

PERCEPTION OCH LJUDMODELLER

*Utdrag från*

**Metoder i syntes och signalbehandling**

(hovedoppgave i musikkvitenskap, universitetet i Oslo 2001)

av Risto Holopainen.

pdf-version: februari 2007

*”Écoutez les sons autour de vous, ils ne programment, ni ne calculent, et cependant le Grand Ordinateur, dans toute sa gloire, n’a jamais chanté comme l’un d’eux.”*

Michel Chion,  
i ett uttalande tillskrivet läromästaren Pierre Schaeffer

*”We may in fact argue that the truly potent texts of our times are certainly not texts like this book, or even true scientific theories, but computer programs themselves.”*

Trevor Wishart

## Sammandrag av uppsatsen

Vilka problem stöter man på när man försöker utveckla en syntesteknik eller metod för ljudprocessering? En algoritm genererar eller processerar en signal som kan avlyssnas eller visualiseras. Genom att variera kontrollparametrarna kan man variera de genererade ljuden, och en av målsättningarna har varit just att komma underfund med hur olika inställningar av parametrarna motsvarar upplevda egenskaper i ljudet. Ibland varierar man hela den generativa algoritmens arkitektur eller modell, och inte bara parametrarna. Valet av modell bestäms både av den typ av ljud man vill åstadkomma, och av vilken typ av kontroll man vill ha över olika aspekter av ljudet.

De tekniker som har diskuterats har till stor del undersökts utifrån en experimentell attityd: Vilken typ av ljud kan man generera med den här algoritmen, och varför låter de som de gör? Hur påverkar kontrollparametrarna ljudet, och finns det något sätt att göra kontrollparametrarna perceptuellt och musikaliskt mera relevanta? Det är inte alltid ett triviellt problem att ställa upp korrelationer mellan en syntestekniks kontrollparametrar och ljudets egenskaper. Att förutsäga resultatet av en syntesteknik är ofta lättare än att förklara varför det blir på ett visst sätt. Föreningen kommer med erfarenheten av att iaktta kopplingen mellan konfigurationer av kontrollparametrar och klingande resultat. Icke desto mindre är det nyttigt att försöka förklara varför en syntesteknik fungerar som den gör, vilket har varit en målsättning i beskrivningen av teknikerna i kapitel 4 och 5. Relationen mellan ljud och upplevd kvalitet är också bitvis komplicerad, och kan beskrivas som en förvrängning. Vi har lånat Schaeffers begrepp *anamorphose* för att beteckna detta förhållande. Men uppmärksamheten är utan tvivel också viktig för hur och vad vi hör, och har en stor betydelse för urskiljandet av olika aspekter av ljud. Att utforska eller utfråga ljud är en central del i arbete med ljudsyntes. Man kan kalla det en hermeneutisk aktivitet, där man växlar mellan att variera ljudets kontrollparametrar och att lyssna på det för att urskilja nya nyanser. Denna process har inget naturligt slut, utan det finns alltid mer att höra efter. Tekniskt och reducerat lyssnande är oundgängligt vid arbete med elektroakustisk musik, men kan bli så närgånget att man för en stund blir blind för viktiga drag i ljudet.

En gammal indelning av syntestekniker är additiv, subtraktiv och ickelineär syntes eller modulation. Additiv syntes kan via fouriertransformen rekonstruera vilket ljud som helst, och kan därför sägas vara den mest generella formen av syntes. Interferenssyntes är en form av additiv syntes, där deltonerna är placerade på exponentiellt ökande avstånd. Denna synnerligen enkla modell ger upphov till ett oväntat komplext beteende<sup>1</sup> med svävningar, amplitudmodulation och glissando.

Ickelineär syntes innefattar bland annat FM och waveshaping. Vi har sett exempel på flera varianter av FM. En felaktigt implementerad oscillator visade sig ha samma funktion som att successivt öka modulationsindexet. Återkopplade FM-modeller kan uppvisa kaotiskt beteende. Kontrollsignaler utgör en förbindelselänk mellan syntesens lägsta nivåer och mera storskaliga förlopp som är musikaliskt relevanta, till exempel kontroll av styrkegrad, tonhöjd och vibrato. För att göra syntesteknikerna mera spelbara är det önskvärt att kunna styra perceptuellt relevanta drag i ljuden genom kontrollfunktioner. Ett användbart och flexibelt gränssnitt för att styra kontrollfunktioner bör möjliggöra kontroll både på flera hierarkiska tidsskalor och av olika aspekter av ljudet samtidigt.

