

Varför låter skivorna så illa?

Del 1: Historisk bakgrund och analys av masteringingrepp

Allt större ingrepp i nya och gamla musikproduktioner gör numera inköp av pop- och rockskivor till en ren chansning. Svante Granqvist och LTS masteringarbetsgrupp arbetar med ett ljudanalysprogram med vilket skivköparen och välljudsvännen kan få värdefull kunskap om hur inspelningen har hanterats. I denna första artikel av två presenteras bakgrund och fem olika sätt att analysera mastrade ljudfiler.

Inom LTS arbetar vi för god ljudåtergivning och testverksamheten handlar av naturliga skäl till stor del om apparater som man har i hemanläggningen. Hemanläggningen är dock bara halva återgivningskedjan. Jag kommer att ägna två artiklar åt några fenomen på inspelningssidan som påverkar ljudkvaliteten som vi till slut får hemma i soffan.

För att beskriva problemet är det enklast att tänka sig en återgivningskedja från mikrofon, via förstärkare, AD-omvandling, eventuell processning, lagring på skiva eller fil, uppspelning via DA-omvandlare, volymkontroll, slutsteg och högtalare i ett rum. För enkelhets skull väljer jag att fokusera på de rena transmissionslänkarna från mikrofonförstärkare till slutsteg.

Enheter i kedjan har alla en brusnivå och en maxnivå. Skillnaden mellan dem kallas för signal-störförhållande eller SNR (signal-to-noise ratio). Brusnivån sätter en lägsta nivå som signalen bör ha utan att ge ett hörbart brus-tillskott. De svagaste delarna av musiksигнаlen bör vara starkare än bruset med en viss marginal. Denna marginal kallas "footroom" och mäts i decibel.

Maxnivån sätts av att apparaten börjar ge en allt kraftigare distorsion när nivån stiger. Halvledarbaserade förstärkare, AD/DA-omvandlare och ljudfiler/CD-skivor har oftast en rätt tydlig gräns över vilken distorsionen ökar plötsligt. Maxnivån är med andra ord tydligt definierad, och alla toppar i signalen som överstiger maxvärdet kommer att kapas av, eller *klippas*, som det kallas. Högtalare, mikrofoner, analoga bandspelare, vinylmediet och många rörförstärkare har en mjukare gräns vilket i stället leder till en gradvis ökande distorsion. Den typen av distorsion blir inte lika illaljudande för örat eftersom distorsionskomponenterna hamnar närmare nyttosignalen spektralt, men den gradvisa ökningen gör valet av signalnivå till en kompromiss snarare än ett enkelt val där man lägger toppnivån strax under maxnivån.

Vid val av nivå måste man se till att musiksигнаlen håller sig under maxnivån i alla leden i återgivningskedjan, i varje fall om man eftersträvar god återgivning. När signalen väl är inspelad är det relativt lätt att styra ut apparaterna fullt, men inte gå över klippgränsen. Under själva inspelningstillfället är det dock klokt att lämna lite "headroom" så att man säkert vet att AD-omvandlaren inte överstyr ifall musikerna blir lite ivriga när det är "skarpt läge".

Sett så här är ljudåtergivning enkelt; man spelar in ljudet med en mikrofon, och låter signalen passera sin förhoppningsvis transparenta utrustning utan överstyrning men på

en tillräckligt hög nivå så att bruset inte blir besvärande. I verkligheten finns förstås hela frågeställningen om hur man ska placera mikrofoner och högtalare, stereoteknik och de fall där man inte har en akustisk ursprungshändelse som ska återges, men vi lämnar det därhän i den här artikeln.

