Zelfs het 'low power' gebruik van complexe filterbanken en fast fourier transformatie (FFT) is mogelijk geworden. Bij het opsplitsen van het signaal in meerdere banden kan verdere processing per band, bijvoorbeeld ter onderdrukking van ruis of rondzingen, steeds succesvoller worden, al is er voor elk type processing wel een maximum waarboven er geen verdere verbetering meer is.

Compressie

Het dynamisch bereik voor geluid bedraagt bij een goedhorende zo'n 120 dB en loopt van de gehoor-

drempel (0 dB SPL) tot de pijngrens (120 dB SPL). Bij de meeste slechthorenden is dit bereik veel kleiner geworden en neemt bovendien vaak af bij toenemende frequentie. In extreme gevallen kan het zelfs dalen tot 10 dB! De oorzaak is dat de gehoordrempel flink toegenomen is, meestal oplopend met de frequentie, terwijl de pijngrens veel minder is toegenomen, vaak zelfs constant gebleven of iets is afgenomen. Kenmerkend voor dit verschijnsel, ook recruitment genoemd, is de uitspraak: 'praat eens wat harder, ik versta je niet goed' om vervolgens te zeggen: 'schreeuw toch niet zo, het doet pijn in mijn oren!'.

Als middel hiertegen wordt de automatische versterkingsregeling (automatic gain control, AGC) veelvuldig toegepast. Deze schakeling vermindert de versterking bij toename van hetingangssignaal. Klassiek wordt een constante compressieverhouding (compression ratio, CR) toegepast. Het aantal dB toename in hetingangssignaal dat nodig is om het uitgangssignaal met 1 dB te doen toenemen (dit is de CR) is dan onafhankelijk van de sterkte van hetingangssignaal (de gain neemt met (CR-1) dB af als de ingang met CR dB toeneemt). Bij toepassing van DSP's komt ook een meer individuele aanpassing van de input-output curve in zicht teneinde de verstoorde luidheidsopbouw van de slechthorende nauwkeuriger te herstellen. In dat geval zal de CR in het algemeen dus variëren met de sterkte van hetingangssignaal.

Omdat signaalafhankelijke versterking uiteraard een niet-lineaire operatie is (voor een lineair systeem geldt: $CR = 1$) zou bij momentane gain-aanpassing een grote, zeer storende harmonische vervorming optreden. Om dit te vermijden worden regeltijden (de zogenaamde attack- en release-tijd) ingevoerd: de versterking is niet afhankelijk van de momentane ingangssignaalwaarde maar van een uitmiddellende maat, bijvoorbeeld de RMS, of een kenmerk van de omhullende zoals het gemiddelde van de absolute waarde of de topwaarde. Daardoor worden de gain-veranderingen veel laagfrequenter en wordt de vervorming naar acceptabele waarden teruggedrongen. Signaalveranderingen veel sneller dan deze regeltijden zullen nu ook niet worden gecomprimeerd zodat de fijnstructuur van spraak niet wordt 'platgeslagen'. Pogingen om deze fijnstructuur zelfs nog op te rekken m.b.v. expansie (CR tussen 0 en 1), teneinde het spraakverstaan te verbeteren, hebben nooit positieve resultaten opgeleverd.

AGC toepassingen als dynamische compressie, output begrenzing en automatische volumerege-

ling dienen er in eerste instantie steeds toe om het inkomende geluid binnen het verkleinde dynamisch bereik van de slechthorende te brengen en te houden en wel met minimale verslechtering van de signaalkwaliteit. De attack-tijd wordt hierbij meestal korter dan de release-tijd gekozen om niets van het begin van een lettergreep, woord of een zin te hoeven missen.

De verschillende AGC toepassingen vragen om verschillende parameterinstellingen:

AGC toepassing	CR	Attack-tijd	Release-tijd
dynamische compressie	1.1 tot 4	0 tot 50 ms	30 tot 150 ms
begrenzing van maximum output	8 tot oneindig	0 tot 20 ms	15 tot 80 ms
automatische volumeregeling	2 tot 4	0.5 tot 1 sec	1 tot 10 sec

Ruisonderdrukking

Uit metingen van de verstaanbaarheid van spraak in de aanwezigheid van achtergrondruis blijkt dat slechthorenden een veel betere signaal / ruis verhouding (S/N) nodig hebben dan goedgehoorden om een even groot percentage te verstaan, gemiddeld ongeveer 10 dB om zinnen te kunnen herhalen [4].

Al jarenlang vindt er wereldwijd uitgebreide research plaats op universiteiten en in de industrie naar methoden om spraak van andere signalen, hier gedefinieerd als ruis, te onderscheiden en vervolgens de relatieve versterking van de gevonden spraak te verhogen. Een extra complicatie is het feit dat in veel situaties de spectrale componenten van ruis en spraak weinig van elkaar verschillen, omdat de achtergrond eveneens uit spraak of geroezemoes bestaat.

Omdat er door toepassing van ruisonderdrukkers ook altijd artefacten ontstaan, neemt de gemeten spraakverstaanbaarheid in ruis meestal niet toe, het luistercomfort soms wel (verstaan kost bijvoorbeeld minder moeite). In veel gevallen is in een gecontroleerde laboratorium-omgeving wel een verbetering van de verstaanbaarheid meetbaar, maar in alledaagse situaties valt het effect vaak tegen. Momenteel wordt de enige uitzondering hierop gevormd door systemen die een grote richtinggevoeligheid hebben en dus 2 of meer microfoons hebben.

Er volgt nu een korte bespreking van een aantal methoden om gewenste signalen (meestal spraak) te scheiden van ongewenste.

Fixed filter:

In situaties waarin de frequentiecomponenten van spraak en ruis veel verschillen kan eenvoudigweg vast ingestelde filtering al helpen. In veel gevallen wordt het eerder genoemde hoogdoorlaat karakter van een hoortoestel bewust wat overdreven t.o.v. het toonaudiogram, omdat storende geluiden in de praktijk meestal laagfrequentier zijn dan de frequenties die voor het verstaan het belangrijkste zijn. Denk hierbij aan verkeerslawaaai, geluid in de auto, huishoudelijke apparaten, machine-gedreun, etc. Een extra reden voor deze overdrijving is het feit dat de 'Upwards Spread Of Masking' (USOM, het verschijnsel dat een laagfrequent toon veel meer hoog maskeert dan andersom) bij slechthorenden meestal groter is dan bij goedgehoorden.

Squelch:

Een squelch is een schakeling die de versterking bij hele zachte geluiden vermindert. Dit kan helpen wanneer de slechthorende zachte achtergrondgeluiden zoals tikkende klokken irritant vindt, bijvoorbeeld omdat hij deze geheel ontwend is geraakt. Een squelch helpt ook als de eigenruis van het hoortoestel als storend wordt ervaren. Bij een goed ontworpen toestel ruist de microfoon echter meer dan de electronica en deze wordt zelden door een slechthorende waargenomen (wel door de normaalhorende dealer die het toestel moet verkopen!).

Voice Activity Detection:

Een al wat meer selectieve methode is voorts de Voice Activity Detector (VAD) die spraak probeert op te sporen aan de hand van het meest karakteristieke kenmerk ervan: het extreem grillige verloop in de tijd van de omhullende of de energie. Bij een relatief rustige omhullende laat men dan de VAD de gain terugregelen. Deze methode zal derhalve beter werken bij continue dan bij fluctuerende ruis. Door op deze manier in meerdere frequentiebanden VAD's te gebruiken kan de frequency-shaping adaptief gemaakt worden teneinde de spraak maximaal te bevoordelen. De regeling moet langzaam genoeg zijn om te voorkomen dat er amplitudemodulatie aan de spraak wordt toegevoegd.

Spectral Subtraction:

Bij 'Spectral Subtraction' wordt geprobeerd om het spectrum van de ruis apart te schatten en daarmee de S/N als functie van de frequentie. M.b.v. deze informatie wordt vervolgens een adaptief filter gestuurd dat die delen van het spectrum verzwakt die een slechte S/N hebben.

Bijvoorbeeld kan in een aantal frequentiebanden gedurende een bepaalde tijd naar het minimum van het signaal worden gekeken. Is het gemiddelde van dit minimum relatief hoog ten opzichte van de RMS-waarde dan wordt verondersteld dat het signaal niet puur spraak was, omdat zuivere spraak hele diepe minima t.o.v. de RMS heeft. De gain in een band wordt vervolgens omgekeerd evenredig met het verschil tussen het gevonden minimum en de RMS teruggegeeld.

Deze methode zal uiteraard pas effect hebben als er niet in elke band evenveel teruggegeeld wordt, dus als er verschillen in S/N per band zijn.

Statistische processing:

Statistische processing methoden, proberen eigenschappen te vinden waarin signalen gemiddeld gesproken verschillen. Van echt witte ruis valt niets te voorspellen, maar van spraak bijvoorbeeld heel veel, o.a. met het Linear Predictive Coding (LPC) model van de generatie van spraak. In zo'n geval is het in principe mogelijk de volgende sample van een gewenst signaal uit een reeks van voorgaande samples te schatten. Er zijn technieken ontwikkeld, als Least Mean Squares (LMS), die uit het verschil van deze schatting met de echte volgende sample-waarde een instelling van een adaptief filter kunnen berekenen, zodanig dat er een criterium, zoals de gemiddelde kwadratische fout, geoptimaliseerd wordt. Dit soort methoden vereisen voor ruisonderdrukking echter een rekenkracht die voorlopig niet in hoortoestellen te vinden zal zijn.

Multi-microfoon technieken:

Multi-microfoon technieken scoren het beste in S/N-verbetering. Deze kunnen een grote richtinggevoeligheid voor het inkomende geluid realiseren. Uitgaande van het feit dat in de meeste situaties het gewenste geluid uit de richting komt waarin men kijkt en het ongewenste geluid uit allerlei richtingen, kan met een eenvoudig fixed microfoon-array al enorm veel bereikt worden. Een adaptief array, dat bijvoorbeeld zelfstandig zoekt naar de sterkste bron in de omgeving, kan in een vergadering misschien handig zijn maar heeft

ook een 'eigen leven' dat voor de gebruiker niet altijd te volgen is en daardoor minder geschikt lijkt voor toepassing in hoortoestellen. Een ander soort adaptief array 'kijkt' wel steeds recht vooruit en kan tegelijkertijd de sterkste bron of bronnen in andere richtingen vinden en verzwakken.

Met grotere arrays, bijvoorbeeld op een bril gemonteerd, is een superieure richtinggevoeligheid te bereiken. Bij lage frequenties (beneden zo'n 800 Hz) is dit zelfs de enige manier om een richtinggevoelig systeem met voldoende S/N te realiseren [5,6].

Rondzingonderdrukking

Rondzingen is altijd een potentieel probleem in hoortoestellen. In een uiterst klein volume is immers een complete geluidsinstallatie met een acoustische versterking van soms wel 80 dB ondergebracht. Geluid dat uit de gehoorgang teruglekt naar de microfoonopening is in de praktijk de meest voorkomende oorzaak van rondzingen. Dit lek kan ontstaan door een niet goed passend gemaakt toestel of oorstukje of door de 'venting', de bewust daarin aangebrachte opening om het oor te beluchten.

Bij kauwen, geeuwen, lachen, huilen, etc. verandert de vorm van het oorkanaal enigszins waardoor er tijdelijk een veel groter lek kan ontstaan. Het externe geluidspad wordt verder sterk beïnvloed door bijvoorbeeld een hand of telefoonhoorn aan het hoofd, een hoed of muts etc. Vooral bij high power toestellen is rondzingen slechts te voorkomen met krappe passingen zonder venting (hetgeen het draagcomfort vermindert) of uiteindelijk door verlaging van de gain. Minder gebruikte methoden zijn toepassing van een zak-apparaat of een cross-toestel (microfoon via een draadje aan het andere oor) wat helpt vanwege de langere luchtweg.

Al met al zijn er dus redenen genoeg om de DSP ook in te zetten tegen rondzingen.

De meest simpele methode, ook wel in analoge toestellen toegepast, is het bij de aanpassing van het hoortoestel verlagen van de gain in de band of banden waar de rondzingmarge te klein blijkt. Dit kan echter gemakkelijk een gain-tekort in de voor het spraakverstaan belangrijke frequentiebanden tot gevolg hebben.

Een meer selectieve methode is toepassing van een of meerdere notch-filters die vast zijn afgestemd op de verwachte rondzing-frequenties. Probleem hiermee zijn de bovengenoemde veranderende

omstandigheden van het externe geluidspad zodat de notches vrij breed moeten zijn.

Beter is het deze notches adaptief te maken en te sturen met een schakeling die op een of andere manier rondzingen detecteert, bijvoorbeeld door signalering van periodieke signalen. Met deze methode worden ook extern gegenereerde binnenkomende periodieke signalen verzwakt, wat afhankelijk van de situatie een voordeel of een nadeel kan zijn. In de praktijk zijn er minstens 2 onafhankelijke notches nodig, maar meer is beter.

