

1. TV-sändning med mångkanalsljud i 5.1-format

I denna uppgift är Erik Lundbeck på SVT källa om inget annat anges.

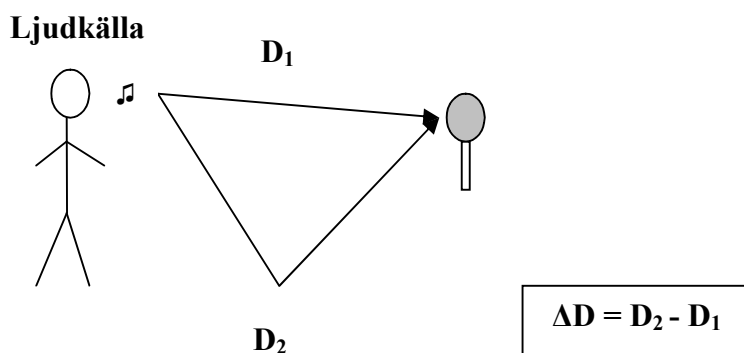
a)

Rundgång: Eftersom högtalare används finns risk för rundgång där mikrofonerna på scen fångar upp det förstärkta ljudet som sedan spelas upp ännu högre.

Balans i ljudbilden: Publikens sång och skrik är viktig för att ge ljudbilden rätt atmosfär, men bör inte vara för dominant. Det finns dock en risk för att ljudet blir för starkt och dominerande och därför tar över ljudbilden. I de flesta fall uppnås den bästa ljudbilden genom att låta centerkanalen dominera och de övriga kanalerna komplettera denna kanal. Balans i ljudbilden uppnås bland annat genom att reglera mikrofonernas dynamik, omfång, känslighet och placering.ⁱ

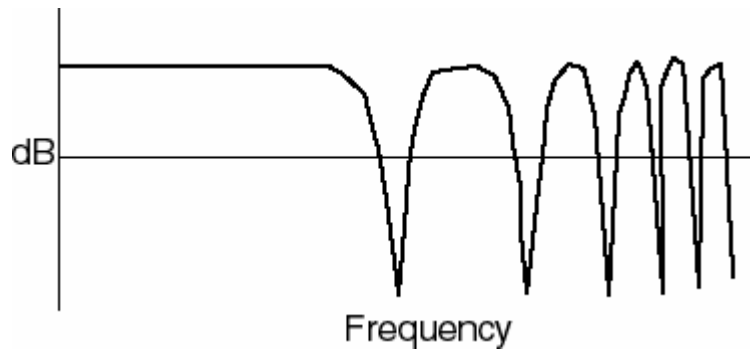
Ofta används två mikrofoner för varje surroundkanal för att publikljudet ska bli mer homogent. Risken finns nämligen att exempelvis en väldigt högljudd person som befinner sig nära en av mikrofonenheterna annars kan få ett oproportionellt stort genomslag i den aktuella kanalen. Använder man två mikrofoner kan man mixa ihop dessa kanaler för att dölja oönskade ljud med hjälp av exempelvis konstruktiv eller destruktiv interferens.

Kamfiltereffekt/överhörning: Kamfiltereffekten är ett interferensfenomen som uppstår då direktljud och reflekterat/fasförskjutet ljud släcker ut varandra. Förutsättningen för att detta ska inträffa är att skillnaden i väg, ΔD , är ett halvt antal våglängder, det vill säga att D_1 har topp när D_2 har dal.



Figur 1. Vägskillnad som kan leda till kamfiltereffekt

Resultatet blir att ljudet släcks ut för alla frekvenser som uppfyller detta krav. För vissa frekvenser kommer skillnaden i väg istället att bli ett helt antal våglängder, vilket ger toppar i spektrat.



Figur 2. Kamfiltereffektens verkan på en signal

I exemplet med inspelning av *Allsång på Skansen* finns det risk för denna typ av utsläckning exempelvis då surroundmikrofonerna i publiken fångar upp PA-ljudet från högtalarna på scen. Det resulterande ljudet får då en svajig karaktär. Samma typ av problem uppstår vid mixning vid efterbehandling. Lösningen på detta problem är att placera mikrofonerna tillräckligt långt ifrån varandra så att nivåskillnaden minskar dalarnas hörbarhet.

b) Dolby Digital är en överföringsdataström, vilket innebär att den är avsedd för att just överför ljudinformationen till mottagaren via ett visst medium. Den kan hantera ett kodat program med upp till sex kanaler som beskrivs med hjälp av en metadataström. Metadata skickas med i dataströmmen och innehåller den information som är nödvändig vid för att kontrollera kodning och avkodning på väg till mottagaren i hemmet (nedströms). Vanligtvis överför den kodade ljudströmmen tillsammans och metadatan som en bitström i två digitala kanaler. SVT använder sig exempelvis av det digitala transmissionsgränssnittet AES/EBU (Audio Engineering Society/European Broadcasting Union). Andra digitala standarder är AES3 och S/PDIF. Genom att använda metadata kan producenterna kontrollera hur ljudet reproduceras och återges hos mottagaren/åhöraren. Exempel på parametrar som kan kontrolleras av metadata är konfigurering av högtalare, kompression, bass management, dynamiskt omfång, nivån för center och surround nermix, rummets akustik, mixningsnivå och olika filter. Åhöraren ställer då in dessa parametrar på sin mottagare efter tycke och smak. Tack vare detta kan lyssnaren exempelvis få jämnare ljudnivåer, då man inte måste styra ut alla program till full utstyrning.^{ii, iii}