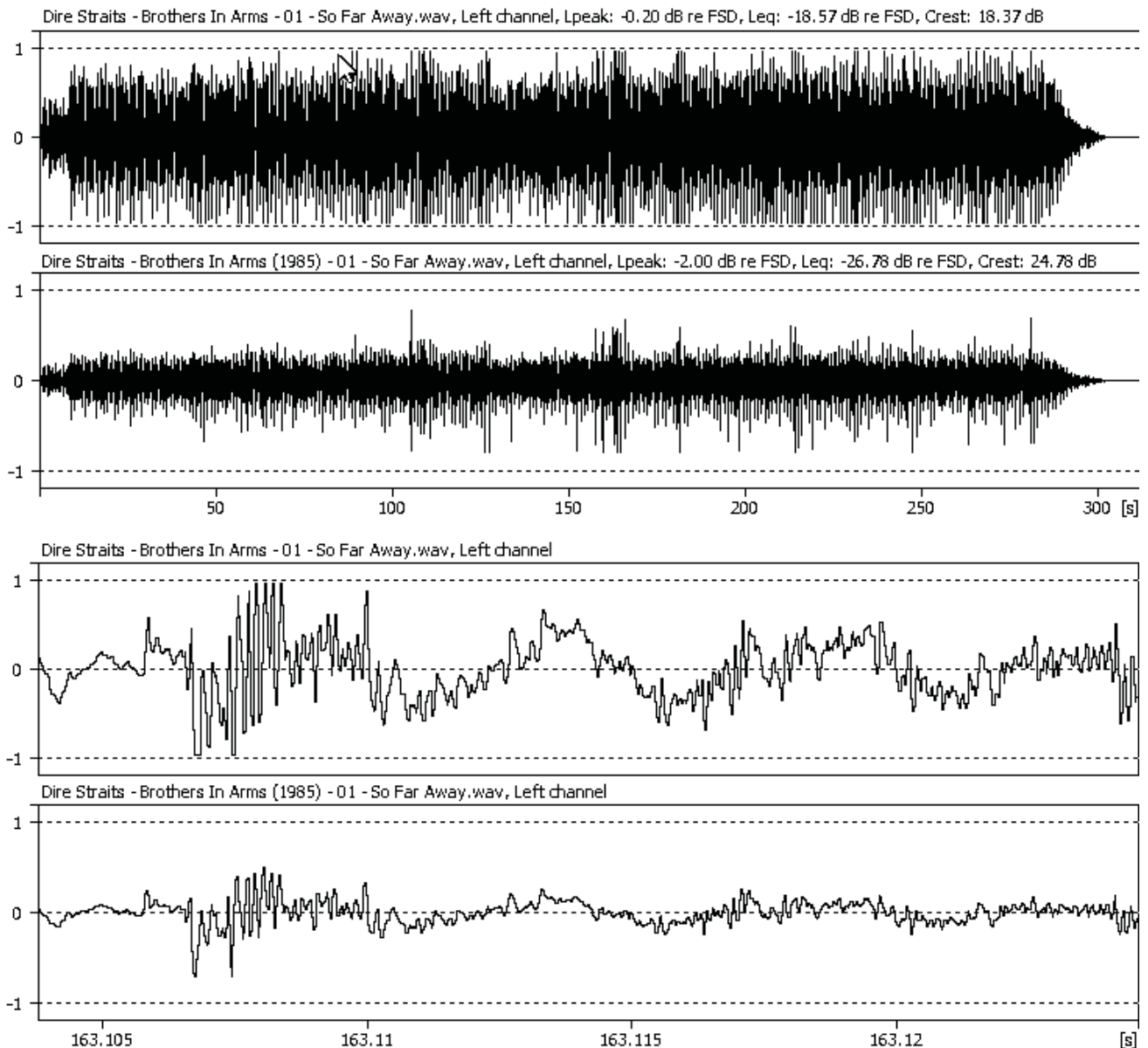
Högre medelnivå trots allt bättre medier

För 40 år sedan var det ganska noga att ställa nivåerna riktigt. Man hade då i bästa fall rullbandspelare att spela in på, med ett SNR på kanske 70 dB. Vinylmediet var sämre, med ett SNR på kanske 60 dB för en ren, knasterfri skiva. Ljudteknikerna lärde sig att utnyttja mediet maximalt, man var noga med att få full utstyrning utan allt för mycket distorsion. För att hålla bruset och knastret på tillräckligt låg nivå relativt musiken var det noga att inte spela in för svagt.