De beste, maar ook rekenkundig duurste methode is toepassing van een adaptief filter dat in staat is automatisch en met voldoende precisie het externe rondzingpad, inclusief de transducers, na te bootsen en door aftrekking van het microfoonsignaal dit pad te onderdrukken. Om te voorkomen dat hiermee ook het inkomende geluid zou verdwijnen moet de uitgang worden 'gedecorreleerd' t.o.v. de ingang, zodat het feedbacksignaal apart kan worden gevonden (zie blokschema fig. 2), bijvoorbeeld door het filter te sturen op minimalisatie van de gemiddelde kwadratische waarde van het verschilsignaal.

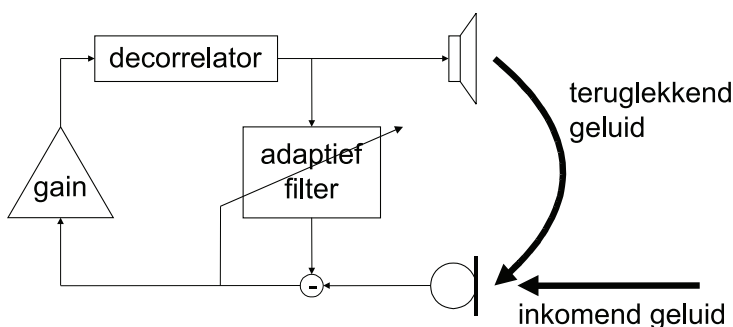


Fig.2 Actieve rondzingonderdrukking m.b.v. adaptieve filtering.

Het decorreleren kan bijvoorbeeld gebeuren d.m.v. tijdsvertraging, kleine frequentieverschuiving (al snel ongeschikt voor muziek), toevoegen van weinig opvallende vervorming of van zachte ruis etc.

Met deze actieve methode ontstaan er in principe geen 'gaten' in de frequentieresponsie zoals bij de andere genoemde methoden wel het geval is.

In alle methoden zijn nadelen aan te wijzen die de geluidskwaliteit van het toestel kunnen verslech-

teren, maar het is natuurlijk al snel beter dan het allesoverheersende rondzingen.

Bij het ontwerp van een hoortoestel moet altijd ook goed gelet worden op het voorkomen van interne mechanische, elektrische en magnetische meekoppelingen. Vooral bij high power toestellen zijn er begrenzungen die meestal slechts in grotere en duurdere toestellen op te lossen zijn. Vanwege de genoemde artefacten zou het getuigen van slecht ontwerp om rondzingonderdrukkers in te zetten tegen intrinsiek instabiele toestellen (misschien m.u.v. de allerswaarste categorie toestellen). Om dezelfde reden mag de aanwezigheid van rondzingonderdrukking nooit een alibi zijn om slechte oraafdrukken of passingen te maken.

4. Philips' tweede generatie dsp-ic voor hoortoestellen

Waarom eigen DSP-IC's ontwikkelen?

Het zal intussen duidelijk geworden zijn dat de allesoverheersende design-parameters van IC's voor hoortoestellen het stroomverbruik en de afmetingen (inclusief de benodigde externe componenten) zijn. De reden waarom de relatief kleine hoortoestelfirma's al jaren, voor hun omvang, enorm kostbare IC-projecten uitvoeren, zelfs inclusief het in eigen beheer ontwikkelen van state of the art AD-converters en DSP-cores, is gelegen in het feit dat 'low power' in deze branche betekent: 0.5 á 1 mW voor een compleet DSP-audio systeem. Dit is dermate laag dat er tot nog toe geen synergie met enige andere branche van betekenis gevonden is. Bijna alle andere draagbare toestellen hebben een veel grotere batterij met ook een hogere spanning aan boord omdat daarmee tevens een grootverbruiker zoals een motor of een GSM-zender gevoed moet worden, waardoor er niet op een paar milliwatt gekeken hoeft te worden. Nieuwe media zoals MP3 spelers bieden hier wellicht openingen voor de toekomst, al zal het waarschijnlijk nog lang zo zijn dat 'ultra low power' en 'general purpose' elkaar praktisch uitsluiten. 'Ultra low power' vereist voorlopig zeker nog gespecialiseerde hardwareblokken voor 'dure' berekeningen.

De benodigde zeer specifieke design kennis (er moeten vrij normale audio-spec's gehaald worden bij een zeer laag energieverbruik) is momenteel dan ook vooral in en rond de hoortoestelfirma's opgebouwd.

Omdat de kleinste afmetingen in IC's nog steeds verder afnemen (zie fig. 1) is de voedingsspanning van 'mainstream' IC's de laatste jaren gestaag richting 1 Volt aan het zakken (vanwege problemen als warmteafvoer en electromigratie ten gevolge van te grote elektrische veldsterkte op de chip). Voordeel hiervan is dat er in steeds bredere kring oplossingen gegenereerd worden voor de specifieke problemen van het ontwerpen bij 1 Volt (en lager) en dat de IC-ontwerp-bibliotheken steeds beter bij deze lage spanningen worden gekarakteriseerd, waardoor het ontwerp beter te simuleren en dus beter te voorspellen is.

Proceskeuze en chipsets

De optimale keuze van het IC-productieproces is van zeer veel factoren afhankelijk: tijd (o.a. tijdige beschikbaarheid van ontwerp-bibliotheken van nieuwe processen), geld, afmetingen, technische eigenschappen, beschikbare proces-opties, bestaande expertise- en vertrouwens-relaties, etc, etc. De verschillende hoortoestelfirma's kiezen dan ook voor heel verschillende oplossingen. Vaak wordt het permanente geheugen (EEPROM) als aparte chip uitgevoerd en sommigen maken ook nog een aparte analoge en digitale chip, of zelfs nog een vierde chip. Vanwege het minimaal aantal externe verbindingen, die ruimte kosten, zou het ideaal zijn om het hele DSP systeem op één chip te zetten (mits die dan niet te groot wordt), maar er zijn nog geen processen beschikbaar waarin alle systeemdelen optimaal gerealiseerd kunnen worden.

Binnen Philips was er gelukkig voldoende kennis aanwezig om ook de analoge delen in CMOS uit te voeren, zodat er geen apart analoog IC (met bipolaire transistoren) nodig was. Vooral de microfoon-voorversterker en de AD-converter zijn hierbij kritisch vanwege de (bij het geëiste, lage stroomverbruik) zeer strenge ruis-eis van 2 μ V input referred (100 Hz tot 8 kHz).

Door zorgvuldig systeem-ontwerp moet bovendien worden voorkomen dat het digitale deel (met de DSP en de geschakelde voedingscircuits) storingen, zoals ruis of toontjes, in het analoge deel kan veroorzaken, hetgeen zowel via de voeding en het substraat als via overspraak op signaalleidingen mogelijk is.

De EEPROM vergt een nogal afwijkend IC-proces waarin teveel concessies voor de analoge en digitale delen gedaan moesten worden. Bovendien zijn de gewenste EEPROM's gewoon op de markt verkrijgbaar, zij het dat de laagste voedingsspanning

momenteel nog 1.8 V is. Deze spanning is echter op de systeem-chip toch al nodig, vooral om in het analoge deel de genoemde ruis-eis te halen.

Systeembeschrijving

Philips' tweede generatie DSP-systeemchip voor hoortoestellen bevat de volgende delen:

- het audio-pad, bestaande uit voorversterkers voor twee microfoons en een luisterspoel, de AD-converter, de DSP met Progam-RAM en State-RAM en een speciale klasse-D uitgangstrap.
- een microcontroller met ROM en RAM
- een I²C interface tussen microcontroller en (externe) EEPROM
- twee klok-oscillatoren met trimregister
- een uitgebreid systeem van voedings- en bias-circuits inclusief bandgap referentie.
- een RS232 interface tussen de microcontroller en de PC die gebruikt wordt tijdens het initialiseren en testen in de fabriek en tijdens het fitten van het toestel bij de dealer
- on-chip EMC condensatoren ter vermindering van hoorbare storing ten gevolge van instraling van mobiele telefoons (GSM, DECT), etc.

[7,8] beschrijven de eerste generatie in detail. Veel aspecten hiervan werden in de tweede generatie overgenomen.

Verder is er een infra rood (IR) ontvanger aanwezig voor optioneel gebruik van een afstandsbediening aanwezig, maar deze zal waarschijnlijk niet commercieel gebruikt gaan worden. Indertijd was een van de redenen deze schakeling op te nemen de angst dat de externe EEPROM niet in de aller-kleinste hoortoestellen zou passen. De afstandsbediening zou dan de rol van permanent geheugen moeten vervullen, zoals in Philips' eerste DSP-generatie standaard was. Het is echter in de loop der tijd gebleken dat een afstandsbediening slechts door ongeveer 5% van de hoortoestel-gebruikers wordt gewaardeerd en dit percentage is zelfs nog lager als deze onmisbaar is bij het opstarten van het hoortoestel.

Audio-pad

De 3 analoge bronnen, 2 microfoons en een luisterspoel (voor het oppikken van magnetische velden van bijv. ringleidingen), worden met programmeerbare gain toegevoerd aan de AD-converter.

Deze is van het type sigma-delta, een zogenaamde éénbits AD-converter en heeft naast de éénbits DA-converter een vierde orde filter in zijn terugkoppellus. De nominale AD-klokkrequentie is 1 MHz, de audio-samplefrequentie (F_s) is 16 kHz ($1 \text{ MHz} = 64F_s$, de AD is dus 64-voudig oversampled). De enorme ruis, die kwantisatie met slechts één bit veroorzaakt, wordt door het lusfilter naar frequenties buiten de band geschoven en in zogenaamde downsample-filters achter de ADC in meerdere stappen weggefilterd, waarbij de samplefrequentie van $64F_s$ naar F_s wordt teruggebracht. Dit ingewikkeld ogend procédé blijkt de meest energie-zuinige oplossing te zijn en bijna alle fabrikanten van digitale hoortoestellen passen het dan ook toe. De kwaliteit van de ADC, in termen van energieverbruik, dynamisch bereik, vervorming en gebruikt IC-oppervlak, blijkt een onderscheidende factor voor de kwaliteit van het hele IC te zijn.

Na downsampling kan de DSP de aangeleverde samples bewerken met een klokkrequentie van 2 MHz. In het algoritme kunnen dus maximaal 128 instructies per audio sample worden uitgevoerd (= $2048/16$). Is het algoritme korter, dan wordt de processor gedurende de resterende tijd gestopt om stroom te besparen. De uitgevoerde instructies zijn van het type VLIW (Very Long Instruction Word) wat betekent dat er vele, in ons geval maximaal ongeveer 10, deel-instructies gelijktijdig, per DSP klokslag van $0.5 \mu\text{s}$, uitgevoerd kunnen worden. Bij deze (parallele) processor is bovendien de hardware qua signaal-flow en energiegebruik geoptimaliseerd voor de meest voorkomende bewerkingen. Zo kosten de berekeningen voor een AGC-detector en voor een adaptor van een wave digital filter (WDF) slechts één enkele DSP-klokslag! Op een standaard general purpose processor zou met name de AGC-detector berekening alleen al vele tientallen klokslagen kosten; bovendien zou daar verder niets parallel mee kunnen lopen.

Dankzij al deze maatregelen is de processor reeds zeer krachtig bij de relatief lage klokkrequentie van 2 MHz. Digitale electronica wordt trager bij lagere voedingsspanning, dus omgekeerd mag de digitale voedingsspanning zakken vanwege de lage klokkrequentie, in ons geval tot zo'n 0.9 V. Precies dat is erg gunstig voor het energieverbruik van de digitale electronica, dat evenredig is met het kwadraat van de voedingsspanning.

In de DSP volgen de signalen een 24 bits breed datapad. Omdat het overall dynamisch bereik van

de chip 77 dB bedraagt lijkt dit misschien overkill. Echter, door de veelheid van bewerkingen in de algoritmes, de gewenste flexibiliteit (we weten nog niet welke algoritmen er in de toekomst nog bedacht gaan worden) en ook door de toepassing van efficiënte teruggekoppelde (IIR) filters, is een tijdelijk hogere precisie noodzakelijk om instabiliteiten te voorkomen en om het gewenste overall dynamisch bereik te halen.

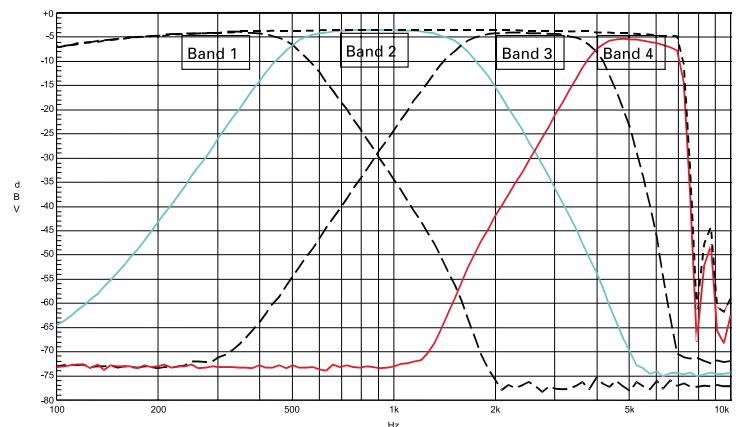
De DSP beschikt over 2 RAM velden: in de program-RAM staat de lijst van uit te voeren VLIW's en de state-RAM wordt gebruikt voor opslag van tussenresultaten.