Ljudprocessering är ett mycket mera öppet område än syntes, eftersom resultatet beror både på vilket ljud man använder, och vilken processering man utsätter det för. Den primära jämförelsen i ljudprocessering är den mellan det ursprungliga ljudet och det resulterande. Många tekniker har tydliga artefakter, som genast avslöjar dem för en införstådd lyssnare. Dessa kännemärken kan lika

---

<sup>1</sup> De märkliga fenomenen uppkommer när avståndet mellan deltonerna är några få cent.

gärna vara eftersökta som önskade. I en situation där man exempelvis vill sträcka ett ljud, kan man välja mellan en mängd metoder, men man bör ta hänsyn till de respektive teknikernas artefakter, eftersom de sätter sin tydliga prägel på ljudet.

För ljudprocessering består systemet av två delar: själva algoritmen för processering, och mängden av alla ljud. Samverkan mellan en förhoppningsvis välkänd algoritm och ett okänt ljud kan bjuda på stora överraskningar, och i princip skulle man kunna hävda att man inte vet vad en processeringsteknik gör, förrän man har provat den på alla tänkbara ljud. I praktiken är det inte nödvändigt, eftersom syftet inte är att veta exakt på förhand vad resultatet kommer att bli. Det räcker ofta med att ha en föräning om resultatet, och en idé om hur man kan variera betingelserna för att successivt närma sig ett önskat resultat.

Vi har sett exempel på amplitudkontrollerad interpolation och decimering, som transponerar och sträcker eller krymper ljuden på sätt som beror på deras momentana amplitud. I likhet med de vågcykelbaserade metoderna involverar de amplitudkontrollerade en primitiv analys av ljudet, vilket gör att man kan få mycket olika resultat beroende på vilket ljud man utgår ifrån. Eftersom de ickelineära filtren inte kan dekomponeras i en lineär filterdel och en ickelineär distorsionsdel, kan de inte undersökas genom traditionell filterteori. De flesta varianterna har dock visat sig vara relativt intuitiva och tacksamma att använda i musikaliska sammanhang. Av de tekniker som opererar på flera samples åt gången, är de granulära både flexibla och enkla att handskas med. Vågcykelbaserade tekniker ger många ovanliga möjligheter, men de sätter också sin tydliga prägel på ljudet. Cellautomater i förbindelse med ljudsyntes är föga utforskade, förutom Karplus-Strong-algoritmen som efterliknar knäppta strängar. Cellautomater kan klassificeras som en kaotisk syntesmodell, men de kan också processera korta ljudfragment.

Jaffes kriterier för evaluering av syntestekniker kan delas in i tre grupper. Den första gruppen kriterier handlar om parametrarnas användbarhet: Hur intuitiva och välartade är de? Har parameterförändringar en perceptuell effekt, och är parametrarna relaterade till fysiska attribut hos instrumentet? Den andra gruppen kriterier har att göra med egenskaper hos de resulterande ljuden: Finns det ett hörbart släktskap mellan ljuden som kan genereras av en viss modell? Kan alla typer av ljud produceras? Finns det analysverktyg? Den sista gruppen kriterier handlar om effektivitet och implementation. För att algoritmen ska fungera i realtid måste effektiviteten tas hänsyn till. Systemet som helhet kan inte förstås enbart utifrån vad man vet om dess delar var för sig. Det konstaterade vi både när det gällde ickelineära filter och i förbindelse med återkopplade FM-modeller. Synergi gäller det oförutsägbara samarbetet mellan delar av ett system; konfluens gäller den förutsägbara samverkan där parametrar är utbytbara med andra.