c) Vid sändning av TV-program i 5.1 sänder SVT information om downmix som metadata. När åhöraren i sitt hem ställer om från surround till stereoformat eller stoppar in stereohörlurar i uttaget använder avkodaren metadatan för att mixa samman de sex kanalerna till de två stereokanalerna. Bland annat innefattar datan information om hur mycket centerkanalen och surroundkanalerna ska dämpas innan de mixas ihop i vänster och höger kanaler (left total och right total). Dämpningen är nödvändig eftersom ljudinformationen i dessa kanaler inte är lika centrala vid återgivning i stereo. Hur mycket signalerna ska dämpas beror på vilket slags ljud det är. Vid konserter och matcher är publiken viktig för att ge rätt atmosfär, varför man ofta väljer att inte dämpa detta surroundljud så mycket vid downmix. Däremot vill man undvika för mycket diffust ljud.

d) Dolby Pro Logic II är en avkodningsteknik som använder matristeknik för att omvandla stereoljud till simulerat femkanaligt surroundljud: vänster, center, höger, vänster surround och höger surround. Surroundljudet är inte diskret och ger därför inte en lika verklighetstrogen och klar surroundbild som exempelvis Dolby Digital Dolby, men ljudet

beskrivs ofta som en av de bästa indiskreta kodningsteknikerna. Dolby Surround är en av de allra äldsta teknikerna för kodning av surroundljud. Vid uppspelning separerar dekodern surroundkanalen från mixen och skickar ljudet via en och samma kanal till en eller två surroundhögtalare. Ljudet som spelas upp i högtalarna är således samma ljud. Den främsta fördelen med Dolby Pro Logic II är att den till skillnad från exempelvis Dolby Surround har ett fullt frekvensomfång i de viktiga bakkanalerna (surroundkanalerna). Vid uppspelning av surroundljud med Dolby Surround kan ljudbilden kantra märkbart i vissa lägen/frekvenser.

Generellt gäller ändå att även vanlig stereo i regel kan avkodas med god behållning. Men får en extra närvarokänsla på grund av att vissa delar av ljudet fyller rummet även bakifrån. Med vanlig Dolby Surround kan dock ljudbilden ha märkbara brister i vissa lägen, trots att det fenomenet motverkas av fördröjning och diskantsänkning på bakkanalen. Detta problem är inte lika märkbart med Dolby Pro Logic II, men vissa kan istället uppleva ljudbilden som överdrivet dynamisk i vissa lägen vilket bland annat kan bero på spänningsstyrda förstärkningar som uppstår vid avkodningen. Dolby Pro Logic II har också fördelen att kunna ge lyssnaren möjlighet att justera ljudbilden efter eget tycke. Parametrar som kan justeras är dimension (flytta ljudscenen framåt eller bakåt), centerbredd (var centerljudet ska återges - endast i center kanalen, endast i vänster och höger eller i båda) och panorama (inkluderar surroundkanalerna i stereokanalerna för att ge en bredare ljudbild).^{iv, v}

e) Dolby E är till skillnad från exempelvis Dolby Digital 5.1 en kodningsteknik för distribution/lagring av ljuddata snarare än transmission. Dolby E omfattar kompressionstekniker som möjliggör att exempelvis ett AES/EBU-kanalpar kan lagras upp till åtta kanaler och en kanal för metadata för efterbehandling och distribution. Varje program lagras diskret, det vill säga åtskilt från de övriga, med egen metadata.ⁱⁱ SVT utnyttjar de tillgängliga ljudspåren på följande sätt:

kanal 1 + 2	stereo (L_T och R_T)
kanal 3 + 4	de sex kanalerna komprimeras bitmässigt med avkodningsstandarden Dolby E och tilldelas varsin kanal

Som tidigare har nämnts möjliggör Dolby E lagring i 8 kanaler. De två ljudkanaler som återstår i spår 3 och 4 utnyttjas inte för tillfället med skulle kunna tänkas användas till att överföra ljud på ett andra språk (exempelvis finska) eller stereoljud i något annat format än Dolby Pro Logic II.

En fördel med Dolby E är att den kodade ljudsignalen kan ändras, avkodas och omkodas flera gånger utan att hörbara försämringar uppstår. Den är också fullständigt kompatibel med Dolby Digital-formatet och tillämpar samma metadatafunktion. Dolby E belastar inte nätverk och överföringsmedium i samma utsträckning tack vare kompression.