Enkele basisbewerkingen die in de DSP gedaan kunnen worden zijn: optellen, aftrekken, tijdsvertraging (sample delay), het berekenen van een signaal-omhullende en deze filteren (opgaande en neergaande flanken verschillend steil), vermenigvuldigen met een constante of met een gefilterde omhullende, afronding, saturatie en bit-shift. De genoemde adaptor, een speciaal hardwareblok ten behoeve van audio-filters, bestaat uit enkele optellingen en een vermenigvuldiging. Het blok heeft 2 outputs (van 24 bits) die zoals gezegd in één klokslag van $0.5 \mu\text{s}$ worden berekend.

In fig. 3 is een voorbeeld gegeven van een vijfde orde, vierkanaals filterbank. Deze kost 19 adaptoren en evenveel sample delays. De steilheid van de hellingen is 30 dB/octaaf voor lage frequenties en loopt op naarmate $\frac{1}{2} F_s$ (= 8 kHz) nadert. Dit effect is bewust niet gecompenseerd omdat steilere hellingen goed van pas komen bij de fitting van steile audiogrammen. Belangrijk is de complementariteit van de filterbank: de sommatie van de outputs dient weer vlak te zijn.

Van de DSP-uitgang gaat het signaal, via het upsample-pad, naar een klasse-D brugversterker.

Fig. 3 Vierkanaals WDF-filterbank, swept sine input, breedband meting van de 4 outputs en hun sommatie.



In de eerste DSP-generatie werd hier nog een DA-converter en een analoge klasse-AB brugversterker toegepast. Belangrijkste redenen voor deze aanpassing zijn de grote winst in IC-oppervlak (een goede DAC is groot op deze schaal) en een aanzienlijke energiebesparing bij gemiddeld grote uitsturing van de receiver, vooral bij de wat zwaardere toestellen.

Opgemerkt moet worden dat het gedeelte na de volume en gain regeling (dus het upsample-pad en de klasse-D trap) een dynamisch bereik van meer dan 90 dB heeft. Dit is noodzakelijk omdat anders de input referred noise van het totale systeem snel te hoog wordt bij het indraaien van deze regelaars.

De microcontroller

De on-chip microcontroller heeft vele functies:

- uitvoering van een uitgebreid opstartprotocol bij het inschakelen van de batterij
- interfacing met de EEPROM, met als belangrijkste taak het laden van het program-RAM van de DSP vanuit de EEPROM (er passen 4 algoritmes van maximale lengte in het gebruikte EEPROM)
- interfacing met de PC tijdens initialisatie, test en fitting
- interupt afhandeling van vier programmeerbare lijnen voor diverse optionele user controls (volumeregeling, programma keuze, signaalbron keuze) en het bijhouden van de bijbehorende (programmeerbare) tabellen en standen
- generatie van (zeer flexibel programmeerbare) beep signalen voor user control acknowledge, battery-low indicatie, etc.
- battery-low indicatie
- decodering en uitvoeren van IR commando's
- uitvoering van beveiligd coderen en decoderen (volgens DES) indien een DSP-algoritme m.b.v. een smartcard via de PC geladen of verwijderd wordt (in de eerste DSP-generatie werd deze methode commercieel gebruikt om prijsdifferentiatie van algoritmes mogelijk te maken)

De microcontroller en zijn ROM en RAM beslaan ongeveer een kwart van het digitale deel van de chip. Het stroomverbruik is hoog (het is een ingekochte standaard-core), maar de tijd dat deze gebruikt wordt is zeer kort. Na elke actie gaat de controller weer in 'sleepmode' en verbruikt dan slechts enkele mA.

Klok-oscillatoren

Er zijn twee identieke klok-oscillatoren met trimregister aanwezig. Omdat een kwarts kristal te groot bevonden werd voor toepassing in hoortoestellen, is gekozen voor on-chip oscillatoren die extern getrimd moeten worden (nodig vanwege de grote spreidingen in het IC-productieproces). Dit kan gebeuren tijdens initialisatie in de fabriek, waarbij iteratief de juiste trimstand wordt bepaald, of tijdens ontvangst van een geldige code van een (kwartsgestuurde) afstandsbediening. De tweede oscillator loopt dan mee met de IR-draag golf zonder dat de hoofd-oscillator verloopt (hetgeen hoorbaar zou zijn voor de gebruiker) en alleen indien de code geldig blijkt wordt de gevonden trimstand overgenomen. De gevonden trimstand wordt in het EEPROM opgeslagen en voortaan gebruikt bij het opstarten van het hoortoestel.

Van de hoofdklok wordt alle timing in de chip (programmeerbaar) afgedeeld, zoals de audio-samplefrequentie, de DSP-klok, de microprocessor-klok, etc. Het verloop van de hoofdosculator als functie van batterijspanning en temperatuur moet klein gehouden worden omdat allerlei belangrijke parameters hiervan direct afhankelijk zijn, zoals de audio-bandbreedte van het gehele toestel, de kantelpunten van alle audio filters, de regeltijden van alle AGC's, en de communicatiesnelheid met de PC. Over de gespecificeerde gebieden van 0 tot 50 graden Celsius en tussen 1.0 V en 1.4 V bedraagt het frequentieverloop slechts enkele procenten, wat goed is voor een systeem zonder kristal of andere externe componenten. Om dit te bereiken bij de geldende spannings- en vermogens-eisen is een zorgvuldig ontwerp van de oscillator en diens voedings- en bias-circuits noodzakelijk.

Voedings- en bias-circuits

In een eerste spannings-conversieslag wordt m.b.v. een spoel en een condensator, uit de batterijspanning de 'analoge spanning' (V_{ana}) gemaakt. Deze is programmeerbaar en wordt meestal op 1.8 V ingesteld. Daarna wordt met hoog rendement de halve spanning hiervan gemaakt voor het digitale deel (V_{dig}) m.b.v. een halveerschakeling met 2 condensatoren. Deze halveerschakeling heeft bovendien een dempende werking op de overspraak naar V_{ana} van de vervuiling op V_{dig} , hetgeen nodig is voor voldoende lage ruis in het audio-pad. V_{ana} wordt in een regellus (met een bandgap-circuit als absolute spanningsreferentie) op peil gehouden, waarbij de batterijspanning

(Vbat) mag variëren tussen zo'n 0.7 V en 1,7 V. De lage waarde van Vbat kan optreden als er plotse harde geluiden verwerkt worden (bijv. bij een dichtslaande deur) en de batterij-klemspanning dan scherp daalt tengevolge van de inwendige weerstand van de batterij (die tijdens diens levensduur nog flink oploopt ook). De bovengrens van Vbat werd gesteld op 1.7 V om ook zilveroxide batterijen, met een nominale spanning van 1.6 V, te kunnen gebruiken, die onder andere in Japan veel gebruikt worden. Om Vana zo schoon mogelijk te houden wordt de eindtrap voor de speaker direct op de batterij aangesloten zodat diens relatief grote stroompieken het regelsysteem niet onnodig belasten.

In het algemeen zijn er bij lage voedingsspanning minder mogelijkheden om de Power Supply Rejection Ratio (PSRR) van schakelingen een hoge waarden te geven. Daarom gelden bij de voeding en de diverse instelstromen, voor o.a. de ADC en de oscillatoren en voor de electreet-microfoon(s) (Vmic), vaak zware eisen qua niveau en spectrum van ruis en rimpel. Zo mag Vmic slechts 5 mV ruis in de audioband bevatten en zal teveel storing op de voeding van de oscillator frequentie-jitter geven, hetgeen weer ruis in het audiopad kan veroorzaken.

In alle voedings- en bias-schakelingen moet voorts ook veel aandacht aan temperatuur-stabiliteit besteed worden.

Een trend in low power design is dan ook dat dit soort ondersteunende circuits steeds meer design effort vragen omdat ze langzaam maar zeker een groeiend deel van het totale vermogen en het IC-oppervlak opeisen.

Batterijen

In hoortoestellen wordt omwille van de beperkte ruimte altijd gebruikt gemaakt van slechts één enkele knoopcel-batterij. Tegenwoordig is dat meestal een zink-lucht (ZnO) cel, met een nominale spanning van 1.3 V. Deze cel bevat tweemaal zoveel capaciteit (mAh) per volume als de enkele jaren geleden nog veelvuldig toegepaste kwikcel, die inmiddels vanwege milieu overwegingen in veel landen verboden is. Nadelen van ZnO zijn wel dat de inwendige weerstand groter is (meer kans op instabiel gedrag van het IC, wat zwaardere eisen aan de voedingscircuits stelt en een lagere maximale uitgangsgeluidsdruk) en het feit dat de batterij, nadat de luchtdichte bewaar-sticker is losgenomen, ook zonder belasting na enkele weken leeggelopen is.

Voor een hoortoestel wordt een normgebruik van 16 uur per dag, 7 dagen per week aangehouden. In de markt wordt het gewaardeerd als de batterij minstens een week (dus ruim 110 uur) meegaat, zodat de gebruiker op een vast tijdstip in de week zijn batterij kan vervangen. De kleinste gangbare batterijen hebben een capaciteit van ongeveer 65 mAh, waaruit een maximaal gewenst stroomverbruik bij de kleinste hoortoestellen volgt van 580 mA. Nog niet veel digitale toestellen zitten op deze lage waarde. Omdat in het Philips-IC het algoritme programmeerbaar is, kan ervoor gekozen worden een korter, eenvoudiger DSP-algoritme in de kleinste toestellen te laden. Indien bij een klant een complexer algoritme noodzakelijk is, maar hij toch dat mooie, kleine toestel preferereert kan alsnog het compromis met het stroomverbruik wat verschoven worden.

Enkele karakteristieke eigenschappen van Philips' tweede generatie DSP-ASIC voor hoortoestellen:

Batterijspanningsrange (operation)	0.7 tot 1.7 V
Batterij-ruststroomverbruik (afhankelijk van DSP-algoritme)	0.75 tot 1 mA
Chip-oppervlak (excl. EEPROM van 5 mm ²)	12 mm ²
Aantal aansluitingen	53
Audio:	
Input referred noise (100 Hz – 8 kHz)	2 µVrms
Maximale ingangsspanning (THD < 1 %)	14 mVrms
(resultierend overall dyn. bereik)	77 dB
Maximale elektrische gain	60 dB
THD	< -45 dB
Dynamisch bereik uitgangstrap (incl. upsample-pad)	> 90 dB
Maximale uitgangsspanning	2.5 Vpp
Maximale uitgangsstroom (gespecificeerd, meer blijkt mogelijk)	6 mA _p

Conditities: Vbat = 1.3 V, Rload = 200 W

SAMENVATTING

Veel fabrikanten van hoortoestellen zijn in de jaren 90 digitale signaalverwerking (DSP) gaan toepassen in hun producten. Pas sinds de intrede van de IC-technologie in het sub-micron tijdperk is dit, qua batterij-stroomverbruik en IC-oppervlak, mogelijk geworden. Doelstellingen hierbij zijn het beter helpen van de slechthorende m.b.v. meer features en vooral door toepassing van meer complexe en ook nauwkeuriger individueel aan te passen signaal processing, bijvoorbeeld ter

onderdrukking van achtergrondruis en rondzingen. Door de toepassing van DSP is de ontwikkeling en het complexiteits-niveau van hoortoestellen in korte tijd in een ware stroomversnelling geraakt.

De zeer hoge en nog steeds stijgende kosten die deze ontwikkeling met zich meebrengt, heeft laatste jaren een fusiegolf onder de fabrikanten doen ontstaan.

REFERENTIES

- [1] The Hearing Aid, its Operation and Development, K. W. Berger, ISBN 0-911472-02-9, derde druk 1984.
- [2] De transistor en zijn toepassing, W. van Bussel, Prisma 1291, ISBN 90-274-0311-2, pag. 12, derde druk 1972.
- [3] Bedrijfsbezoek Philips Hoorapparaten op 3 september 1999, ir. L. H. Geerdink en ing. P. Termeer, NERG-Tijdschrift, jaargang 64 nr. 4, pag. 129-134, 1999.
- [4] SRT for sentences as a function of age and noise level. R. Plomp, A.M. Mimpen, JASA 66(5), pag.1333-1342, nov 1979.
- [5] Improvement of Speech Intelligibility in Noise, W. Soede, Proefschrift TU-Delft 1990, ISBN 90-9003763-2.
- [6] Binaural application of microphone arrays for improved speech intelligibility in a noisy environment, I. Merks, Proefschrift TU-Delft jan. 2000, ISBN 90-9013417-4.
- [7] A Single Battery, 0.9V-Operated Digital Sound Processing IC Including AD/DA and IR Receiver with 2mW Power Consumption, H. Neuteboom et al., Paper TP 6.4, Proceedings ISSCC, San Francisco, Febr. 6, 1997.
- [8] A DSP-Based Hearing Instrument IC H. Neuteboom et al., IEEE, JSSC, Vol. 32, No. 11, pag. 1790-1806, Nov 1997.

CV

Jos R.G.M. Leenen werd geboren in 1959 te Weert en studeerde Electrotechniek aan de Technische Universiteit Eindhoven. Na voltooiing van deze studie trad hij in 1985 als wetenschappelijk medewerker in dienst van Philips Research, waar hij o.a. werkte aan datareductie en (compatibele) decodering van High Definition Television (HDTV) signalen.