I början av 80-talet blev CD-mediet allmänt spritt, med sin tydligt definierade maximala toppnivå och ett brusgolv som ligger teoretiskt 96 dB under maximala toppnivån. Nu finns det möjlighet att lagra ljudet utan nämnvärd distorsion och med ett brus som ligger på i de flesta fall ohörbart låga nivåer. Till en början användes också mediet på just det viset. Man såg till att ha ett litet, egentligen onödigt headroom på skivorna och stornivån sattes ofta av bakgrundsbuller, mikrofoner och förstärkare snarare än mediet.

Parallellt med detta startade ett fenomen i framför allt de kommersiella radiokanalerna. Undersökningar visade att när folk växlar mellan radiokanalerna så har de högre benägenhet att "fastna" om kanalen spelar med en hög medelnivå. Det var bra för reklamintäkterna, och följaktligen började radiostationerna leta efter sätt att få det att låta starkare vilket i praktiken innebar att man behövde höja medelnivån. Medelnivån, snarare än toppnivån, bestämmer ju i hög grad hur starkt det låter. En typisk "rå" ljudmix har en toppnivå som ligger kanske 20 dB högre än medelnivån. Denna siffra kallas toppfaktor (eng. crest factor). Uppenbart är att om man lyckas sänka topparna och/eller höja nivån på musiken mellan topparna så kommer medelnivån att öka. Det kan ske genom att man låter topparna klippas eller mjukklippas, och genom att använda kompression.

Detta leder till att dynamiken i materialet reduceras och att medelnivån kan höjas. Radiostationerna i USA började med sådan processning redan på 60-talet i syfte att attrahera lyssnare. Man ska för övrigt vara noga med att skilja dy-



Figur 1. Vänster kanal av två olika utgåvor av förstaspåret på Dire Straits "Brothers in arms". Det översta spåret är en ommastrad utgåva från 1996, det näst översta är originalutgåvan från 1985. Man kan se hur "loudness race" har påverkat den senare utgåvan. Man har tagit bort det headroom på 2 dB som finns vilket är helt ok, men man har dessutom limiterat och/eller komprimerat signalen så att medelnivån ligger på -19 dB i stället för -27 dB. Medelnivån på den senare utgåvan är alltså 8 dB starkare, vilket lätt låter "bättre" vid en direkt jämförelse. Detta har fått som följd att toppfaktorn har sjunkit från 25 till 18 dB. De två undre spåren visar effekten som processningen har fått på en av topparna som verkar vara driven i klippning. Detta exempel är ganska tidigt och påverkan låter inte så hemskt illa som många mastringar av senare datum. Exemplet visar ändå att hämningarna att låta en signal gå i toppklippning har släppt till förmån för en högre medelnivå.

namikreduktion från datareduktion eller snålkodning, till exempel flac- eller mp3-kodning, som olyckligtvis ofta kallas kallas kompression även den.

Man kan spekulera i varför radiolyssnare fastnar för komprimerat ljud. Det är ju i strikt mening dålig återgivning att processa ljudet på det sättet. Det skulle kunna förklaras med FM- och AM-sändningarnas höga störnivå. Dessutom lyssnar många på radio i bilen vilket ger ytterligare bakgrundsbuller. Musiken får då en bättre förmåga att överrösta störningarna. Kanske föredrar man också en jämn nivå, det kan upplevas jobbigt om nivån varierar för myck-

et speciellt om man använder musiken som bakgrundsljud. Det kan också vara så att man lyssnar på musiken på en lägre ljudstyrkenivå än det ursprungliga ljudet och då hjälper kompressionen de svaga ljuden att hamna över hörtöskeln.

Man kan göra många sådana här spekulationer, men den troligaste anledningen till benägenheten att föredra en hög medelnivå är nog ändå att den helt enkelt låter starkare. Det är ett välkänt fenomen att om man gör en direkt jämförelse mellan två signaler och den ena är starkare, så upplevs den starkare typiskt som "bättre". Och det är ju precis

det som inträffar när lyssnaren växlar mellan stationerna, man jämför dem och den med högst nivå får då en konkurrensfördel.