Nackdelar med Dolby E är som för andra kompressionstekniker att viss data går förlorad. Återgivningen av ljudet blir således inte exakt. Genom att tillämpa psykoakustiska principer/kunskaper är det dock ytterst sällan som dessa dataförluster är märkbara för det mänskliga örat. En annan nackdel är att Dolby E, till skillnad från exempelvis Dolby Digital, kräver att ljuddata ”dekomprimeras” innan det mixas eller ändras.^{vi}

2. Rumslig ljudåtergivning

I detta avsnitt är fakta hämtad från Dr. Damian Murphys föreläsningssanteckningar^{vii} om inget annat anges.

a) Ambisonic-system används för att återge surroundljud i 2D eller 3D och bygger på vetenskapliga teorier om ljudperception. Systemet skapades i England redan på 70-talet och var främst tänkt att användas för återgivning av musik. Det som kännetecknar systemet är att det baseras på att återge ljudet i en exakt punkt i ljudrymden. I verkligheten är detta omöjligt, främst på grund av begränsningar i mikrofonerna, men genom att tillämpa matematiska formler är det möjligt att utifrån inspelat material beräkna fram ljudet i varje punkt. Med hjälp av dessa matematiska formler är det i Ambisonics-system möjligt att skapa surroundljud på syntetisk väg.

5.1-system används liksom Ambisonics för att återge surroundljud. Skillnaden mellan de båda systemen är att 5.1-systemet i sitt normala tillstånd endast återger surroundljudet i 2D. Dessutom har systemet främst utvecklats för återgivning i kombination med visuell feedback, såsom film. Till följd av detta har dessa ljudåtergivningssystem utvecklats utifrån teorin om att det finns ett fram och bak i rummet, där du som åhörare är riktad framåt. 5.1-systemen tillämpar följaktligen en bärande ljudscen framför åhöraren och en kompletterande surroundscen bakom åhöraren. Ambisonics har i motsats till detta valt att vikta all riktningssinformation lika, det vill säga en 360° ljudbild.

b) Vid inspelning av Ambisonics-ljud används vanligtvis en soundfield-mikrofon eller liknande. En Soundfield-mikrofon är konstruerad särskilt för inspelning av 3D-ljud och består av fyra dubbelkardioider ("åttor") som bildar en tetrahedral kring en punkt. Surroundljudet spelas oftast i så kallat B-format, som består av fyra signaler, X (fram - bak), Y (väster - höger), Z (upp - ner) plus en monosignal W som spelas in med en rundupptagande mikrofon och anger summan av vänster, höger, fram, bak, upp och ner. X, Y och Z-signalerna har alla förstärkta med 3 dB i någon riktning i förhållande till W-signalen (0 dB). Signalerna står alla i ett inbördes förhållande till varandra och ger i kombination med varandra information om tryck och hastighet i alla riktningar i rummet. Som tidigare nämnt är det möjligt att med de matematiska formler som tillämpas inom Ambisonics koda om en monofonisk signal till B-format-signaler i polär form.

Vid inspelning av surroundljud i 5.1 format används en femkanalig mikrofonvektor där de fem kanalerna spelas in separat. Mikrofonerna placeras ut för att ge en så naturtrogen och dynamisk ljudbild som möjligt. Vid inspelningstillfället måste man således noga överväga var mikrofonerna ska stå i förhållande till varandra och den ljudbild som ska spelas in. Vid mixning är det viktigt att tänka på att två eller fler mikrofoner kan ha fångat upp samma ljud, vilket kan leda till att signalerna släcker ut eller förstärker varandra.

c) Lagringsförfarandet är kanske det mest kännetecknande för Ambisonic-system. Systemen skiljer sig från andra ljudåtergivningssystem genom att kodningen och avkodningen är separerade och oberoende. Vid inspelning kodas fyra kanaler, kallade W, X, Y, Z, vilka kan avkodas för i princip ett valfritt antal högtalare i 2D eller 3D vid återgivning. Ambisonics har fördelen av att tillämpa ett mono- och stereokompatibelt, hierarkiskt avkodningsschema kallat UHJ som möjliggör att Ambisonics-material kan överföras via två eller flera kanaler. Fördelen med detta är att det tillför flexibilitet och ger maximal effekt utifrån antalet kanaler som används för tillfället.

I 5.1-system är kodning och avkodning inte separerade. Om ljudet är kodad för sex kanaler avkodas det också för sex kanaler. Signalerna kan antingen lagras diskret, det vill säga separat, eller mixas ner till stereo. Informationen i surroundkanalerna och centerkanalen överförs då enligt vissa algoritmer till de båda stereohögtalarna, L och R. Ofta överförs ljudet både i 5.1 och stereo för att anpassa mediet för alla mottagare. Ett alternativ är att använda en Dolby Pro Logic II kodare för kodning av surroundljudet. Vid återgivning kan då en Pro Logic II avkodare användas för att återskapa eller simulera surroundljudet enligt specifika algoritmer.

d) Avkodning av Ambisonics-ljud är relativt enkelt. De fyra B-formatsignalerna passerar genom ett filter för anpassas till psykoakustiska aspekter, varpå de kombineras utifrån valda avkodningsmatriser. Vid avkodningen är det möjligt att välja vilken virtuell mikrofonpolaritet (rundupptagande - åtta) som man vill att ljudet som återges av respektive högtalare ska ha spelats in med.

Vid uppspelning av Ambisonics-ljud avkodas de fyra informationskanalerna för att reproducera en fullständig 3D-ljudbild över ett valfritt antal högtalare. *Dr. Damian Murphy* nämner i sitt kompendium fyra fördelar med detta tillvägagångssätt.