In 1989 stapte hij over naar de afdeling Pre-development van Philips Hearing Instruments in Eindhoven (eind 1999 gefuseerd met Beltone Electronics Corporation, Chicago tot Beltone Netherlands b.v.). Hier werkte hij aanvankelijk aan richtinggevoelige microfoon-systemen en later aan IC's met een analogo signaalpad, digitaal programmeerbare parameters en afstandsbediening.

In 1993 werd hij projectleider van Philips' eerste generatie DSP-ASIC voor hoortoestellen die in 1997 op de markt kwam. Tot begin 1999 was hij eveneens projectleider van de tweede generatie, die eind 1999 op de markt kwam; daarna werd hij benoemd tot groepsleider 'DSP and IC Development' binnen R&D. Deze groep werkt momenteel, samen met diverse groepen van Philips, aan de derde generatie DSP-ASIC voor hoortoestellen.



Verliesvrij coderen van audiosignalen

Fons Bruekers
Philips Research
Prof. Holstlaan 4
5656 AA Eindhoven
Email: fons.bruekers@philips.com



Samenvatting

Met gebruikmaking van enkele basistechnieken zoals predictie en entropiecodering is het mogelijk om audiosignalen efficiënt te coderen, zodanig dat na decoderen het oorspronkelijke signaal weer exact teruggekregen wordt. Dit zogenaamde verliesvrij coderen staat tegenover het niet-verliesvrij coderen waarbij kleine vervormingen of verliezen zijn toegestaan. Voor twee soorten audiosignaal-representaties, multi-bit PCM en 1-bit DSD, worden in dit artikel de principes van verliesvrij coderen behandeld.

Inleiding

Dat een bericht naast informatie vaak ook nog redundantie bevat is een inzicht dat al heel lang bestaat. In de vorm van het spreekwoord “Een goed verstaander heeft maar een half woord nodig” zien we al heel vroeg een melding van dit inzicht. Anderhalve eeuw geleden heeft Samuel Morse bij de definitie van het naar hem vernoemde alfabet voor het versturen van teksten in wezen al gebruik gemaakt van het begrip redundantie. Per letter is een code toegekend waarvan de lengte afhangt van de frequentie van optreden in de taal. Hoe vaker een letter voorkomt hoe korter het codewoord en omgekeerd. Op deze manier is het voor de marconist mogelijk om met minder tikken op de seinsleutel een bericht te versturen. Rond het midden van de vorige eeuw heeft uiteindelijk Claude Shannon de basis gelegd voor wat wordt aangeduid als de “informatietheorie”. Begrippen als informatie, redundantie, opslag en communicatie heeft hij toen zodanig vorm gegeven dat de informatietheorie een nieuwe tak van de wiskunde is kunnen worden. Veel zaken die we tegenwoordig zo gewoon vinden zijn ontworpen met gebruikmaking van deze theorie, denk hierbij bijvoorbeeld aan de CD-speler.

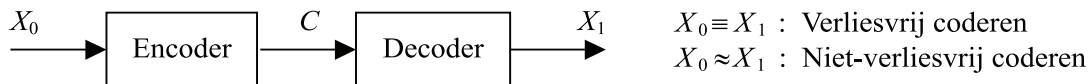
Als in dit artikel gesproken wordt over “coderen” dan wordt bedoeld dat een bericht zodanig wordt bewerkt dat dit efficiënt opgeslagen of getransporteerd kan worden. We hebben het dus niet over coderen om berichten onleesbaar te maken voor derden (cryptografie) of om berichten robuust te maken tegen verminking (foutcorrectie).

Gebruikmakend van de informatietheorie zijn er talloze codeeralgoritmen ontwikkeld die later meer of minder populair zijn geworden. In dit artikel wordt ingegaan op codeeralgoritmen voor toepassing op audiosignalen waarbij na het decoderen het oorspronkelijke signaal exact wordt teruggekregen. Deze zogenaamde verliesvrije codering staat tegenover de niet-verliesvrije codering (in het Engels respectievelijk: “Lossless coding” en “Lossy coding”) waarbij feitelijk ook een deel van de informatie in het signaal wordt verwijderd. De reden dat we ingaan op verliesvrije codering van audiosignalen is dat momenteel een grote ontwikkeling plaatsvindt van deze algoritmen voor nieuwe producten die we straks weer heel vanzelfsprekend zullen gaan vinden.

Coderen

Het codeerproces (figuur 1) wordt in de regel opgesplitst in het encodeerproces dat uitgaande van het oorspronkelijke bericht of signaal, X_0 , een code maakt, C . Vervolgens wordt in het decodeerproces uitgaande van de code, C , een nieuw bericht of signaal gemaakt, X_1 . Indien geldt dat $X_0 \equiv X_1$ dan noemen we het codeerproces verliesvrij en indien $X_0 \approx X_1$ noemen we het codeerproces niet-verliesvrij.

Niet-verliesvrij coderen kennen we van systemen als bijvoorbeeld MPEG, JPEG, PASC en ATRAC.



Figuur 1: Het codeerproces

Die elementen van een video- of audiosignaal die voor de mens niet waarneembaar zijn (irrelevant) worden bij het encoderen verwijderd en vervolgens niet opgeslagen of verstuurd [1]. Dat op deze wijze het signaal efficiënter gerepresenteerd wordt is eenvoudig in te zien, maar het ontwerpen van een dergelijk algoritme vraagt veel kennis van perceptie. Het toepassen van niet-verliesvrij coderen op bijvoorbeeld een stukje tekst kan natuurlijk wel. Bij het lezen van:

“Allen infrmati di vn belang is mot verstuur wrdn.”

zal na even puzzelen duidelijk zijn dat hier bedoeld is:

“Alleen informatie die van belang is moet verstuurd worden.”

Met name als een tekst op de letter geïnterpreteerd moet worden, denk aan een computerprogramma of een salarisbriefje, is het onacceptabel als de interpretatie niet eenduidig is. Daarom kennen we voor dergelijke toepassingen al lang verliesvrije codeerprocessen als bijvoorbeeld: WINZIP, PACK en COMPRESS. Hoe het mogelijk is een efficiënte representatie te maken zonder daarbij iets van de informatie te verwijderen, dus uitsluitend door het verwijderen van redundantie, zal verderop nog uitgebreid aan bod komen. Omdat bij niet-verliesvrij coderen méér uit het signaal verwijderd wordt dan bij verliesvrij coderen, is de mogelijke codeerwinst voor niet-verliesvrij coderen vanzelfsprekend groter. In principe kunnen de eerder genoemde verliesvrije codeerprocessen voor tekst en data, ook toegepast worden voor audiosignalen. Universele verliesvrije codeerprocessen adapteren

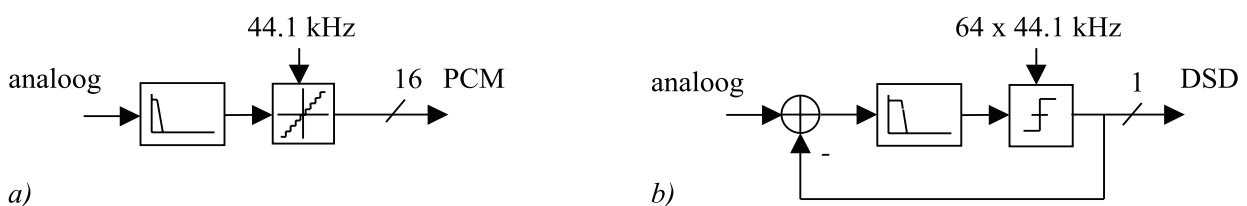
aan de karakteristieken van het aangeboden bericht of signaal. Echter een codeerproces dat speciaal ontworpen is voor een bepaald type bericht of signaal is in vergelijking met een universeel codeerproces minder complex of behaalt voor een vergelijkbare complexiteit een hogere codeerwinst.

Audiosignaalrepresentaties

Tot nu toe is niet expliciet vermeld dat we bij het verliesvrij coderen van audiosignalen er van uitgaan dat deze digitaal zijn. Zeker wanneer bewerkingen exact ongedaan gemaakt moeten kunnen worden, is het wezenlijk dat de signalen digitaal zijn. In geval van analoge signaalbewerking is het onvermijdelijk dat ruis wordt toegevoegd waardoor nimmermeer het oorspronkelijke signaal exact teruggewonnen kan worden. Aangezien de audiobronnen doorgaans een analogo karakter hebben, zullen deze voordat ze verliesvrij gecodeerd worden, omgezet worden naar digitaal met een zogenaamde analogo/digitaal-omzetter. Het is vervolgens dit digitale signaal dat verliesvrij gecodeerd wordt en het is de kwaliteit van dit digitale signaal die behouden blijft. Momenteel zijn twee representaties voor digitale audiosignalen van belang: Pulse Code Modulation (PCM) en Direct Stream Digital (DSD), figuur 2.

Bij PCM wordt op regelmatige tijdstippen de momentane waarde van het signaal (monster) weergegeven met een bepaald aantal bits. In het geval van de CD is de bemonsteringsfrequentie 44.100 Hz en het aantal bits per monster 16. Voor het DVD-audio systeem is gekozen voor PCM met de mogelijkheid van een hogere bemonsteringsfrequentie en een groter aantal bits per monster dan bij CD.

Figuur 2: Bij analogo naar digitaal conversie voor PCM (a) wordt na filtering het signaal bemonsterd en gekwantiseerd. Bij conversie naar DSD (b) wordt na filtering slechts het tekenbit gegeven waarna dit wordt teruggekoppeld naar de analoge ingang

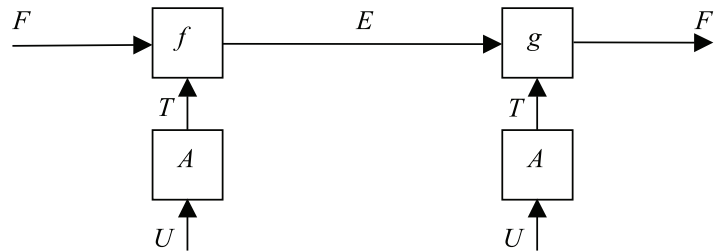


Bij DSD, ook wel bekend onder de naam "bit-stream", wordt veel vaker dan bij PCM een monster uitgegeven dat uit precies 1 bit bestaat. Doordat hierbij gebruik wordt gemaakt van een zogenaamde sigma-delta modulator, waarin de waarde van het uitgegeven monster wordt teruggekoppeld, kan een digitaal signaal van zeer hoge kwaliteit gemaakt worden. Een "1" vertegenwoordigt de waarde +1 en een "0" vertegenwoordigt de waarde -1. Sigma-delta modulatoren worden al geruime tijd toegepast in hoogkwaliteit (PCM) digitaal/analoog-omzetters [2]. Voor het Super Audio CD systeem is gekozen voor DSD met een bemonsteringsfrequentie die 64 maal hoger is dan die van CD.

Voor beide signaalrepresentaties is het mogelijk om met behulp van een verliesvrij codeer algoritme een efficiënte opslag op een disk te realiseren. Beide algoritmes zullen verderop nog aan de orde komen.

Verliesvrije structuren

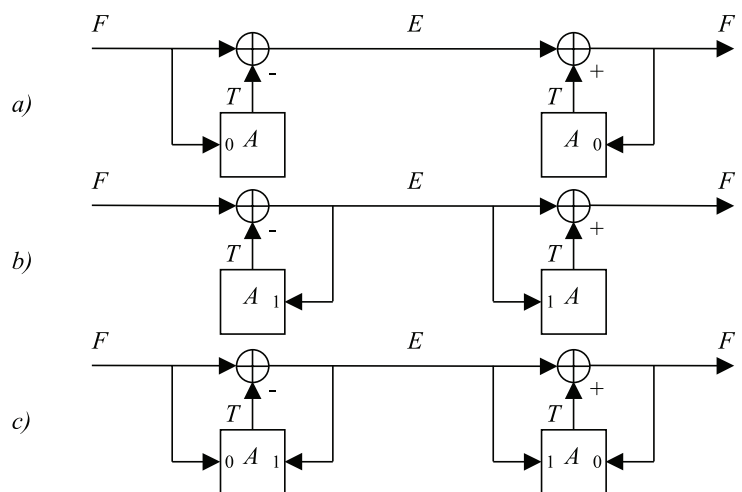
In een verliesvrije coder treffen we allerlei bewerkingen aan die zorgen voor codeerwinst. De belangrijkste eis die gesteld wordt aan deze bewerkingen in de encoder is dat ze in de decoder exact ongedaan gemaakt moeten kunnen worden, dus verliesvrij zijn. Hierbij moeten we vooral letten op kwantisatie die gebruikt wordt om het aantal bits voor het weergeven van resultaten van een berekening in de hand te houden. Bij het kwantiseren wordt in zijn algemeenheid informatie weggegooid maar als het kwantiseren op een speciale wijze gebeurt kan deze bewerking exact ongedaan gemaakt worden. Om dit toe te lichten zien we in figuur 3 het mogelijke uitgangspunt van een verliesvrije structuur. Voor een willekeurig ingangssignaal F en hulpsignaal T zijn de operaties f en g elkaars inverse: $g(f(F, T), T) = F$. Voor f en g kunnen bijvoorbeeld respectievelijk een aftrekker en een opteller of een deler en een vermenigvuldiger gekozen worden. Het is hierbij belangrijk dat bij de realisatie van de functies f en g geen informatie verloren gaat. Voor het geval dat de signalen F en T slechts 1 bit zijn, bijvoorbeeld DSD, kan voor f en g elk een "exclusive-or" gekozen worden. Blok A vertegenwoordigt een willekeurige functie (dus ook een kwantisator) die van signaal U het signaal T maakt. Op basis van deze eenvoudige structuur kunnen zelfs zeer complexe functies volledig verliesvrij geïmplementeerd worden [3]. Indien de signalen U en de blokken A in de encoder en de