Komprimering på CD-mediet

Vid 80-talets början mastrades CD-skivorna med visst headroom, men snart spred sig önskan om en hög medelnivå även hit. På radion spelades ju flera skivor efter varandra och lät det starkt så gillades låtarna mer och skivförsäljningen ökade. Att ta bort det headroom som skivorna ofta hade är okontroversiellt. Om man har en mix som är inspelad med fler än 16 bitar ökar SNR en aning och tillför ingen distorsion eftersom CD-mediet har en skarp klippgräns. Ganska snart upptäckte man dock att det inte hörs så hemskt mycket om man även låter de starkaste partierna bli lite toppklippa, speciellt inte om man mjukklipper med en limiter. Starkaste punkten inträffar ju bara på ett ställe i en inspelning och förmodligen passerar det de flesta lyssnarna utan reaktion om det blir någon enstaka klippning på en låt.

Vid det här laget började skivbolag och artister bli medvetna om att en hög ljudnivå är viktigt för skivförsäljningen och följaktligen började det som vi idag kallar "loudness race" eller "loudness war". Resan från 90-talets början fram till idag har inte varit särskilt upplyftande. Idag har en stor andel av de skivor som ges ut grav toppklippning och kraftig kompression. Toppfaktorn ligger ofta under 10 dB vilket inte sällan innebär att stunsen i trumslagen är borta och att musiken har förvandlats till en grötig smet. I en del produktioner är det så illa att man verkar ha bytt ut dimensionen svagt-starkt ljud mot rent-distorderat ljud. Man låter med andra ord ljudet vara någorlunda rent (men fullt utstyrt) under de partier som egentligen borde vara svaga, och låter sedan ljudet bli toppklippat och illaljudande för att signalera stark ljudnivå.

Det finns även andra sätt än kompression att öka den upplevda ljudstyrkan. Man kan till exempel höja diskanten. Eftersom örat är som mest känsligt i området 2-6 kHz så ger de frekvenserna "mer utdelning" i hörstyrka än andra frekvenser. På vinylskivans tid var det svårt att höja diskanten så mycket eftersom RIAA-korrekturen redan innebar en ganska kraftig diskantshöjning på mediet. Ytterligare höjning ledde till överstyrning. CD-mediet har inte den begränsningen och diskantnivån har därmed kunnat höjas ytterligare. Ibland beskär man också basen eftersom den inte höjer hörnivån nämnvärt. Detta var dock vanligt på vinylskivor.

En ond cirkel

Hela den här proceduren med att höja medelnivån hade en teknisk anledning när medierna var dåliga. Man behövde en hög medelnivå för att överrösta störningarna. Det behöver vi inte idag, utan enda anledningen att höja medelnivån är att hävda sig i konkurrensen med andra. Och eftersom alla gör det så vinner ingen på det, utan resultatet blir bara att fonogrammen låter illa, trots att vi har ett medium som tillåter så mycket bättre ljudkvalitet.

Hela orsaken till fenomenet är att mediets gräns för maximal signalnivå gäller topparna och inte medelnivån. Vore det medelnivån som var begränsad i mediet så skulle det

inte finnas någon anledning att platta till topparna. För att komma tillrätta med problemet skulle man kunna standardisera en maximal medelnivå i stället för den av CD-mediet satta maximala toppnivån på inspelningarna. Det finns flera möjliga tekniska sätt att åstadkomma detta. Inom radiodistribution pratas det bland annat om att införa en medelnivåbegränsning i distributionsledet. För att åstadkomma samma sak för hemhifi skulle det behövas ett nytt medium, eller spelare som känner av medelnivån och justerar den så att den blir lika för olika låtar, snarare än toppnivån. Vissa mjukvaruspelare har en finess som kallas "replay gain" som man kan slå på. Den mäter nivån på antingen enskilda låtar eller på hela album och anpassar nivån så att hörnivån ska vara lika stark för låtarna eller albumen. Problemet med sådana finesser är att de inte löser några problem förrän de flesta har finessen aktiverad och skivproducenterna inser att de därmed inte kan höja medelnivån med hjälp av limitering och kompression.