Även om en syntesmetod kan betraktas som ett instrument i den meningen att den har ett fixerat dispositiv av möjligheter, gäller inte denna inskränkning för den elektroakustiska musikens praxis i sin helhet, eftersom man i princip har alla tänkbara ljudkällor, syntesmetoder och processeringar att tillgå. I den mån man kan kalla en dator ett musikinstrument, är det ett instrument utan egentliga begränsningar. De eventuella begränsningarna beror i så fall på användaren. Den skenbart självklara frågan om hur verktygen påverkar resultaten är långt ifrån självklar, om man betänker att det står komponisten fritt att välja och bygga sina verktyg själv. Utveckling av syntesmodeller har liknats vid instrumentbygge. På denna hantverksnivå tycks det olämpligt att tala om estetik, eftersom estetiska teorier handlar om konst och konstverk. Ändå är det frestande att fråga sig om ett instrument eller ljudverktyg i sig har några estetiska implikationer. Vilken typ av musik lägger instrumentet till rätta för?

Den akusmatiska musiken har bidragit med en oerhörd differentiering på det morfologiska och spektrala planet. Analogiprincipen går ut på att behandla olika musikaliska dimensioner som tonhöjd, varaktighet, timbre, dynamik och rumslig position som helt likvärdiga. I elektroakustisk komposition kan analogiprincipen föras till sin extrem, eftersom det finns praktiskt taget obegränsade möjligheter att ställa upp musikaliska dimensioner som kan kontrolleras minutiöst.

Vi har sett vad det kan innebära att tillverka sina egna verktyg på en tämligen låg nivå – C-programmering. Det är krävande, men ger också en stor frihet till att konstruera originella

instrument. Och det finns mycket kvar att undersöka inom syntes och signalbehandling. Eftersom programmen ska vara flexibla nog att fungera med olika input, kan programmets komplexitet bli för stor för att man ska kunna kontrollera korrektheten i alla tänkbara situationer. När slutprodukten, en ljudfil, är avsedd att ingå i ett stycke musik, är det dock bara hur det låter som verkligen räknas. Men det är alltid önskvärt att programmet faktiskt gör vad det ska. Programmets korrekthet är inte främst viktig för att generera intressanta ljud, utan för att skapa användarvänliga program som gör vad de utger sig för att göra.

De stora kraftfulla texterna av idag är inte böcker som denna, utan datorprogram, som Wishart konstaterar i det inledande citatet. Det har han säkert rätt i, men samtidigt kan det vara klokt att ta till sig Chions uppmaning att lyssna på ljuden i omgivningen. De varken programmerar eller kalkylerar, men ändå har den Stora Datorn i all sin ära aldrig sjungit som de.

\*\*\*

*Följande utdrag är en något redigerad version av det som ursprungligen utgjorde kapitel 3 av Metoder i syntes och signalbehandling från 2001.*

# PERCEPTION OCH LJUDMODELLER

Hur kommer det sig att vi uppfattar ljud på olika sätt, vad sker i våra öron och hjärnor? Det är en av sakerna som kommer att belysas i första avsnittet av detta kapitel. Även om mer och mer börjar bli känt om hur sinnesintryck kodas i nervsignaler och vad hjärnan gör med dem, är själva övergången från dessa signaler till medveten erfarenhet fortfarande en djup gåta. Om man inte förmår förklara upplevelsernas uppkomst i egentlig mening, så kan man i varje fall ställa upp korrelationer mellan ljud och upplevd kvalitet. Denna relation är bitvis komplicerad, och kan beskrivas som en förvrängning. Vi kommer i avsnitt 2 att låna Pierre Schaeffers begrepp *anamorphose* för att beteckna detta förhållande, och beskriva några viktiga relationer mellan ljudsignal och ljudupplevelse. Ett andra tema är lyssningen som sådan. Vi mottar inte bara ljud passivt, utan vi riktar uppmärksamheten mot olika aspekter av dem allt efter omständigheterna och våra intressen. Även här kommer huvudsakligen Schaeffers idéer till vår hjälp, men vi kommer också att diskutera mekanismer för gruppering av ljud, som studeras i auditory scene analysis (avsnitt 3). Ett problem som är aktuellt när man försöker tala om ljud är hur man beskriver dem, hur de kan klassificeras. De olika försök jag har samlat under rubriken "ljudmodeller" (avsnitt 4) angriper detta på olika sätt. Detta är ett omfattande kapitel, och det kan kanske se ut som ett sidospår från temat i boken. De resultat som presenteras här, har förvisso mycket mera generell tillämpning, men de är ändå helt centrala för arbete med ljudsyntes och signalbehandling inom musik, eftersom lyssning är en så viktig informationskälla för att diagnostisera syntes- och processeringsmodeller. Lyssning ses inte som någonting subjektivt och ovetenskapligt — utan precis lika objektivt som den information man kan dra ur en visuell representation av ett ljud — förutsatt att man tar hänsyn till anamorfoserna mellan akustisk signal och perception. Man kan med fördel tänka på experimenteringen med syntes och ljudprocessering som informella perceptionspsykologiska studier: Det rör sig i båda fallen om att finna övergången från ett känt, kontrollerat fysiskt fenomen, bestämt av kontrollparametrarna och syntesmodellen, till perception. Många av de resultat som har visats i vetenskapliga studier, kan lätt efterprövas av vem som helst med tillgång till en dator med något ljudsyntesprogram. (Vissa försök kräver dock flera försökspersoner eller lite mer avancerade förberedelser.)