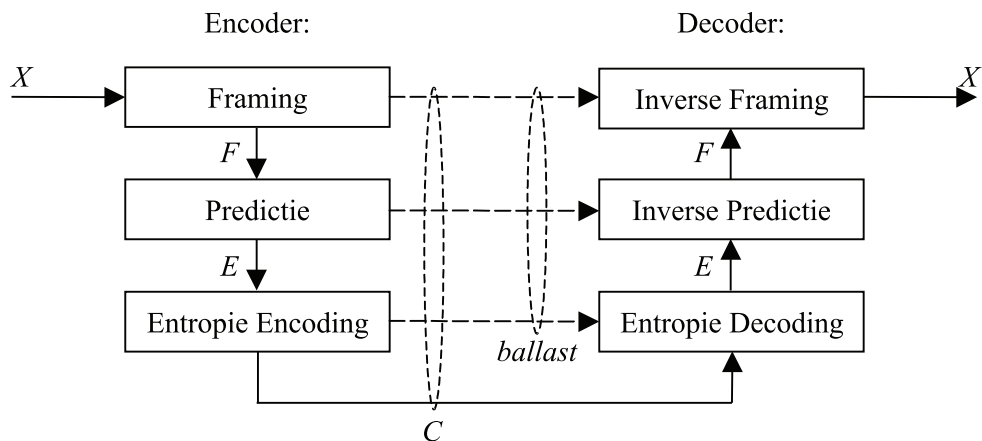
Figuur 3: Indien geldt: $g(f(F, T), T) = F$, dan is deze structuur verliesvrij voor elk willekeurig blok A en signalen F en U

decoder identiek zijn, is de totale structuur verliesvrij.

Uitgaande van het schema uit figuur 3 wordt in figuur 4 de verliesvrije structuur verder gespecificeerd. Voor f en g worden nu respectievelijk een aftrekker en een opteller gebruikt, maar een "exclusive-or" voor elk kan ook. In figuur 4a is voor U het signaal F gekozen. Dit mag op voorwaarde dat blok A voor het bepalen van de huidige waarde van T uitsluitend gebruik maakt van vorige waarden van F . Dit is de belangrijke eis van causaliteit. In tegenstelling tot de Baron von Münchhausen die zichzelf aan zijn haren omhoog kon trekken, is niet mogelijk om voor het berekenen van de huidige waarde deze waarde zelf al te gebruiken. Behalve F kan ook E , of kunnen beiden als invoer dienen voor blok A (figuur 4b en 4c).



Figuur 4: De functies f en g zijn hier vervangen door respectievelijk een aftrekker en een opteller. Verder zijn drie mogelijkheden gegeven om blok A van een ingangssignaal te voorzien. De twee ingangen van blok A zijn aangeduid met 0 en 1



Figuur 5: Het codeerproces verdeeld in de stappen: framing, predictie en entropiecodering

Codeerschema

Ofschoon er duidelijke verschillen zijn tussen de codeerschema's voor PCM en DSD signalen zijn er ook belangrijke overeenkomsten. Een belangrijke overeenkomst is de opdeling van het schema in drie stappen, te weten: framing, predictie en entropiecodering (figuur 5). Deze stappen worden hierna verder toegelicht.

Framing

Met framing van een signaal wordt bedoeld dat een signaal wordt opgedeeld in frames (stukjes van het signaal) die daarna afzonderlijk geëncodeerd worden. In de decoder worden de frames, na decoderen, weer verliesvrij samengevoegd tot het oorspronkelijke signaal. Er zijn drie redenen om framing toe te passen: toegankelijkheid, robuustheid en stationariteit. Overigens gelden deze voor zowel codeersystemen voor PCM als DSD signalen en voor zowel verliesvrije- als niet-verliesvrije codeersystemen.

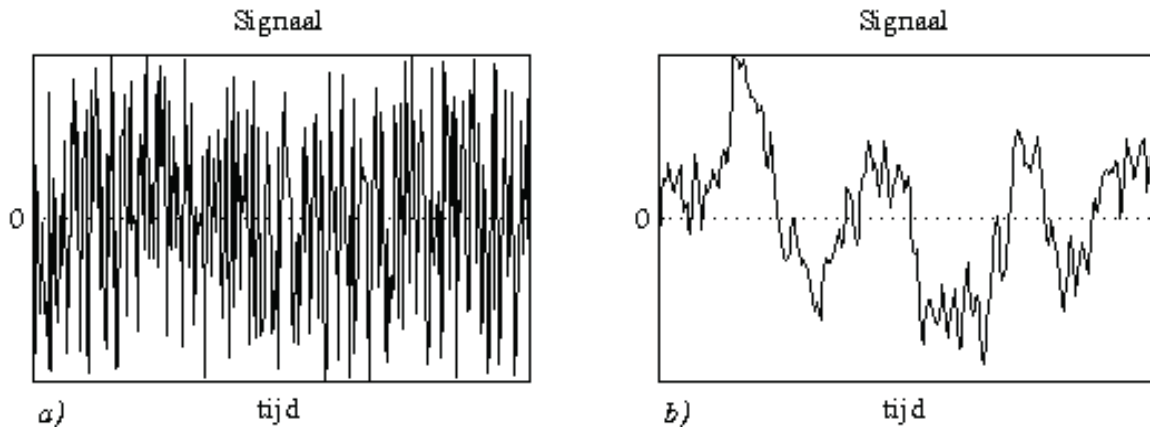
Toegankelijkheid: als van een stuk geëncodeerde muziek slechts het laatste deel beluisterd gaat worden, dan is het ongewenst indien vooraan met decoderen begonnen moet worden, omdat het decoderen doorgaans net zo lang duurt als de muziek zelf. Indien framing is toegepast, volstaat het om een frame op te zoeken nabij het begin van het gewenste fragment en daar met decoderen te beginnen.

Robuustheid: een goede verliesvrije encoder laat nagenoeg geen redundantie achter in de code, waardoor in geval van een vermindering van de code dit niet ontdekt en zeker niet hersteld kan worden. De decoder kan van een verminkte code geen cor-

rect signaal maken. Als de encoder elk frame afzonderlijk heeft geëncodeerd kan de decoder met de komst van een nieuw frame na een vermindering een correct signaal afgeven. De vermindering kan namelijk niet tot het einde van het stuk muziek doorwerken (foutpropagatie). Overigens wordt in de moderne audiosystemen gebruik gemaakt van krachtige foutcorrigerende codes waardoor een zeer betrouwbare opslag of communicatie wordt verkregen.

Stationariteit: zoals in de sectie over entropiecodering verderop aangegeven wordt, is de statistiek van het te coderen materiaal van belang en als deze in de tijd varieert, is het nodig om het codeerproces hieraan aan te passen. Door korte frames te gebruiken is binnen een frame de statistiek constant, we spreken dan van stationariteit, en verhoogt dit de codeerwinst.

Omdat in elk frame opnieuw begonnen kan worden met decoderen, treffen we in elk frame extra gegevens aan voor de decoder. Deze ballast, die de codeerwinst drukt, wordt naar verhouding méér naarmate een frame korter wordt. Het kiezen van de optimale framelengte voor een systeem is in principe een compromis tussen enerzijds toegankelijkheid, robuustheid en stationariteit en anderzijds de ballast. Nu blijkt dat voor heel veel muziek stationariteit en ballast elkaar over een groot gebied ongeveer in evenwicht houden, waardoor de keuze van de framelengte wat dit betreft niet zo kritisch is. In audiocodeersystemen treffen we dan ook framelengtes aan variërend van enkele milliseconden tot enkele tientallen milliseconden.



Figuur 6: Voorbeeld van een signaal gegenereerd met een toevalsgenerator (a) en een "normaal" audiosignaal (b)

Predictie

Als met behulp van een toevalsgenerator een signaal gegenereerd wordt, dan ziet dat er echt anders uit dan een "normaal" audiosignaal (figuur 6). De waarde van een nieuw monster uit een toevalsgenerator heeft niets te maken met de waarden van vorige monsters, terwijl bij een audiosignaal het op basis van de waarden van vorige monsters heel goed mogelijk is om de waarde van een nieuw monster te schatten. Denk bijvoorbeeld aan een zuivere sinustoon waarbij we alle volgende monsterwaarden zelfs exact kunnen bepalen, als enkele vorige waarden bekend zijn. Omdat audiosignalen niet slechts uit sinustonen bestaan zal een schatting altijd een restfout of residu laten. Het schatten van een signaal, doorgaans predictie genoemd, is een reeds lang bestaande techniek [4]. Aan de hand van PCM signalen zullen we de predictie nader toelichten. Voor DSD signalen is in beginsel dezelfde techniek toepasbaar.

Een bruikbare predictie, $T[n]$, van de waarde van het n^{de} monster in een frame, $F[n]$, is een gewogen som van enkele vorige monsterwaarden, formeel is dit:

$$T[n] = \sum_{i=1}^N a_i F[n-i].$$

De weegfactoren a_i , ook wel de predictiecoëfficiënten genaamd, worden elk frame opnieuw met bijvoorbeeld de autocorrelatiemethode bepaald. Hierbij wordt de energie van het residusignaal, $E[n] = F[n] - T[n]$ geminimaliseerd. Het aantal vorige monsterwaarden van F dat gebruikt wordt voor de bepaling van T wordt de predictieorde, N , genoemd. Naarmate de predictieorde toeneemt

wordt de predictie beter en daarmee de predictiefout kleiner. Dat een kleinere predictiefout van belang is voor een hoge codeerwinst is eenvoudig in te zien. Veronderstel dat met een bepaalde serie predictiecoëfficiënten een signaal exact voorspelbaar is, dus met een residu identiek aan 0. In dat geval is het voldoende om in de decoder dezelfde serie predictiecoëfficiënten te hebben en daarmee het oorspronkelijke signaal te reconstrueren. Is het residu niet exact 0 dan moet dit eveneens overgedragen worden aan de decoder, zodat deze zijn predictie kan corrigeren met behulp van dit residu. Naarmate er minder afhankelijkheden bestaan tussen opeenvolgende monsters in het signaal, voegt een predictor minder toe aan het codeerschema. De mogelijke codeerwinst is dan ook sterk afhankelijk van de systematiek in het signaal.

Wordt de lineaire predictiefunctie ondergebracht in blok A van figuur 4a dan kunnen we met behulp van systeemfuncties eenvoudig laten zien dat de totale structuur verliesvrij is. De respectievelijke systeemfuncties in de encoder en decoder zijn namelijk:

$$1 - z^{-1}A(z) \quad \text{en} \quad (1 - z^{-1}A(z))^{-1}.$$

Entropiecodering

Om te beginnen zullen we aan de hand van een voorbeeld demonstreren dat de aanwezigheid van redundantie gebruikt kan worden om een bericht efficiënter te representeren. Er worden dan in principe precies genoeg bits gebruikt om alleen de informatieinhoud of entropie te coderen. Veronderstel dat alle berichten die we willen

beschouwen zijn samengesteld uit 4 symbolen, te weten: "a", "b", "c" en "d". Het is voor de hand liggend om elk van de 4 symbolen "rechttoe rechtaan" met een 2-bit codewoord te representeren, bijvoorbeeld: "00", "01", "10" en "11" respectievelijk. De gemiddelde codewoordlengte is hier dus 2 bit. Als echter bekend is dat de 4 symbolen ongelijke kansen van voorkomen hebben, worden, net als in het Morse-alfabet, korte codewoorden gebruikt voor de representatie van de meest voorkomende symbolen en omgekeerd. Zijn de kansen van optreden van de symbolen respectievelijk: 0,5, 0,25, 0,125 en 0,125 dan is een mogelijke toewijzing van de respectievelijke codewoorden bijvoorbeeld "1", "01", "001" en "000". De gemiddelde codewoordlengte is nu:

$$0,5 \times 1 + 0,25 \times 2 + 0,125 \times 3 + 0,125 \times 3 = 1,75 \text{ bit.}$$

Voor een lang bericht wordt op deze manier een codeerwinst behaald van $\eta = 2 / 1,75 \approx 1,14$. Zodra de codewoorden verschillende lengten kunnen hebben moet er op de een of andere manier voor gezorgd worden dat een aaneenschakeling van codewoorden decodeerbaar is. In het voorbeeld is dit voorzien door de codewoorden "pre-fix free" te maken. Dit wil zeggen dat geen enkel codewoord voorkomt als het beginstuk van een langer codewoord.