Det finns mycket att säga om detta, men jag nöjer mig med att konstatera att höjningen av medelnivån på dagens toppnivåbegränsade fonogram bara är av ondo. Det finns ingen anledning att ha den idag. Skivbolagen borde i stället verka för att få en överenskommelse till stånd om att inte låta medelnivån på skivan överstiga till exempel -20 dB. Det är inte skivbolagens sak att bestämma hur starkt slutanvändaren ska lyssna. Lyssnaren äger och kan hantera volymkontrollen. Däremot borde det vara skivbolagens ansvar att leverera hög ljudkvalitet.

Ovanstående behandling av inspelningar är ganska lätt att upptäcka med örat om man väl har blivit varse att den finns. En stor andel av befolkningen har dock aldrig reflekterat över den, eller har kanske ens hört något annat. Många går aldrig på helakustiska konserter och vet därför inte hur sådana låter i verkligheten. Det växer upp en hel generation människor som tror att det ska låta som det gör på dagens fonogram.

Upplysning om problemet försvåras av att vårt språk för att beskriva ljudkvaliteter är väldigt okalibrerat lyssnare emellan. Det visar sig när man undersöker vad ljudbeskrivande ord betyder för olika människor att de ofta betyder väldigt olika saker. Man kan få någorlunda god överensstämmelse för grova termer som "mörkt" eller "ljust", men så fort beskrivningarna ska vara lite mer detaljerade måste lyssnarna kalibreras mot varandra genom att lyssnarna har tillgång till samma ljudstimuli och enas om gemensamma ord för ljudkvaliteterna. Detta är ett stort problem som vi delar med till exempel röstforskning och logopedi och som gör diskussioner om ljudkvalitet svåra. Man kan dock söka stöd i den fysiska verkligheten och på så sätt få ett gemensamt, kalibrerat diskussionsunderlag.

Analys av inspelningar

Eftersom vårt språk inte riktigt räcker för att beskriva ljudkvalitet så är det svårt att förmedla sådan information i ord i till exempel en skivrecension. Det enda man med någorlunda säkerhet kan få fram med ord är om man som recensent gillar ljudkvaliteten eller inte. För att få en konstruktiv debatt om ljudkvalitet behöver den subjektiva upplevelsen få stöd i objektiva mätdata. Frågan är vad vi kan mäta och vad som skulle lämpa sig för att detektera misshandlade fonogram.

Jag har under den senaste tiden roat mig med att titta och mäta på vågformen i en ljudeditor och korrelera det med mina lyssningsintryck. En del effekter av loudness race är sålående enkla att se och mäta medan andra är svårare.