1. Endast fyra kanaler behövs för att kunna återge ett fullständigt surroundljud i 3D. På samma sätt räcker det med tre kanaler för att återge ljud i 2D.
2. Tack vare att de fyra signalerna redan är korrelerade i tid underlättas och förbättras mixningen.
3. Samma material kan avkodas för fyra kanaler för hemmet som för åtta i biosalongen. Om dagens surroundformat ändras med avseende på antalet kanaler/högtalare, kommer det således inte att vara nödvändigt att ändra ljudmaterialet.
4. Till skillnad från 5.1-systemen bidrar alla högtalare till att generera den kompletta ljudbilden.

Vid uppspelning av surroundljud i 5.1 krävs som bekant sex kanaler och således också sex högtalare, vänster (L), höger (R) och center (C) som placeras längst fram, två surroundhögtalare (S_L respektive S_R) som placeras till vänster respektive höger bakom åhöraren samt en kompletterande lågfrekvenskanal (subwoofer) som återger de allra lägsta frekvenserna. I 5.1-systemen är kodning och avkodning beroende av varandra; kodas ljudet för sex kanaler måste det också avkodas för sex kanaler. En nackdel med 5.1-system är att de, till skillnad från Ambisonics, ställer mycket specifika krav på högtalarnas placering, både i förhållande till varandra och till åhöraren för att surroundljudbilden ska bli bra. Det är viktigt att åhöraren befinner sig i "sweet spot". Ljudet återges dessutom endast i 2D, eftersom specifik höjdinformation saknas. Ytterligare ett problem, som delvis också kan förekomma vid återgivning av Ambisonics-ljud är att det uppstår "hål" emellan högtalarna, där ljudbilden vacklar. Problemet med "sweet spot" är dock inte lika kritiskt i Ambisonics-system.

Det finns naturligtvis även rekommendationer för inställning och placering av högtalare vid återgivning av Ambisonics-ljud.

På webbsidan www.ambisonics.net^{viii} kan man läsa följande:

- Högtalarna för ha samma effektnivå och ha samma fasrespons. Det bästa sättet att uppnå detta på är att använda identiska högtalarenheter.
- Alla förstärkare och högtalare bör täcka hela spektrat av frekvenser. Till skillnad från andra surroundljuds-system finns det inga särskilda surroundkanaler och signalerna är inte filtrerade.
- Om du använder fyra högtalare bör dessa placeras i en rektangel med långsidan i riktningen fram-bak. Alla bör vara riktade mot en punkt i mitten av rummet.
- Om du använder fem högtalare bör dessa placeras så att de bildar en femhörning, riktade mot mitten av rummet.
- Om du använder sex högtalare ska dessa bilda en sexhörning. Även i detta fall ska högtalarna riktas mot en punkt i mitten av rummet.
- En högtalare kan flyttas närmare mittpunkten men bör då matas med en fördröjd signal så att inte ljudbilden rubbas.

Som tidigare nämnts, har man valt att i Ambisonics-system implementerat avkodnings-schemat UHJ för att ljudet ska vara kompatibelt med olika återgivningssystem. Schemat omfattar liksom B-formatet fyra kanaler. Skillnaden är dock att två av dessa är kompatibla med stereoåtergivning. UHJ gör det således möjligt att utesluta den fjärde och tredje kanalen och på vis skala bort höjdinformation respektive horisontal surroundinformation. De resterande två kanalerna kan sedan användas för återgivning i stereo.

En stor nackdel med både UHJ och B-formatet är de kräver en särskild avkodare för att surroundljudet ska upplevas som tillfredsställande. Man har därför fått upp ett intresse för att kombinera Ambisonics med 5.1-formatet genom att koda om B-formatsignalerna så att de direkt, utan en särskild avkodare, kan spelas upp i ett 5.1-system (G-format). Problemet med detta tillvägagångssätt är att surroundljudet inte längre kan avkodas för ett valfritt antal högtalare.^{viii}

e) Idag kan 5.1-systemet sägas vara standard för återgivning av surroundljud. Tekniker har bland annat implementerats i DVD och HDTV. Det är sällan man som konsument hör talas om Ambisonics. Ovan har ett antal fördelar med Ambisonics-system presenterats, där en av de främsta fördelarna är att surroundljudet kan återges av ett valfritt antal högtalare. Varför har då detta system inte slagit igenom på marknaden? *Dr. Damian Murphy* nämner i sitt kompendium att en av anledningarna kan vara att tekniken olyckligtvis kopplades ihop med surroundljuds-systemet *Quadraphonics* som inte alls fungerade tillfredsställande. När Ambisonics introducerades på marknaden hade Quadraphonics precis floppat kommersiellt. Kanske var producenterna inte redo för ett nytt surroundljuds-system. Vidare backades systemet aldrig upp av några stora företag. Dessutom är det möjligt att Ambisonics lite mindre ”häftiga” tillämpningsområden kan ha spelat in. Som tidigare har nämnts marknadsfördes systemet främst för återgivning av musik, medan 5.1-system förknippas med hemmabiosystem och biografer.^{viii}