## 1. Från akustisk signal till nervsignal

För att förstå hur det kommer sig att vi uppfattar en given akustisk signal på ett visst sätt, behöver vi veta vad som sker med signalen i alla överföringsled. Alla ljud vi uppfattar är vibrationer i luft (eller eventuellt vatten och andra medier) som vi uppfångar med öronen, eller genom benledning. Sin egen röst hör man bara delvis genom ytteröronen, en stor del av ljudet fortplantar sig direkt genom skallbenen till innerörat. Om det rör sig om ljud förmedlade via högtalare, eventuellt inspelade och bearbetade, eller syntetiserade, behöver man också veta något om vad som sker med signalen i alla dessa stadier; hur högtalaren utstrålar ljud och vad som sker med det i olika slags rum. I all elektroakustisk omvandling kan det uppstå distorsion av olika slag (Sundberg 1989 s. 199): Den harmoniska distorsionen genererar heltalsmultiplar av den ingående frekvensen, dvs. harmoniska övertoner, som ju ändå brukar förekomma i många ljud. Därför är det kanske den minst störande förvrängningen. Harmonisk distorsion kan också utnyttjas som en syntesteknik i form av vågformning (*waveshaping*). Differensstonsdistorsionen skapar undertoner och övertoner till två ingående frekvenskomponenter enligt sambanden

$$(1) \quad f_n = n f_1^{(n-1)} f_2 \quad \text{och}$$

$$(2) \quad f_n = n f_2^{(n-1)} f_1, \quad n = 1, 2, 3, \dots$$

för undertoner respektive övertoner. Sådana kombinationstoner uppstår också i örat. Intermodulation ger kombinationstoner vid

$$(3) \quad f_n = f_2 \pm n f_1, n=0, 1, 2, \dots$$

där  $f_1$  är en lägre frekvens än  $f_2$ . Enligt Yates (1995 s. 69) kan man om två sinustoner presenteras samtidigt för ett öra, uppmäta en tredje komponent i den yttre hörselgången vid frekvensen  $2f_1 - f_2$ , kallad «the cubic intermodulation frequency», på grund av att den komponenten uppstår om överföringsfunktionen är ett tredjegrads (kubiskt) polynom. Det skulle alltså motsvara differensstonsdistorsion (ekv. 2) med  $n = 2$ .

Det är ett slående faktum att man finner liknande förvrängningar i den elektroakustiska omvandlingen, dvs. i mikrofoner, högtalare och förstärkare, som i örat. Rummet ljudet färdas igenom innan det når örat har också en inverkan. Ljudintensiteten avtar med kvadraten på avståndet till ljudkällan, alltså 6 dB per avståndsfördubbling. På stora avstånd inverkar dessutom luftabsorptionen. Högfrekventa ljud dämpas med ca. 1 dB per 100 m vid 10 kHz, och ju högre frekvens, desto större dämpning. Vidare bidrar alla reflexioner från fasta ytor ljudet stöter emot till ekon och efterklang, vilket kan sudda ut detaljer och göra ljudet diffust. Mjukare ytor absorberar däremot ljud, framför allt i de höga frekvenserna. Stående vågor, som uppstår i de flesta rum, gör att toner med vissa frekvenser kan ha mycket olika intensitet på olika platser i rummet. Detta är några faktorer som gör att framföringar av elektroakustisk musik faktiskt inte är lika från gång till gång, om de äger rum i olika lokaler. Det är också en grund till att man i många perceptionspsykologiska försök använder hörlurar för att kunna kontrollera vilka signaler som faktiskt når lyssnarens öron.