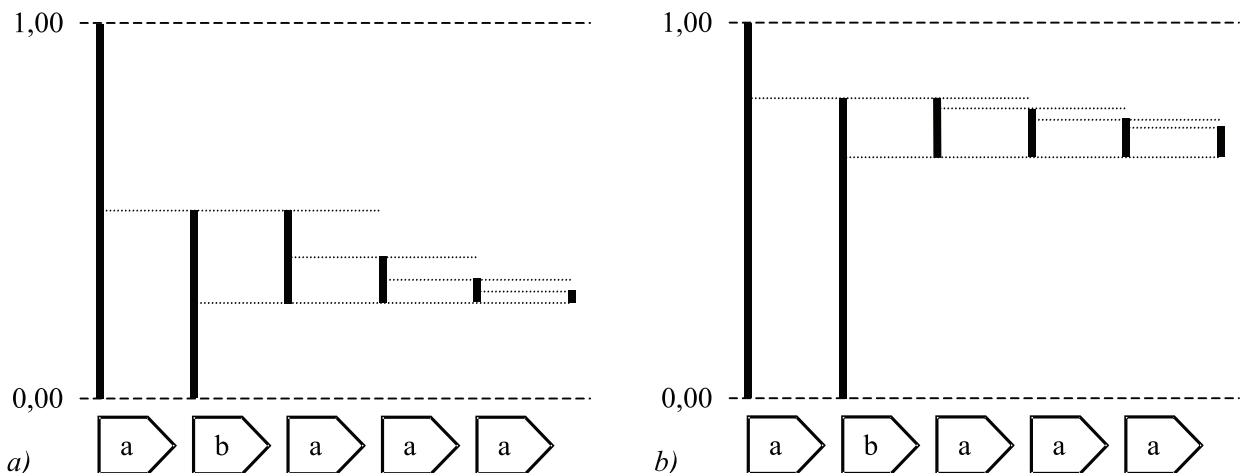
Shannon heeft samen met Fano een methode ontwikkeld om voor een gegeven kansverdeling van de voorkomende symbolen een verzameling codewoorden te ontwerpen (Shannon-Fano code). Later (1952) heeft Huffman een alternatief bedacht dat, doordat het efficiënter is, uiteindelijk veel populairder is geworden [5]. Voor het geval de voorko-

mende kansen allemaal machten van 2 zijn, zoals in het eerdere voorbeeld, is de Huffman-code optimaal. Voor een symbool met de kans p van optreden, is de theoretisch optimale codewoordlengte: $-\log_2(p)$, dit is dus niet perse een geheel aantal bits.

Bij een bepaalde kansverdeling behoort een verzameling codewoorden. Lopende het encodeerproces hoeft een kansverdeling en daarmee ook de optimale verzameling codewoorden, niet constant te zijn. Het is voorstelbaar dat de kansverdeling van de symbolen aan het begin van een lang bericht afwijkt van de kansverdeling aan het eind van het bericht. Wordt de verzameling codewoorden bijgesteld aan de hand van de veranderende statistiek, dan kan daarmee de codeerwinst vergroot worden. Ook op de korte termijn kan de kansverdeling sterk veranderen. Als in een Nederlandse tekst een woord met de letter "q" begint dan is de tweede letter van dat woord altijd een "u" en hoeft dus in principe niet overgestuurd te worden. Door de kansverdeling en daarmee de verzameling codewoorden afhankelijk te laten zijn van de context kan de codeerwinst ook vergroot worden.

Codewoorden bestaan in het geval van Huffman-codering altijd uit een geheel aantal bits, terwijl het optimale aantal, de entropie, dat meestal niet is (toevallig wel in het voorbeeld). Als er slechts twee symbolen zijn, bijvoorbeeld "a" en "b" dan biedt de Huffman-codering met een minimale codewoordlengte van 1 bit weinig soelaas. Een alternatief codeeralgoritme, het zogenaamde "arit-

Figuur 7: Illustratie van het basisidee van aritmetisch coderen. In a) wordt het bericht "abaaa" "rechttoe rechtaan" gecodeerd door het interval herhaaldelijk in gelijke subintervallen te verdelen. In b) wordt aritmetisch gecodeerd door, het interval herhaaldelijk in ongelijke subintervallen te verdelen. Het bijbehorende codewoord is in beide gevallen een getal in het meest rechtse subinterval



metisch coderen" kan in principe minder dan 1 bit per symbool toekennen [6].

Bij het aritmetisch coderen wordt het complete bericht gebruikt om ergens tussen de waarden 0 en 1 op de reële as, een interval aan te duiden. Het bijbehorende codewoord is dan een getal ergens in dat interval. Hoe groter het interval hoe minder bits nodig zijn voor het codewoord. De grootte van het interval wordt bepaald door de kansen van voorkomen van de symbolen. De plaats van het interval wordt bepaald door de kansen van voorkomen én de volgorde van de symbolen.

Aan de hand van het voorbeeldbericht "abaaa" tonen we in figuur 7a eerst wat intervallen betekenen bij het "rechttoe rechtaan" coderen en in figuur 7b bij het aritmetisch coderen. Herhaaldelijk zal bij het "rechttoe rechtaan" coderen een interval in twee gelijke subintervallen verdeeld worden en zal vervolgens een van de twee subintervallen gekozen worden afhankelijk van het te coderen symbool. In figuur 7a wordt het interval $[0, 1)$ om te beginnen verdeeld in $[0, 0,5)$ en $[0,5, 1)$ voor de respectievelijke symbolen "a" en "b". Omdat het eerste symbool een "a" is wordt verder gewerkt met het interval $[0, 0,5)$ wat vervolgens wordt verdeeld in de subintervallen $[0, 0,25)$ en $[0,25, 0,5)$. Omdat het tweede symbool een "b" is wordt nu verder gewerkt met het interval $[0,25, 0,5)$. Als uiteindelijk alle symbolen verwerkt zijn, resulteert een klein interval met de afmeting $0,5^5$. Voor elk ander bericht dat uit 5 symbolen ("a" of "b") bestaat zal uiteindelijk een interval met dezelfde afmeting resteren, maar het interval zal een andere plaats hebben. Elk van de mogelijke intervallen kan met $-\log_2(0,5^5) = 5$ bit gecodeerd worden. Hoe groter een interval hoe minder bits nodig zijn om het te coderen.

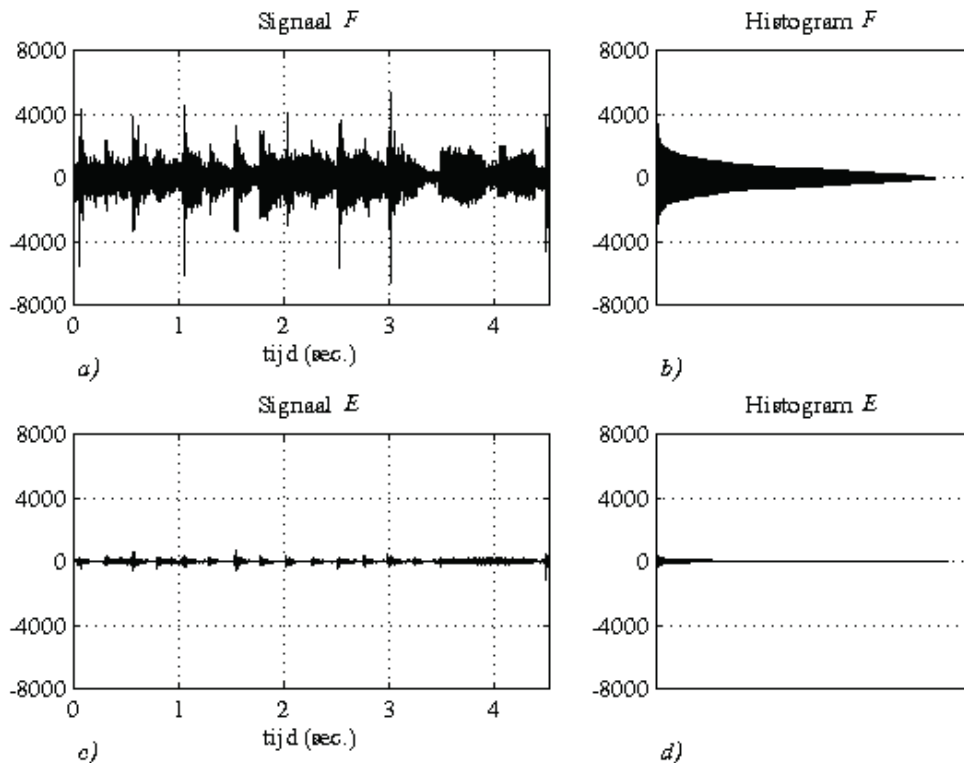
Hiervoor was aangenomen dat beide symbolen gelijke kansen (0,5) van optreden hebben waardoor alle bijbehorende codewoorden dezelfde lengte hebben gekregen. In het voorbeeldbericht hebben de symbolen "a" en "b" echter een respectievelijke kans van optreden van 0,8 en 0,2. Bij aritmetisch coderen wordt een interval niet perse in twee gelijke subintervallen verdeeld, maar worden de afmetingen gekozen afhankelijk van de kans van optreden van de symbolen. In figuur 7b wordt het interval $[0, 1)$ om te beginnen verdeeld in $[0, 0,8)$ en $[0,8, 1)$ voor de symbolen "a" en "b" respectievelijk. Omdat het eerste symbool een "a" is wordt verder

gewerkt met het interval $[0, 0,8)$ wat vervolgens wordt verdeeld in de subintervallen $[0, 0,64)$ en $[0,64, 0,8)$. Omdat het tweede symbool een "b" is wordt nu verder gewerkt met het interval $[0,6, 0,8)$. Als uiteindelijk alle symbolen verwerkt zijn, resulteert een klein interval met de afmeting $0,8^4 \times 0,2^1$. Voor elk ander bericht dat uit 5 symbolen ("a" of "b") bestaat zal uiteindelijk het interval een andere plaats en wellicht ook een andere afmeting hebben. Ook hier geldt dat hoe groter een interval hoe minder bits nodig zijn om het te coderen. Men kan aantonen [6] dat in een lang bericht het gemiddelde aantal bits per symbool gelijk is aan: $-0,8 \times \log_2(0,8) - 0,2 \times \log_2(0,2) \approx 0,72$ bit. Dit komt overeen met een codeerwinst van $\eta = 1 / 0,72 \approx 1,38$.

Hoewel het aritmetisch codeeralgoritme in principe niet moeilijk is, heeft men bij een implementatie onder andere te maken met een zeer grote nauwkeurigheid bij de berekening van de intervallen. Vandaar dat er heel veel werk is gedaan om dit algoritme praktisch bruikbaar te maken.

Voorbeeld

Aan de hand van een voorbeeld laten we zien wat de invloed van zowel predictie als entropiecodering is bij het coderen van een PCM signaal. In figuur 8a zien we voor 1 kanaal van een CD opname 1 frame, F , ter lengte van 200000 monsters, dit komt overeen met ongeveer 4,5 seconde speelduur. Op de CD wordt voor elk monster een 16-bits representatie gebruikt wat een signaalbereik van -32768 tot 32767 toestaat. In dit voorbeeld zien we echter dat $|F| < 7612$ waardoor het voor dit voorbeeldframe voldoende is om maar 14 in plaats van 16 bit per monster te gebruiken. Een triviale stap is dus om per frame aan te duiden hoeveel bits per monster nodig zijn. Als vervolgens de verdeling van mogelijke waarden van F bekeken wordt (figuur 8b), dan blijkt duidelijk een voorkeur voor waarden dicht bij de 0. Wordt voor deze verdeling een Huffman-code ontworpen dan resulteert dit, naast een nagenoeg verwaarloosbare hoeveelheid ballast, in een gemiddelde codewoordlengte van ongeveer 11,5 bit. Toepassing van lineaire predictie, hier orde 10, levert een residusignaal E (figuur 8c) met $|E| < 1658$ waarvoor in principe 12 bit per monster voldoende is. Wordt nu voor de verdeling van mogelijke waarden van E (figuur 8d) een Huffman-code ontworpen dan resulteert dit naast een nagenoeg verwaarloosbare hoeveelheid



Figuur 8: Van enkele seconden audiosignaal F (a) resteert na 10^{de} orde lineaire predictie het residu E (c). De bijbehorende histogrammen zijn respectievelijk (b) en (d)

ballast, voor een gemiddelde codewoordlengte van ongeveer 7,8 bit.

In geval van framing en uitsluitend entropiecodering is in dit voorbeeld een codeerwinst, $\eta = 16 / 11,5 \approx 1,4$ behaald. Wordt eveneens predictie gebruikt dan is de codeerwinst, $\eta = 16 / 7,8 \approx 2,1$. Dit voorbeeld leert dat predictie een wezenlijke bijdrage levert aan de codeerwinst.