3. Bit-rate reduction of digital audio signals

a) TRUE. Lossless compression is characterized by the fact that the original bit stream can be reproduced exactly from the compressed data. The lossless codec algorithms reduce redundancy and inefficiencies in the input rather than information. It is therefore theoretically possible to cascade lossless codecs without losing quality. However, as Watkinson points out, audio compression systems are usually not completely lossless and therefore cascading compression usually leads to generation loss, especially when the codecs are not the same.^{ix x}

b) FALSE. Lossless codes do not remove *irrelevancy* from the signal. Rather, as has been explained above, they reduce *redundancy*. Lossy codes however, remove irrelevant data based on psychoacoustics. The entropy is a measurement of the quantity and randomness of the information of a data source. All data files contain a certain amount of information and, based on the probabilities of the different symbols, a certain randomness (disorder). Shannon has proved that there is a relation between entropy and the performance of the code. The best performance attainable in terms of encoded bits/symbol for any code, lossless algorithms included, is given by the entropy function. A coding algorithm can never achieve an average number of encoded bits/symbol smaller than the entropy. A good, lossless algorithm can however achieve an average number close to the entropy value. Still, it does *not* remove entropy.^{xi, xii}

c) FALSE. Transparent quality means that quality of the decoded signal is indistinguishable from the original signal.^{xiii} Some lossy compression techniques remove irrelevant data without making the changes detectable **by** the human ear by taking psychoacoustics into account.

d) TRUE (assuming that code = compression algorithm). There are several compression algorithms that utilize statistical modeling and run-length coding. The so called Lempel-Ziv approach parses the information stream into strings of symbols that occur frequently and encodes them using a variable-length code. When a certain pattern is repeated in the data sequence, the pattern is replaced by a pointer to the first occurrence of the pattern and a value that indicates the length of the pattern. The LZ algorithm is adaptive in the sense that the patterns are automatically identified as the data sequence is processed. Hence, a strongly varying audio signal would generate a lower compression factor than a relatively uniform signal.

John Watkinson discusses variable degrees of compression in PCM in his book *The art of sound reproduction*. Watkinson points out that the compression ratio in PCM is determined by the bit allocation system. It is therefore not too complicated to change the output word length parameter to achieve a different compression factor. If the parameter is changed the allocation system just repeats the process until the new block size is filled. Further the decoder can quite easily be reprogrammed to handle a change in word length. As a result the same codec can use different compression factors with the same hardware.^{ix}

e) TRUE. Most sounds have a varying spectrum where the relative signal power to noise power differs according to frequency. *Sub-band coding* splits the spectrum into several frequency bands in order to take advantage of the varying spectral energy and masking levels of the sound signal. Each band can now be quantized separately with a different quantizer and hence with different numbers of quantization levels (non-uniform quantization). As a

result, since the number of levels in the quantizer determines the power of the quantization noise and consequently also the SNR, the SNR level can be controlled, based on the power of the individual subbands.^{ix, xi} The idea is in other words to minimize the number of quantizing steps for each subband, without making the noise level exceed the masking level of the signal.

4. A/D- och D/A-omvandling

*I denna uppgift kommer fakta från John Watkinsons bok *The art of sound reproduction* om inget annat nämns.*

a) De två vanligaste teknikerna som man brukar använda för att undertrycka kvantiseringsbrus i moderna A/D-omvandlare är översampling och brusformning. De båda teknikerna används ofta i kombination med varandra, men kan också tillämpas separat.

Översampling. Översampling innebär, liksom namnet delvis antyder, att samplingsfrekvensen vid A/D-omvandling höjs betydligt över den minimifrekvens som Nyquist samplingsteorem rekommenderar. Metoden används för att komma runt vissa praktiska problem som uppstår vid A/D-omvandling. John Watkinson nämner tre huvudsakliga fördelar med översampling i sin bok^{ix}:

1. Möjliggör användandet av mindre skarpa antiviknings- och rekonstruktionsfilter
2. Möjliggör en utökad upplösning hos omvandlaren
3. I kombination med brusformning möjliggör tekniken en disproportionell förbättring av omvandlarens upplösning

Översampling fungerar på så sätt att det sprider det kvantiseringsbrus som uppstår vid A/D-omvandlingen samt signalens frekvensband över ett större frekvensområde. På detta vis kommer brusets energitäthet att minska. Genom att i ett senare steg i omvandlingsprocessen låta signalen och bruset passera ett lågpasfilter kommer den del av bruset som ligger utanför det önskvärda frekvensområdet, det vill säga spärrbandsfrekvenserna, att filtreras bort, samtidigt som det brus som finns kvar har blivit lägre. Genom att öka samplingsfrekvensen blir det därför möjligt att använda omvandlare som använder ett färre antal bitar utan att informationsmängden minskar och kvalitén försämras.