Vågform

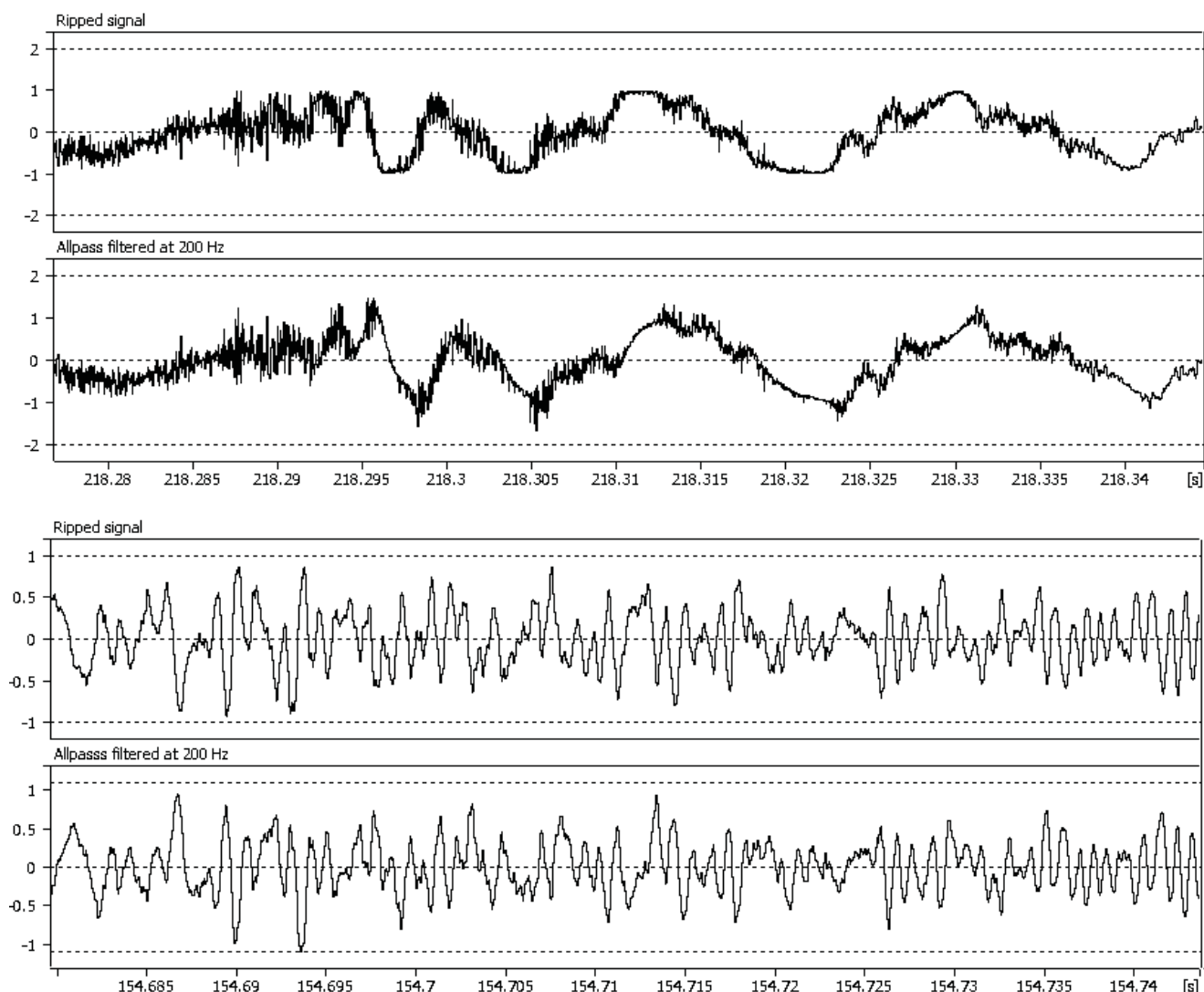
Öppnar man en med dagens mått mätt normalt mastrad inspelning i en ljudeditor så ser den ut att gå i klippning. Med vågformen utzoomad över en hel låt kan det se ut som att signalen ligger i konstant klippning. Zoomar man in lite kan det dock se ut på olika sätt. Antingen kan topparna verkligen vara klippta, eller så är de bara nivåjusterade via kompression så att de alltid precis når taket. Ingetdera låter särskilt bra. Det förstnämnda ger distorsion, och det andra tar bort musikens naturliga dynamik. Här syns också en tydlig skillnad mellan olika utgåvor av samma skiva, där de nyare remastrade versionerna nästan undantagslöst har en högre signalnivå.

Toppfaktor

Toppfaktorn anger förhållandet mellan den starkaste toppen och signalens RMS-värde. Den är enkel att mäta. Man kan mäta den för hela låten eller i kortare snuttar och på så sätt få en bild av hur den varierar under låten. Hög toppfaktor är i sig inte ett mål, det finns ljud med en naturligt låg toppfaktor, men det verkar ändå som att det är svårt att göra riktigt välljudande fonogram med en toppfaktor på under kanske 13-16 dB.

Histogram

Ett histogram talar om hur vanliga olika värden i vågformen är. I en oprocessad akustiskt inspelad signal ser histogrammet oftast ut som en topp med jämna sluttande flanker åt sidorna. Ett sådant histogram innebär att de starka värdena är ovanligare än de svaga. I nivåmaximerade inspelningar slutar histogrammet abrupt, eftersom man har



Figur 2. De översta två spåren visar en limiterad signal och samma signal som har passerat genom ett allpassfilter. När en toppklippt signal körs genom ett allpassfilter ökar toppens höjd till ett värde som är högre än klippnivån. De två understa spåren visar samma sak med material som inte är toppklippt. Signalen som inte är toppklippt behåller ungefär samma maximala topphöjd efter allpassfiltrering. Detta fenomen kan användas för att detektera toppklippning och limitering.

använt någon teknik för att så ofta som möjligt ”slå i taket”. Ibland kan man till och med se en liten topp före det abrupta fallet. Ofta är det ett tecken på toppklippning eller kraftig limitering.