Följande mycket summariska genomgång av signalens gång från örat till hörselbarken bygger mest på Wallin (1982, kapitel 1-2) och Sundberg (1989, kapitel 3). I ytterörat uppstår en resonans vid 3 - 4 kHz, som förstärker signalen i det området. Dessutom kan formen på öronen leda till något olika filtrering av ljud från olika riktningar, vilket kommer till användning som en ledtråd för riktningssuppskattning. I trumhinnan sker det en kompression av starka ljud, som utlöses av en muskelreflex. Men vid plötsliga starka ljud hinner inte den mekanismen med, vilket kan leda till hörselskada. I örats känsligaste område, omkring 4 kHz, kan man uppfatta en vibration på  $10^{-9}$  cm, något mindre än en vätemolekyls diameter. Då är trumhinnans utslag lika stort som luftpartiklarnas, men vid starkare ljud reduceras dess relativa amplitud. I mellanörat sker ytterligare en kompression. Hörselbenen (hammaren, som står i förbindelse med trumhinnan, städet, och stigbygeln som står i förbindelse med det ovala fönstret) verkar som en serie hävstänger som vid svaga ljud kan ge en förstärkning på upp till 3 gånger. Dessutom har trumhinnan en 15 – 30 gånger större yta än ovala fönstret, vilket kan förstärka tryckvariationerna med upp till 90 gånger vid svaga ljud (Sundberg 1989 s. 49). Det ovala fönstret är en öppning mot hörsnäckan, som är fylld med en lättflytande vätska. Om stigbygeln fotplatta som står i kontakt med ovala fönstret svänger fram och tillbaka uppkommer en deformation av basilarmembranet i form av en vandrande våg. Men vid långsamma tryckvariationer pressas vätskan igenom en trång passage, helicotrema, och får det runda fönstret som också gränsar mellan innerörat och mellanörat att bukta ut.

Olika positioner på basilarmembranet ger störst utslag vid olika frekvenser, längst ut vid basen de högsta frekvenserna och längst in vid spetsen de lägsta. Basilarmembranet är tunnast (0,1 mm) vid basen, men samtidigt tjockast, och bredast (0,5 mm) i spetsen. Detta gör membranet styvast längst ut, vilket påverkar den vandrande vågens utformning. På basilarmembranet sitter det cortiska organet med hårceller (stereocilier). När basilarmembranet deformeras kommer dessa sinneshår i kontakt med en benlist, tectorismembranet, vilket utlöser cellernas aktivitet. Nervcellerna svarar med en hel serie nervimpulser på ett stimuli från receptorcellen som svarar mot intensiteten i stimuli, från några få per sekund till flera hundra. Alla nervimpulser har samma amplitud, så intensiteten kodas därför genom antalet impulser per tidsenhet (frekvenskodning).

Nervbanan från örats receptorer via olika kärnor till hörselbarken är tonotopiskt organiserad,

vilket innebär att närbelägna receptorer har cellbanor som leder till närbelägna celler i alla stadier fram till hjärnbarken, där de är ordnade i isofrekventa band. Signalen från det cortiska organet går till cochleariskärnorna, och vidare till det främre olivkomplexet, där signalerna från vänster och höger öra kombineras. Via ytterligare kärnor når signalen hjärnbarken. Denna anatomiska organisation gör det möjligt att avgöra om vissa perceptuella attribut har sitt upphov i den perifera eller centrala delen av nervsystemet. Ett exempel är tonhöjdsuppfattning i grundtonslösa harmoniska spektra. En ton sammansatt av tredje och fjärde deltonen ur ett harmoniskt spektrum ger en tonhöjdsförnimmelse på den frånvarande grundtonen, till och med om de presenteras för varsitt öra (det fungerar dock inte på låg volym, och ger ingen speciellt påfallande effekt). Därför måste tonhöjdsupplevelsen i sådana ljud uppstå på en central nivå, det vill säga efter att signalerna från vänster och höger öra har kombinerats i olivkomplexet. Långt ifrån alla detaljer är ännu kända om hur nervimpulser tolkas som tonhöjd. Men man har funnit att olika områden av hörselbarken behandlar ljud av olika komplexitet och innehåll, exempelvis fonem eller ackord (Tervaniemi 1999).