Översampling öppnar också upp för tillämpning av mindre skarpa analoga antivikningsfilter och rekonstruktionsfilter. Det mesta av filtreringen flyttas istället över till den digitala världen, efter omvandlingen, där det är lättare att åstadkomma skarpa, faslinjära filter (FIR-filter) som filtrerar bort oönskade frekvenser. I och med att bandbredden på detta sätt minskar ökar signal-brusförhållandet.^{xiv}

För att ge läsaren en uppfattning om metodens inverkan presenteras här ett exempel. Vi tänker oss att vi har en signal och ett vitt brus som sträcker sig över hela basfrekvensbandet till och med Nyquists samplingsfrekvens, $f/2$. Signalen översamlas med en faktor fyra. Om basfrekvensbandets bandbredd nu med hjälp av ett digitalt filter reduceras med samma faktor fyra, så att signalen återfår sin ursprungliga bandbredd, kommer brusets bandbredd också att reduceras med en faktor fyra. Detta leder till att brusets energimängd kommer att vara en fjärdedel av den mängd som producerades i samband med kvantiseringen. En minskning av energimängden med en fjärdedel innebär en halvering av energins nivå, vilket i sin tur leder till att signal-brusförhållandet ökar med 6 dB. 6 dB motsvarar en extra bit i kvantiseringen. På detta sätt ökar omvandlarens upplösning.

Brusformning. Exemplet som avslutade avsnittet om översampling ovan gäller endast vitt brus som har en konstant spektraltäthet. Energireduktionen blir därför mer eller mindre proportionell. För att kunna reducera så stor energimängd som möjligt kan det vara önskvärt att förskjuta så mycket som möjligt av bruset till högre frekvenser, helst så att det hamnar utanför basfrekvensbandet. Tekniken som används för att åstadkomma detta kallas brusformning. Brusformning kan endast förskjuta bruset utanför signalbandet då det kombineras med översampling. I annat fall kan tekniken användas för att exempelvis forma bruset utifrån de kunskaper som vi har om hörselsystemets ojämna känslighet för olika frekvenser. Det kan med andra ord handla om att förskjuta bruset till de frekvensband som örat inte uppfattar så bra.

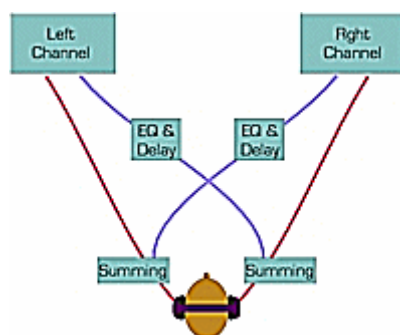
Enkel brusformning kan åstadkomma genom att i samband med kvantisering implementera felhantering som försämras vi högre frekvenser, det vill säga ju snabbare den inkommande informationen ändras. Kvantiseringsbruset kommer då att öka med frekvensen och ha sin topp vid Nyquists samplingsfrekvens, som bestäms av antalet kvantiseringssteg. Exempel på sådan felhantering är rekvantiserare där kvantiseringsbrus uppstår då bitar av lägre ordning av någon anledning tappas. Genom att lägga till dessa bitar till nästa sampel undviker man att informationen går förlorad. Tekniken fungerar bäst vid låga frekvenser. Vid höga frekvenser ökar fördröjningen och återkopplingsmekanismen blir mindre effektiv.

b) När översampling används tillsammans med brusformning uppnås således den mest effektiva reduceringen av kvantiseringsbrus. Med hjälp av översampling minskas först brusets spektraltäthet, varpå bruset genom brusformning förskjuts utanför basfrekvensbandet. Slutligen filtreras det högfrekventa kvantiseringsbruset bort med ett tidsdiskret, digitalt filter.

Brusformning i kombination med översampling användes bland annat i Philips allra första CD-spelare. I dessa spelare översamplades signalen med en faktor fyra. Översampling möjliggör, som tidigare har nämnts, att färre bitar används vid kvantiseringen. Tillsammans med brusformning medför översamplingen att bruseffekten reduceras med en faktor $1/n^2$, i detta fall $1/16$, varpå brusets nivå minskar med en faktor fyra. Samtidigt ökar signalbrusnivån med 12 dB, vilket motsvaras av två bitar. Använder vi 16 bitars pulskodmodulering från CD-skivan blir det därför möjligt att reducera antalet bitar med två och således använda en 14 bitars omvandlare med bibehållet signal-brusförhållande. För att lösa problemet med bortklippta bitar tillämpade Phillips i sina första CD-spelare den återkopplingsteknik som presenterades i avsnittet ovan. De två bitar som måste klippas bort adderas till nästkommande sampel.

5. Binauralt till varje pris

a) Den huvudsakliga skillnaden mellan stereoljud och tvåkanaligt binauralt ljud ("konsthuvudstereo") är att stereoljudet två signaler båda är avsedda att nå båda öronen medan de två binaurala signalerna separat sänds ut till respektive öra. Medan den stereofoniska ljudbilden bygger på de tids- och intensitetsskillnader som uppstår mellan de båda signalerna, är binauralt ljud helt beroende av att denna information skickas med de separata signalerna. Om ljudet spelas in binauralt, exempelvis genom att placera två mikrofoner strax *utanpå* öronen på en docka eller en människa, behöver man inte bekymra sig så mycket över detta eftersom filtreringen då redan har skett^{xv}. Om man däremot har spelat in ljudet i stereo och sedan vill konvertera ljudet till binauralt ljud, det vill säga simulera binauralt ljud, måste vissa justeringar göras. För att åstadkomma detta kan man exempelvis använda sig av en så kallad *shuffler*.^{ix} Det shufflern gör är att tillföra överhörning genom att sända en fördröjd och filtrerad kopia av respektive signal till motsatt kanal.