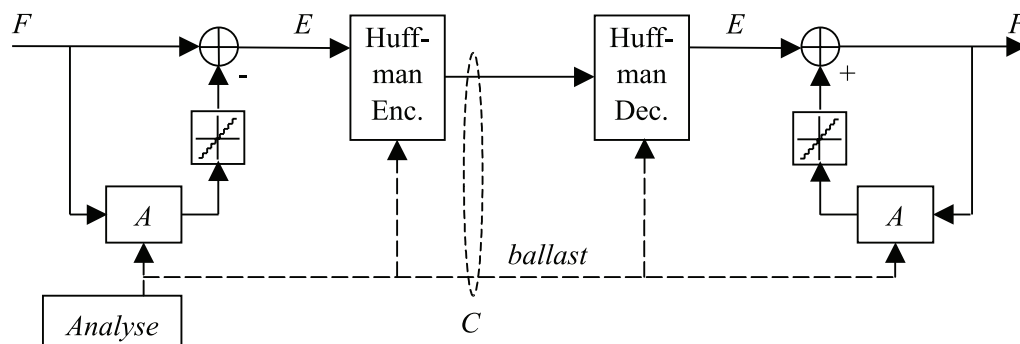
Toepassingen

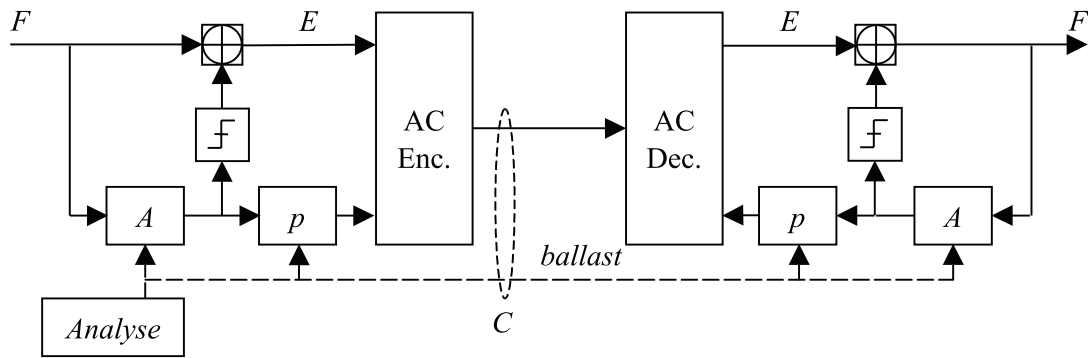
Momenteel kennen we twee populaire codeerschema's: Meridian Lossless Packing (MLP), dat wordt toegepast voor PCM signalen in DVD-audio en

Direct Stream Transfer (DST), dat wordt toegepast voor DSD signalen in Super Audio CD. Van beide schema's worden hierna slechts de elementaire kenmerken besproken en komen typische systeemaspecten niet aan de orde. In [7] en [8] worden beide codeerschema's gedetailleerd beschreven. Ofschoon er heel veel overeenkomsten tussen beide schema's bestaan zijn er ook diverse significante verschillen.

Met behulp van het predictiefilter A in figuur 9 wordt een schatting gemaakt van de geheeltallige PCM monsterwaarde. Daar de coëfficiënten van het filter niet geheeltallig zijn is de schatting dat in

Figuur 9: Principeschema voor het verliesvrij coderen van PCM signalen





Figuur 10: Principeschema voor het verliesvrij coderen van DSD signalen

beginsel ook niet. Door de schatting af te ronden naar een gehele waarde wordt voorkomen dat onnodig veel bits nodig zijn voor de representatie van het residu, E . Een teveel aan bits werkt in de encoder contraproductief. In figuur 10 wordt eveneens een niet geheeltallige schatting afgegeven maar nu voor een 1-bit DSD monster. Van de schatting wordt hier slechts het tekenbit gebruikt, een positieve schatting resulteert in een geschatte "1" en een negatieve schatting resulteert in een geschatte "0" van het DSD signaal. Hier wordt nu een "exclusive-or" gebruikt om de schatting en het signaal te combineren tot het 1-bit residu, E .

Voor PCM (figuur 9), wordt de verdeling van het residu geanalyseerd om, zoals eerder al is verklaard, de Huffman-code te ontwerpen. Voor DSD (figuur 10), wordt naast het residu ook de grootte van de schatting door het filter A , gebruikt om de kansverdeling, p , te bepalen en zo aan te passen aan de context. Omdat voor een DSD signaal het residu uit slechts 1 bit bestaat is een Huffman-code, met de minimale codewoordlengte van 1 bit, hier niet bruikbaar en wordt er daarom gebruik gemaakt van aritmetische codering (AC).

In beide gevallen hebben de decoders de inverse structuur en maken ze gebruik van parameters zoals die door de encoder zijn bepaald en overgestuurd. Dit oversturen is extra ballast in de code maar bespaart op de complexiteit van de decoders

die nu niet zelf de parameters hoeven te bepalen. Dit is met name van belang bij toepassing van de decoders in consumentenproducten.

Conclusie

Met behulp van relatief eenvoudige inzichten is het mogelijk om digitale audiosignalen verliesvrij te coderen hetgeen betekent dat de kwaliteit van het signaal in het codeerproces op geen enkele wijze wordt aangetast. Wat er uit komt is identiek aan wat er in is gegaan. Verliesvrije coders voor gegevensbestanden bestaan al veel langer, maar het coderen van signalen kan eenvoudiger en efficiënter met een speciaal daartoe ontworpen coder. Ook de gekozen representatie van het audiosignaal, PCM of DSD, heeft invloed op het codeerschema. Het gebruik van codering laat vervolgens een efficiënter gebruik van een communicatiekanaal of opslagmedium toe, voor audiosignalen is een codeerwinst van ongeveer 2 doorgaans goed te realiseren. De codeerwinst is echter heel sterk afhankelijk van de inhoud van het aangeboden signaal, voor stilte kan een veel hogere codeerwinst behaald worden dan voor luide ruisachtige signalen. De codeerwinst van een verliesvrij codeersysteem is ruim lager dan die voor de niet-verliesvrije codeersystemen maar hier staat tegenover dat de kwaliteit onberispelijk is.

Literatuur

- [1] M. Breeuwer. The Basics of Source Coding. *Tijdschrift van het Nederlands Elektronica- en Radiogenootschap*, 61(1): 15-18, 1996.
- [2] P. Nuijten en C. Dijkmans. Bitstream A/D- en D/A-conversie: Perfecte lineariteit met 1-bit-conversie. *KIVI & NIRIA: Digitale-Analoge Technologie*, 3(1/2): 28-35, 1993.
- [3] A.A.M.L. Bruekers and A.W.M. van den Enden. New Networks for Perfect Inversion and Perfect Reconstruction. *IEEE Journal on Selected Areas in Communications*, 10(1): 130-137, 1992.
- [4] S.L. Marple, Jr. *Digital Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [5] D. Huffman. A Method for the Construction of Minimum Redundancy Codes. *Proceedings of the IRE*, 40(10): 1098-1101, 1952.
- [6] I.H. Witten, R.M. Neal and J.G. Cleary. Arithmetic Coding for Data Compression. *Communications ACM* (30): 520-540, 1987.
- [7] M.A. Gerzon, P.G. Craven, J.R. Stuart, M.J. Law and R.J. Wilson. The MLP Lossless Compression System. *Proceedings of the AES 17th International Conference on High Quality Audio Coding*, 61-75, Florence, September 1999.
- [8] A.A.M.L. Bruekers, A.W.J. Oomen, R.J. van der Vleuten and L.M. van de Kerkhof. Improved Lossless Coding of 1-bit Audio Signals. *The AES 103rd Convention*, preprint 4563, New York,

CV

Fons (A.A.M.L.) Bruekers, geboren op 23 september 1957 in Nederweert, behaalde het HTS-E diploma in 1979 in Eindhoven en startte in 1981 op het Natuurkundig Laboratorium van Philips. Na 5 jaar gewerkt te hebben in een Local Area Network project veranderde hij van onderwerp en ging verder met onderzoek op het gebied van digitale signaalbewerking in het algemeen en digitale filtering in het bijzonder. Sinds eind 1995 doet hij onderzoek op het gebied van verliesvrije codering van audiosignalen.



Uitreiking Vederprijs



Het bestuur van het Vederfonds heeft besloten om de Vederprijs 1998 toe te kennen aan de heer dr. ir. Dignus-Jan Moelker.

Het onderzoek, op basis waarvan de toekenning is geschied, verschaft duidelijkheid omtrent de vraag of het mogelijk is, om met voldoende nauwkeurigheid en extreme bedrijfszekerheid verkeersvliegtuigen met behulp van satellietplaatsbepaling onder zeer slechte zichtomstandigheden veilig te kunnen laten landen. De heer Moelker werd geconfronteerd met het gegeven, dat de uiterst zwakke spread-spectrum satelliet signalen, die ongeveer 20 dB onder het natuurlijke ruisniveau liggen en vanuit de ruimte worden ontvangen, door allerlei natuurlijke en "man made" stoorbronnen, gemakkelijk onbruikbaar kunnen worden. Deze effecten dreigen een streep te halen door de talrijke toepassingsmogelijkheden van satellietnavigatie, wanneer daar risico's ten aanzien van veiligheid of milieu aan verbonden zijn. Vergelijkbare interferentieproblemen komen aan de orde bij spread-spectrum communicatiesystemen, vooral bij mobiele toepassingen. Moelker heeft de moeilijke, maar uiteindelijk vruchtbaar gebleken weg gekozen om deze interferentieproblematiek in drie fasen aan te pakken.

- De eerste stap was het modelleren van interferentie-effecten op de synchronisatieprestaties van spread-spectrum ontvangers. Daarbij heeft hij vooral aandacht geschonken aan de theoretische grenzen en de mogelijkheden van realistische ontvangerconcepten.
- Vervolgens heeft hij modellen ontwikkeld voor de capaciteit van co-existerende breedband- en smalband mobiele communicatiesystemen. Omdat spread-spectrum technologie gebruikt zal gaan worden voor toekomstige mobiele telecommunicatie, heeft hij met name aandacht geschonken aan de co-existentie van spread-spectrum (CDMA) en smalband-concepten zoals GSM.
- Tot slot heeft hij interferentie-onderdrukkings-technieken geoptimaliseerd voor satelliet-plaatsbepalings-systemen zoals het Amerikaanse global positioning system (GPS) en het Russische global navigation satellite system (GLONASS). De mate van succes heeft hij daarbij niet uitgedrukt in parameters als dB, maar in voor navigatie-toepassingen meer relevante grootheden als nauwkeurigheid van de positie-bepaling.

Met zijn aanpak heeft Dr. Moelker tevens duidelijk gemaakt, dat radiocommuni-

catie- en radio-plaatsbepalings-systemen zeer nauw verbonden zijn. Het grootste onderscheid tussen beide klassen van systemen is, dat bij telecommunicatie vooral informatie over-gebracht wordt, terwijl bij navigatie veeleer bekende informatie wordt verstuurd, maar dat daarbij de propagatietijd de eigenlijke informatie in zich bergt. Opvallend is, dat hij interferentie in verschillende domeinen al dan niet gelijktijdig aanpakt. Daarbij moeten we denken aan het tijd-, frequentie-, modulatie-, code-, amplitude- en tot slot het spatiële domein.



De heer Moelker tijdens zijn presentatie.

De heer Moelker heeft bewezen, dat grondige theoretische analyse van interferentie en de gevolgen daarvan belangrijk zijn, en vervolgens de wegen aangegeven hoe op basis van deze grondige kennis interfe-

rentie bestreden kan worden. Dat wil zeggen hoe de scheiding tussen gewenste informatie en ongewenste interferentie geoptimaliseerd kan worden. De scheiding van signaal en interferentie neemt in belangrijkheid sterk toe, doordat de mate waarin het scheiden van "vriend en vijand" slaagt, in belangrijke mate gaat bepalen wat de praktische capaciteit van het beschikbare frequentiespectrum zal zijn. Het behoeft daarom geen betoog wat de economische impact kan zijn

van efficiënte en effectieve interferentiebestrijding.

De prestaties van Dr. Moelker zijn inmiddels ook internationaal bekend geworden, wat blijkt uit het feit, dat hij in 1997 een Best Paper Award kreeg van het prestigieuze Amerikaanse "Institute of Navigation".

Gezien deze overwegingen is het Vederfonds zeer verheugd, dat in de persoon van dr. ir. Moelker de Vederprijs ook in 1998 weer een waardig prijswinnaar kent, die representatief is voor het hoogwaardige wetenschappelijk

onderzoek dat op radiogebied in Nederland plaatsvindt.

En dan verzoek ik thans de voorzitter van het Vederfonds, Mevr. Françoise Olthof-Kosters, om de prijs te overhandigen aan de heer dr. ir. Dignus-Jan Moelker.



Mevr. Françoise Olthof-Kosters, voorzitter van het Vederfonds, overhandigt de prijs aan de heer dr.ir. Dignus-Jan Moelker.



Voorwoord van de redactie

In nummer 4 van 1999 van het Tijdschrift van het NERG is er tot onze grote spijt een paragraaf weggevallen uit het artikel "Bedrijfsbezoek PHILIPS Gehoorapparaten", geschreven door ir. L.H. Geerdink en Ing. P. Termeer. Het ontbrekende stuk is hieronder alsnog geplaatst.

Slechthorendheid en soorten hoorapparaten

Ing. Peter Termeer

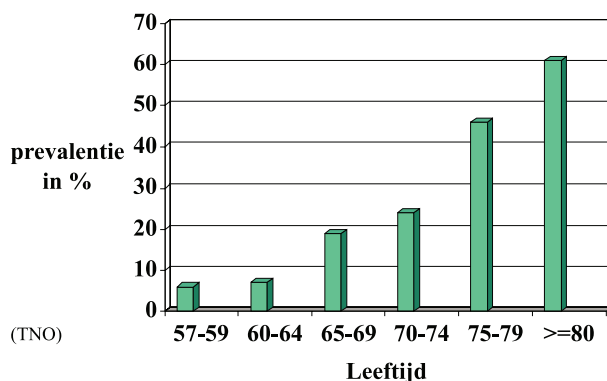
Beltone Netherlands b.v./Philips Hearing Instruments
Afdeling R&D, Eindhoven.