## 2. Anamorfoser

Det är inget enkelt samband som råder mellan ljudet i form av fysiska vibrationer och dess uppfattade kvaliteter. Tvärt om utmärks detta förhållande av icke-linearitet. Det är vad Schaeffer fångar in med sitt begrepp '*anamorphose*': «Au sens propre, le terme anamorphose se rapporte à la déformation que subit dans un miroir courbe l'image d'un objet par rapport à cet objet. Nous l'utilisons ici dans un sens figuré, pour désigner certaines "irregularités" remarquables, dans le passage de la vibration physique au son perçu, faisant penser à une espèce de déformation psychologique de la "réalité" physique, et dont nous verrons qu'elles traduisent simplement l'irréductibilité de la perception à la mesure physique.» (Schaeffer 1977, fotnot s. 216) Schaeffer liknar alltså övergången från fysisk signal till perception vid de deformationer som uppkommer i en skev och buktande spegel, och talar hellre om korrelationer mellan dessa domäner än direkta korrespondenser. Man kan skilja på två slags "anmärkningsvärda skevheter", nämligen, för det första de tämligen regelbundna, som där en exponentiell variation av en fysisk kvantitet motsvaras av en lineär variation av perceptionen, såsom är fallet med frekvens och tonhöjd, eller amplitud och uppfattad intensitet. (Vi ska nyansera dessa påståenden senare.) För det andra finns det trösklar som när de passerar ger upphov till vitt skilda perceptioner. En sådan är integrationströskeln<sup>2</sup> vid ca 20 Hz. Över den kan vi uppfatta en tonhöjd i en periodisk signal, men under den hör vi bara ett periodiskt ljud (om det har komponenter i det hörbara området). Det finns också många trösklar för hur lång tid något måste pågå för att en viss typ av perception ska uppstå. Alla upplevda egenskaper är emergenta, och kräver en viss utsträckning i tid för att uppstå.

Benjamin Libet har gjort några studier som har blivit kända för sina märkliga resultat; förenklat uttryckt, tyder de på att medvetandet är försenat med upp till en halv sekund. Så här beskrivs det av Wallin (1982 s. 183):

«Ett subduralt ("under den hårda hjärnhinnan") stimulus till den primära sensoriska hjärnbarken (SI) ger ett input, som är annorlunda beträffande bana och mönster än vad som genereras av ett perifert sensoriskt stimulus. Libet och medarbetare lyckades utarbeta en metod, där medvetna sensoriska erfarenheter med "naturliga" sensoriska kvaliteter skapas genom en noggrann reglering av stimulus-parametrar (intensitet, pulsfrekvens och varaktigheten för pulsgrupper eller "pulstrains" till värden, som ligger nära tröskelvärdena).

Undersökningarna visade på två tidsmässiga egenskaper hos de processer, vilka leder till medveten sensorisk erfarenhet: (a) för att ett adekvat tillstånd, som möjliggör en medveten sensorisk upplevelse, skall kunna uppkomma, krävs en viss minimiperiod för aktivering av de neuronala

<sup>2</sup> 'Integrationströskeln' är kanske inte en vedertagen term på svenska. Den är lånad direkt från franskans "seuil d'integration."

systemen, dvs en avsevärd försening (upp till 0.5 s), innan neuronal tillräcklighet uppnåtts; (b) en mekanism, vilken retroaktivt återför den subjektiva tidssättningen ("timingen") av erfarenheten till en tidpunkt, som ligger nära början av den sensoriska signalen; eller annorlunda uttryckt: då den neuronala tillräckligheten uppnåtts, en subjektiv hänvisningspunkt bakåt i tiden, som fördaterar upplevelsen så att den svarar mot tidpunkten för primära kortikala svar på specifika afferenta projektionssignaler.»