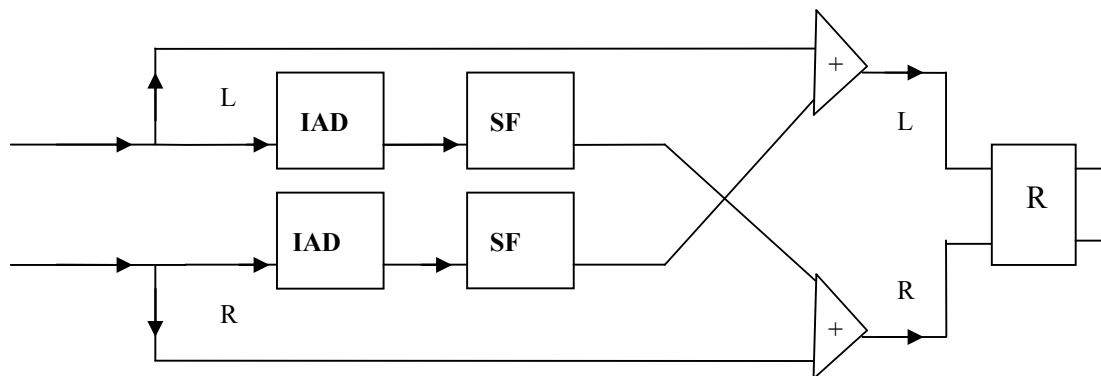
Figur 3. Simulering av naturlig överhörning

Anledningen till att signalen filtreras är att man genom att filtrera bort höga frekvenser kan simulera den filterfunktion som huvudet och övriga kroppen har. Resultatet blir en virtuell stereoljudkälla som uppstår framför lyssnaren. Det man med andra ord gör är att implementera *Huvudrelaterade transferfunktioner (HRTF)* i den simulerade ljudbilden. Det kan också tänkas att man tillämpar någon typ av (spatial) expanderare som med hjälp av genklang expanderar ljudrymden så att ljudet får en mer spatial karaktär.^{xvi, xvii}

Det finns flera metoder för att få fram lämpliga parametrar för att skapa binauralt ljud. Dels kan man utföra mätningar på en människa eller docka, dels kan man rent matematiskt beräkna parametrarna utifrån statistik och tidigare experiment. En metod som implementerar rumsakustik i det binaurala ljudet är *auralisering*, det vill säga akustiskt modellering. Idag finns avancerade datorprogram för att genomföra denna typ av modellering. I detta fall skapas en digital modell över huvudets och örats akustiska egenskaper. Utifrån denna modell genereras sedan det syntetiska, spatiala binaurala ljudet.^{xviii, xix}

b) Om jag skulle implementera approximativ stereo-till-binaural konvertering med hjälp av ett analogt mixerbord skulle kopplingen se ungefär likadan ut som den i figur 3 ovan. Grundtanken för att skapa binauralt ljud är att överföra en del av signalen i respektive kanal till motsvarande kanal för att på så vis simulera naturlig crosstalk. Jag skulle således inleda med att mata en "kopia" av vardera stereokanalen ut från mixerenheten, genom effect send-utgången, till effektenheten. I effektenheten matas signalerna först genom varsin EQ i vilken de högre frekvenserna, det vill säga diskantfrekvenserna, reduceras med ett filter. Här bör man beakta att filtreringen påverkar på den tilltänka ljudkällans position. Om ljudkällan är

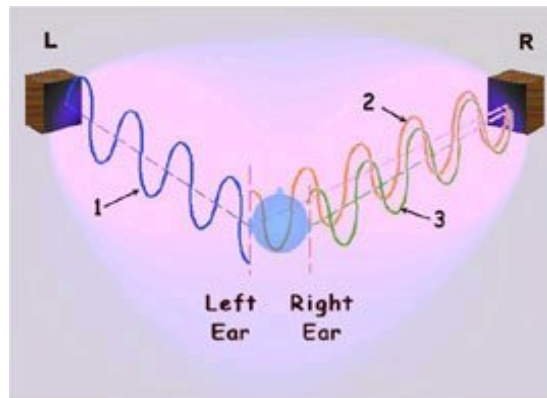
tänkt att finnas rakt framför åhöraren bör båda signalerna filtreras lika mycket. Nästa steg blir att fasförskjuta (fördröja) signalerna, eftersom det tar längre tid för ljudet från höger än från vänster att nå vänster öra och vice versa. Detta genomförs med hjälp av tidsprocessorn som kan användas för att fördröja ljudsignaler. Signalerna är därmed färdigbearbetade och överförs nu tillbaka till mixerbordet för att mixas samman med motstående signal för att på så vis skapa överhörning. En sista åtgärd är att mata det simulerade stereoljudet genom en genljuds-enhet för att skapa en känsla av rymd i ljudbilden. Alternativt kan man välja att endast göra detta på de ”kopior” som manipuleras i effektenheten. Här nedan presenteras en förenklad bild av kopplingsschemat.