Detektion av masterkompression med hjälp av allpassfilter

Att platta till toppar i tidsdomänen sänker toppfaktorn. Men det visar sig att nästan vilken ytterligare processning som helst av den tillplattade signalen gör att den återfår lite av sin ursprungliga toppfaktor. Det är inte så konstigt egentligen: om signalen optimerats mot låg toppfaktor och sedan processas ytterligare bör man ju röra sig bort från den man har optimerat emot. Detta kan utnyttjas för att detektera om signalen har utsatts för toppfaktorminskning. Om man applicerar någon form av klangneutral processning och toppfaktorn ökar, kan det ses som ett tecken på att ursprungssignalen var toppfaktorsänkt. En form av processning som är nästan helt neutral för signalens klangliga egenskaper är fasvridning. Genom att utsätta materialet för fasvridning kan man alltså få en metod som detekterar om toppfaktorsänkande processning har applicerats. Fasvridningen kan ordnas med ett allpassfilter. Ett sådant filter fasvrider signalens olika frekvenser olika mycket, men påverkar inte deras amplitud.

Fördelen med den här metoden jämfört med en ren toppfaktordetektion är att det är svårt att hävda att skillnaden i toppfaktor mellan signalen på skivan och den allpassade versionen beror på något annat än att man har tagit ner toppfaktorn i syfte att höja medelnivån. Man kan däremot komma undan med att hävda att ljud med en låg toppfaktor är en del av ett konstnärligt uttryck.

Spektrum

Det är betydligt svårare att skapa en bild av om man har applicerat tonkurveförändringar på materialet. Visserligen är det lätt att mäta signalens frekvensinnehåll med spektralanalys, men eftersom man inte vet hur originalets frekvensinnehåll var så går det inte att säga mycket om processningen från ett enkelt spektrum. Spektrum talar om hur

klangbalansen är i låten, men även om det säger något om hur det låter det är svårt att dra några slutsatser om spektrumets utseende kommer från musikinstrument, spelstil, inspelningsrum eller från ingrepp vid inspelning/mix/mastering. Tillsammans med genreinformation kan dock spektrum ibland avslöja om materialet är basbeskuret eller har utsatts för andra kraftiga tonkurveingrepp. Olika utgåvor av samma inspelning kan också jämföras med varandra spektralt för att påvisa skillnader mellan dem.

Vad kan inte analyseras enkelt?

Ovanstående parametrar är ganska basala och fungerar för att detektera det värsta skräpet. Det finns dock en hel del kvalitetsskillnader mellan högkvalitativa fonogram som är betydligt svårare att mäta sig fram till. Till det räknas stereobild, rumsinformation, mikrofonposition, och små spektrala skillnader.

Nu är inte de svårigheterna något skäl att låta bli att mäta det man kan. Det är bättre med en diskussion med en svag verklighetsförankring än med en som helt saknar förankring. Man måste börja någonstans, och det finns gott om dåliga mastringar som förtjänar att uppmärksammas. Här fungerar även de enklaste signalanalysmetoderna väl.

Datorprogram för analys

Jag har tillsammans med LTS mastringsgrupp tagit fram ett datorprogram som utför ungefär de analyser som jag skissar på ovan. Programmet är avsett att förmedla information till en tredje part, på ett sätt som alltid blir likadant. Det finns inga inställningar att ändra, så om man ser en bild från programmet så kan man förstå vad den betyder och man vet att informationen inte är vinklad.

Avsikten är att programmet ska kunna laddas ner helt fritt från LTS hemsida och publicering av de mätprotokoll som programmet genererar ska uppmuntras i alla former. Det blir mer om detta i ett kommande nummer av MoL.

Svante Granqvist



Ljudtekniska Sällskapet

Box 112

184 22 Åkersberga

Hemsida:

www.lts.a.se