Ongeveer 1 miljoen Nederlanders hebben in beide oren een gehoorverlies van tenminste 35 dB (fig. 1). Dit verlies is gemeten met zuivere tonen in het frequentiegebied van 500 tot 4000 Hz, een gebied dat van groot belang is voor het verstaan van spraak. Die toondrempel van 35 dB is bij de ziekenfondsen een criterium om voor eventuele hoortoestelvergoeding in aanmerking te komen.

De meest frequente oorzaken van minder goed horen zijn presbycusis (ouderdomsslechthorendheid), otosclerose (vastgroeien van een van de gehoorbeentjes) en blootstelling aan lawaai. Otosclerose kan meestal met een operatie worden verholpen. Verder zijn er nog tientallen minder frequente oorzaken van slechthorendheid.

Figuur 1

Slechthorendheid vs leeftijd



Niet iedereen met 35 dB gehoorverlies heeft in dezelfde mate problemen met het verstaan van spraak. De feitelijke spraak-analyse-functie van het gehoor kan bij personen met gelijke toondrempels toch heel verschillend zijn. Ook zijn de luisteromstandigheden niet voor iedereen gelijk. De een moet tijdens het werk veel communiceren met mogelijk nog storend geluid op de achtergrond; de ander is vaak thuis en heeft zijn of haar kaartavondjes in een rustige en goed gestoffeerde (en daardoor voor spraakverstaan gunstige) huiskamer. Maar voor de meesten wordt de communicatie toch enigszins bemoeilijkt. Hoortoestellen zijn dan onmisbare hulpmiddelen.

Afgezien van de verschuiving van de toondrempel (verzwakking) treedt bij de meeste vormen van slechthorendheid ook verlies van frequentie- en tijdoplossend vermogen op. Dit verslechtert het spraakverstaan. Ook is vaak de oorspan (het bereik tussen het zachtste waarneembare en het luidste nog acceptabele geluid) aanzienlijk verkleind. Dan is er nog maskering. Dit is het effect dat sommige (gewenste) geluiden onhoorbaar worden gemaakt door andere (ongewenste) geluiden die gelijktijdig of kort ervoor aanwezig zijn. Vaak blijkt dat bij slechthorenden deze maskering sterk is vergroot. Een hoortoestel kan bovengenoemde effecten verminderen of zelfs geheel compenseren.

Hoortoestellen kunnen worden ingedeeld naar o.a. de plaats waar ze op of in het lichaam worden



figuur 2.

gedragen. Het kasttoestel (fig. 2) draagt men op de borst en is door een snoertje verbonden met de oortelefoon. Dit toestel is door zijn grote knoppen en penlite batterij geschikt voor mensen met een beperkte motoriek. Ook vanwege zijn zeer hoge versterking wordt het toegepast. Verder wordt het toestel gebruikt in landen waar geen kleine batterijtjes te koop zijn. Dan is er het achter-het-oor-toestel (fig. 3).

Dit is in Nederland het meest populaire type. Behalve een microfoon heeft het ook een luisterspoel, waarmee via inductie geluiden uit een ringleiding worden weergegeven. Ze kunnen heel klein zijn waardoor ze niet erg opvallen. Het in-het-oor-toestel (fig. 4) wordt in de oorschelp en de gehoorgang gedragen. Zeker de kleinste typen, die (bijna) geheel in de gehoorgang zitten, hebben geen luisterspoel. De batterijtjes voor deze miniatuur toestellen zijn erg klein en daardoor iets duurder in het gebruik. Dan is er nog het beengeleider hoortoestel, meestal in de poot van een bril geplaatst. Hierbij wordt geluid via een trilplaat aangeboden aan het harde bot achter de oorschelp, dicht bij het slakkenhuis. Er bestaat een uitvoeringsvorm waarbij de trilplaat op een schroef in het bot achter de oorschelp wordt geklikt (fig. 5). Dit verbetert de overdracht.



figuur 4.

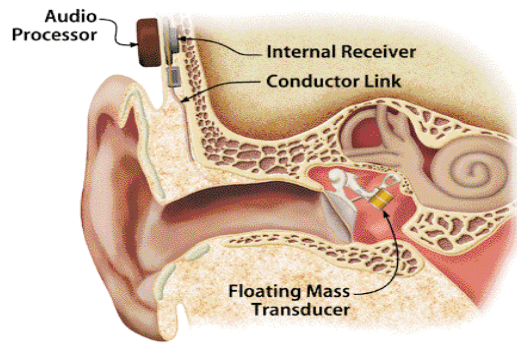
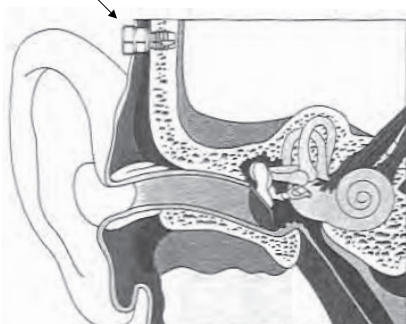


figuur 3.

De batterijtjes voor deze miniatuur toestellen zijn erg klein en daardoor iets duurder in het gebruik. Dan is er nog het beengeleider hoortoestel, meestal in de poot van een bril geplaatst. Hierbij wordt geluid via een trilplaat aangeboden aan het harde bot achter de oorschelp, dicht bij het slakkenhuis. Er bestaat een uitvoeringsvorm waarbij de trilplaat op een schroef in het bot achter de oorschelp wordt geklikt (fig. 5). Dit verbetert de overdracht.

figuur 5: Botverankering voor beengeleider-hoortoestel van Entific

Schroef voor bevestiging beengeleider



figuur 6: Middenoor implantaat Symphonix

In nog experimentele vorm zijn er middenoorimplantaten. Hierbij wordt een trilling direct op de gehoorbeentjes overgedragen (fig. 6). Het deel van het hoortoestel waarin microfoon, elektronica en batterij zitten wordt extern gedragen. Tenslotte is er het cochleaire implantaat. Wordt bij alle eerdergenoemde typen het oor akoestisch of door middel van trilling gestimuleerd, bij het cochleaire implantaat gebeurt dat door middel van elektrische stroompjes. Hierbij wordt het microfoonsignaal gecodeerd en aangeboden aan een veelkanalige elektrode in het slakkenhuis (fig. 7). De elektrode zendt elektrische stroomimpulsen uit die kunnen worden opgevangen door de gehoorzenuw. Dit systeem wordt toegepast bij zeer ernstig slechthorenden, die met een akoestisch hoortoestel onvoldoende te helpen zijn.



- 1 = slakkenhuis-elektrode
- 2 = ontvanger
- 3 = spraakprocessor
- 4 = zender
- 5 = microfoon

figuur 7: Cochleair implantaat van Cochlear

In moderne hoortoestellen zoals die van de SpaceLine van Philips wordt digitale signaalbewerking ingezet om de eerdergenoemde problemen te verhelpen of te verlichten. De digitale techniek heeft o.a. het voordeel dat het geluid beter kan worden aangepast aan de specifieke stoornissen in het oor. Verschillende frequentiegebieden kunnen ieder hun eigen signaalbewerking ondergaan, passend bij die van het individuele slechthorende oor. De verstoorde luidheidsopbouw (de sterkte

waarmee een geluid wordt waargenomen) kan hierbij optimaal worden hersteld. Het gedrag van compressors in het tijddomein kan op de individuele behoefte worden ingesteld. Tenslotte zorgt ruisonderdrukking ervoor dat zachte spraak wel wordt versterkt, maar dat zachte omgevingsgeluiden niet op de voorgrond treden. Hoewel slecht-

horendheid over het algemeen een zeer complexe verandering van de gehoorfunctie teweegbrengt, helpen digitale technieken ons om stap voor stap verbetering te brengen in de spraakwaarneming door slechthorenden.

CV

Peter Termeer behaalde het diploma HTS-E in 1966. Na militaire dienst begon hij in 1968 als ontwikkelaar (electrisch/akoestisch) bij Philips Hoor toestellen. Hij volgde o.a. cursussen audiologie, automatisering en digitale signaalbewerking en liep stages in audiologische centra. Zijn huidige werkzaamheden zijn geheel gericht op de audiologische aspecten van hoortoestellen en computerprogramma's ter ondersteuning van het aanpasproces



ANNOUNCEMENT AND FINAL CALL FOR PAPERS

for the

21st Symposium on INFORMATION THEORY in the BENELUX

Thursday, May 25, and Friday, May 26, 2000

NIAS Conference and Study Facilities, Wassenaar, NL



The program will consist of two invited lectures and a number of both oral and poster presentations contributed by researchers in the Benelux. The two invited lectures will address the following topics:

- “The TU-Delft Research Program ‘Ubiquitous Communications’”,
by Prof.dr.ir. R.L. Lagendijk, Information and Communication Theory Group, Faculty ITS, TUD
- “Video Processing for Multimedia Systems”,
by Dr.ir. G. de Haan, Philips Research Laboratories, Eindhoven

Prospective authors are encouraged to submit papers according to the timing schedule and deadlines listed below. Papers are solicited in the following (or related) areas:

- * Shannon theory and applications of information theory
- * Multi-user information theory
- * Source and channel coding, data compression
- * Cryptography and data protection issues
- * Audio/video signal processing and compression
- * Speech/signal processing, pattern recognition
- * Signal detection and estimation
- * Terminals for ubiquitous communication
- * Wired and wireless communication networks for ubiquitous communications
- * Audio/visual signal and application processing for ubiquitous communications

Timing schedule and deadlines:

Feb. 4, 2000 Deadline for submission of
1 page summary
Feb. 25, 2000 Notification of acceptance

April 7, 2000 Deadline for submission of full
paper
April 15, 2000 Deadline for registration
May 25-26, 2000 Symposium

To submit a paper and present your work at the symposium, please send a one-page summary to the symposium chair and specify topic area and preferred type of presentation (oral or poster). Electronic submission is also possible by email to J.Biemon@its.tudelft.nl specifying “Benelux IT-2000” on the subject line.

Symposium Organization and Local Arrangements:

Jan Biemon (chair)
Delft University of Technology
Faculty of Information Technology and Systems
Dept. of Mediamatics
Mekelweg 4
2628 CD Delft
Phone : +31 152 784695
FAX : +31 152 781843
Email : J.Biemon@its.tudelft.nl

Annett Bosch
Delft University of Technology
Faculty ITS
Dept. of Mediamatics
Mekelweg 4
2628 CD Delft
Phone : +31 152 786052
FAX : +31 152 781843
Email : A.BoschvanDijk@its.tudelft.nl

Program Committee:

J. Biemond
A. Hanjalic
E.A. Hendriks
R. Heusdens
R.L. Lagendijk

J.C.A. van der Lubbe
M.T.J. Reinders
J. Simonis
J.H. Weber

Web site:

Updated information on this symposium is available on <http://www-ict.its.tudelft.nl/benelux>

Co-operation:

This Symposium is organized in co-operation with the IEEE Benelux Chapters on Information Theory, Signal Processing, Consumer Electronics, and Vehicular Technology and Communication. Please, pass this information on to your colleagues

Note: Due to space limitations participation is limited to a maximum of 50 persons.

Werkgemeenschap voor Informatie- en Communicatietheorie in de Benelux

The purpose of the Werkgemeenschap voor Informatie en Communicatietheorie (WIC) is to further the development and application of information and communication theory. The WIC shall further this goal by

- stimulating the exchange of knowledge, experience, and insights between professionals working in information and communication theory.
- supporting the preparation and realization of projects in her field of interest.

The main activities

- *Benelux Symposium on information and communication theory.* This two days symposium is held yearly in the month of May. The purpose is to allow researchers in the Benelux to meet and present their recent research.
- *Midwinter meeting.* These yearly meetings are organised in January. They are of an introductory and tutorial nature and treat a topic within our field that is of current interest to a broad audience.
- *Informant.* The Informant is the annual newsletter of the WIC. Its main goal is to inform the members about the current research of their colleagues, their fields of expertise, recent publications and other scientific achievements. It also reports on the relevant

events of the last year and gives an agenda of upcoming conferences, scientific visits planned by or to the members.

- *WIC WWW pages.* This is like an electronic version of the Informant. It contains more recent announcements and other information relevant to members and other persons interested in information and communication theory.

Members

Membership is open to everyone who is or has been active in the field of information and communication theory in the Benelux. The membership of the WIC will bring you in contact with researchers in information, communication and coding theory in the Benelux. You will receive our newsletter, the Informant (mainly in English), and you will receive timely invitations to meetings and symposia (co)-organised by the WIC. Currently the WIC requires no membership fee.

Contacting the WIC

The WIC's Secretary is
Dr.ir. L. Tolhuizen,
Philips Research Laboratories, WY 61.
Prof. Holstlaan 4, 5656 AA Eindhoven, Nederland.
Telephone: +31 40 2742069 Fax: +31 40 2744660.
Email: tolhuizn@natlab.research.philips.com.