IAD = Inter-aural delay
 SF = Shading filter
 R = Reverb

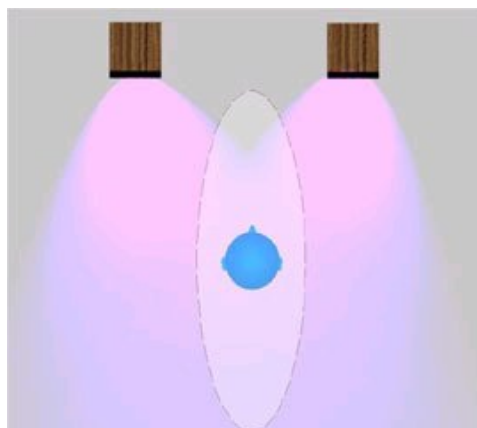
Figur 4. Shuffler-koppling^{ix}

c) När binauralt ljud anpassas för återgivning över högtalare kallas det för transauralt ljud. Den vanligaste metoden för att åstadkomma detta kallas på engelska för *crosstalk cancellation*, det vill säga utsläckning av överhörning. Problemet vid återgivning av binauralt ljud över högtalare är att ljuden från höger respektive vänster högtalare kommer att interferera med varandra (överhörning), vilket förstör den binaurala ljudbilden. Överhörningsutsläckningen isolerar de båda signalerna genom att eliminera, eller åtminstone reducera, interfererande ljudvågor. Dolby beskriver på sin hemsida hur deras virtuella högtalare tillämpar tekniken i sina virtuella högtalare. Tekniken går ut på att sända en fasförskjuten, spegelvänd kopia av ljudet från höger högtalare tillsammans med ljudet från vänster högtalare. På detta sätt kommer det ljud som når åhörarens vänstra öra från den högra högtalaren att släckas ut. På samma vis sänds en fasförskjuten, spegelvänd kopia av vänster signal ut av den högra högtalaren. Metoden är således motsatsen till syntetisk stereo som behandlas i delfrågorna a och b.



Figur 5. Utsläckning av överhörning

Det största problemet med transauralt ljud som tillämpar utsläckning av överhörning är att det är väldigt viktigt att lyssnaren befinner sig precis i mitten av det område som på engelska kallas *sweet spot*, på symmetriskt avstånd från respektive högtalare, och dessutom inte rör på huvud för mycket. Detta eftersom tekniken bygger på att ljudvågor ska släcka ut varandra precis. Åhöraren är således begränsade till en viss position.



Figur 6. En åhörare som befinner sig i *sweet spot*

Det pågår idag forskning, bland annat hos Dolby, där man försöker att öka åhörarnas rörelsemöjligheter genom att öka storleken på *sweet spot*. Dolby har delvis lyckats, men tekniken lämpar sig fortfarande bäst i situationer där åhörarna är stilla, exempelvis framför TV:n eller datorn.^{xx}

De filter som tillämpas för att skapa transauralt ljud är ofta konstruerade utifrån en förenklad modell av ett mänskligt huvud. Risken finns därför, liksom i andra virtuella ljudrymder att upplevelsen blir olika för olika åhörare. En del producenter strävar därför efter att ta fram allt mer komplexa modeller för att kunna anpassa upplevelsen efter individuella fysiska förutsättningar.^{vii}

-
- ⁱ <http://www.wired4music.com/Wired4Surround.htm> (2004-12-15)
- ⁱⁱ http://www.dolby.com/assets/pdf/tech_library/18_Metadata.Guide.pdf (2004-12-15)
- ⁱⁱⁱ <http://www.audiolab.se/aes/reports.html> (2004-12-15)
- ^{iv} http://hometheater.about.com/od/beforeyoubuy/a/surroundsound_4.htm (2004-12-15)
- ^v http://www.dolby.com/resources/tech_library/index.cfm (2004-12-14)
- ^{vi} <http://www.broadcastpapers.com/sigdis/PhilipsMCDA08.htm> (2004-12-14)
- ^{vii} Föreläsning: *You are surrounded* med Dr. Damian Murphy
- ^{viii} <http://www.ambisonic.net> (2004-12-11)
- ^{ix} *The art of sound reproduction*, John Watkinson (Focal press, 1998)
- ^x *Digital teknik i ljudproduktion*, Sveriges Radio, sida 52 (2003)
- ^{xi} *Communication Networks*, Leon-Garcia & Widjaja (McGraw-Hill, 2000)
- ^{xii} <http://www.data-compression.com/theory.html> (2004-12-13)
- ^{xiii} Lecture: *Audio coding* by Renat Vafin
- ^{xiv} <http://www.edacafe.com/books/phdThesis/Chapter-7.2.php> (2004-12-15)
- ^{xv} http://en.wikipedia.org/wiki/Binaural_recording (2004-11-12)
- ^{xvi} <http://www.headwize.com/faqs.htm#surround> (2004-12-12)
- ^{xvii} http://www.ambiophonics.org/Ch_2_ambiophonics_2nd_edit.htm (2004-12-15)
- ^{xviii} <http://www.akustikon.se/sve/software.html> (2004-12-14)
- ^{xix} <http://www.noogenesis.com/binaural/binaural.html> (2004-12-13)
- ^{xx} http://www.dolby.com/consumer/technology/virtual_speaker.html (2004-12-15)