Även om det vid direkt stimulering av områden i hjärnan kunde behövas upp till 0.5 sekunder vid tröskeln för de svagaste uppfattbara känselintrycken, uppstod motsvarande upplevelse redan av den kortaste beröring av huden. Libet har också påpekat några implikationer av sina försök, vilka Wallin (ibid, s. 190) refererar: «En nödvändig avsevärd period av neuronal aktivitet kunde fungera som ett "filter" genom att hindra en mängd pågående sensoriska inputs från att nå medveten nivå, under förutsättning att dessa endast utlöser kortvariga aktiveringar av hjärnbarken.» Vidare är det tänkbart att fördröjningen av den perception som håller på att bildas gör det möjligt för andra processer i hjärnan – inhiberande eller förstärkande – att påverka den resulterande perceptionen. En annan punkt är vilka konsekvenser det får för viljestyrda handlingar: «En väsentlig försening av den verkliga produktionen av en medveten sensorisk upplevelse, således inte dess subjektiva tidssättning, kunde innebära att snabba beteendesvar, t.o.m. sammansatta och ändamålsenliga, skulle kunna uppstå omedvetet; detta kan begränsa urvalet av handlingssätt, i vilka man kan utöva ett fritt viljestyrt val.» (ibid, s. 191.)

En studie av Tervaniemi (1999) visar att en del förmedveten (pre-attentive) processering äger rum vid tonhöjdsuppfattning bland annat av grundtonslösa spektrum. Individer med bättre resultat på ett musikalitetstest visade sig ha högre utslag på en nervimpuls kallad MMN (mismatch negativity), som utlöses av ljud med egenskaper som avviker från andra ljud intill. Ljud som är lätta att skilja utlöser MMN med större amplitud och kortare latenstid än ljud som är mera lika varann. En bättre förmedveten processering, som var vanlig hos de musikaliska försökspersonerna, leder också till bättre förmåga att medvetet urskilja sådant som tonhöjd eller ordningsföljd i en tonsekvens.

Men huvudsaken i det här sammanhanget är att perceptionen byggs upp över ett tidsintervall, och att ljudets kvaliteter är emergenta egenskaper. Egenskapernas emergens är dock inte bara ett psykologiskt fenomen; osäkerhetsrelationen styr hur precist man kan bestämma en frekvens på basis av ett begränsat utsnitt av en signal. Vi ska nu se på några områden där perceptionen i större eller mindre grad präglas av anamorfoser.

## 2.1 Maskering

Cochleans funktion liknas gärna vid en korttids-fourieranalys: sammansatta akustiska signaler delas upp i sina komponenter genom att den vandrande vågen i basilmembranet når sin topp på ett avgränsat område som svarar mot en viss frekvens. Via hårcellerna utlöser det nervimpulser i de nervtrådar som är mottagliga för den aktuella frekvensen vid en viss amplitud. Varje nervfiber har en kurva för tröskelvärde som är V-formad, med en bästa frekvens där den är som mest känslig. På båda sidor om bästa frekvensen måste man alltså öka amplituden för att det ska ge något utslag. I tillägg till denna cochloetopiska organisationen eller "position-kodningen" kan frekvenser också kodas genom nervimpulsernas timing. Impulserna sammanträffar i så fall företrädesvis med en viss fas av vågformen, men det gäller mest för låga frekvenser, och för högre frekvenser avtar den funktionen. Detta fenomen, som kallas faskoppling, tycks kunna förklara vissa interferenstoner som uppkommer i örat, utan att de är närvarande i ljudets spektrum, närmare bestämt differenstoner i harmoniska spektrum med frånvarande grundton (Wallin 1982 s. 95).

Örat har en viss grad av frekvensupplösning, som avgör om två frekvenskomponenter tydligt låter sig urskiljas och vilken dissonansgrad eller strävhet de har (sensorisk dissonans, till skillnad från musikteoretisk). Maskering kan användas för att studera hörselns frekvensupplösningens förmåga:



«Masking may be regarded as reflecting the limits of frequency analysis [= upplösning]. If a sound of a given frequency is masked by another sound with a different frequency, then the auditory system has failed to resolve the two sounds. Hence, by measuring when one sound is just masked by another, it is possible to characterize the frequency analysis capabilities of the auditory system.» (Moore 1995 s. 162) Ett sätt att mäta maskeringen av en sinuston, är att kombinera den med bandpassfiltrerat brus, centrerat kring tonen. Medan man successivt ökar brusets bandbredd, mäter man signalens tröskelvärde (i dB) där den just kan uppfattas. Till en början ökar signalens tröskelvärde när man ökar brusets bandbredd, men efterhand flackar kurvan ut, och ytterligare ökning av bandbredden har ingen signifikant effekt på tröskelvärdet.

För att förklara detta fenomen tänker man sig det hörbara frekvensregistret täckt av en serie överlappande bandpassfilter. När man ökar brusets bandbredd, passerar mera av bruset genom ett av dessa filter, så länge som brusets bandbredd inte överstiger filtrets. Men ytterligare ökning av brusets bandbredd ökar inte mängden brus som passerar filtret. Den bandbredd där signalens tröskelvärde upphör att öka kallas ett kritiskt band.

Ett annat sätt att mäta maskering är att hålla det maskerande bruset konstant och variera signalens frekvens. På så sätt får man en kurva som beskriver hur mycket olika frekvenser maskeras av detta brus. Ett viktigt resultat som framkommer av dessa försök är att maskeringens avtagande är mindre brant för högre belägna frekvenser än för lägre, och denna skillnad blir mera markant om man ökar det maskerande brusets amplitud. Det kritiska bandets storlek varierar också med frekvens. Enligt en ofta citerad gammal uppskattning är ett kritiskt band 100 Hz för frekvenser upp till 500 Hz, och därifrån en liten ters. Moore (1995 s. 175) ger en uppdaterad formel där ett kritiskt band har storleken

$$\text{kritiskt band} = 24,7 (4,37F + 1) \text{ Hz,}$$

där F är centerfrekvensen i kHz. En sådan enhet motsvarar också det konstanta avståndet 0.89 mm på basilmembranet. Effekten av maskering är inte alltid så lätt att förutsäga utifrån kända resultat. Det kan till exempel hända att kombinationen av två maskerande ljud ger upphov till något större grad av maskering än vad man hade fått om man bara summerar deras effekt. Ett ljud kan också maskera en kort signal som kommer högst 100-200 ms efteråt, s.k. framlänges maskering. Baklänges maskering kan också förekomma, där det maskerande ljudet uppträder efter en kort signal.

## 2.2 Frekvens och tonhöjd: sinustoner

Tonhöjdsförnimmelser kan uppstå på flera sätt. Vi ska först se några exempel på sinustonernas förhållande till tonhöjd, och därefter komplexa toner. I försök med sinustoner har man låtit försökspersoner justera en ton så att den låter dubbelt eller hälften så hög som en referenston. På så sätt har man fått fram den så kallade melskalan, där 1000 Hz har värdet 1000 mel, och en ton som upplevs som dubbelt så hög 2000 mel. Det märkliga är att melskalan inte är en lineär funktion av frekvens, så en oktav motsvarar inte alltid en fördubbling i mel. Däremot finns det en direkt koppling mellan melskalan och kritiska band, eftersom tonhöjd uppmätt på detta sätt motsvarar bestämda positioner på basilmembranet (Houtsma 1995; Sundberg 1989). Melskalan ses sällan använd i praktiken, vilket kan ha flera förklaringar. Musiker är mera vana vid Hz och relativa mått som cent och kromatiska toner. En annan sak är att melskalan endast gäller för sinustoner, och skulle man åstadkomma en liknande skala för andra typer av ljud, så skulle man förmodligen vara tvungen att utföra en hel serie nya försök för varje slags ljud.

Tonhöjden hos sinustoner kan också variera med intensitet. Låga toner, under 1000 Hz, tycks sjunka i tonhöjd om man ökar deras amplitud, medan höga toner, över 2000 Hz, ökar i tonhöjd. För korta toner (mindre än 40 ms) tycks det alltid gälla att ökning av intensiteten sänker

tonhöjden, oavsett frekvens (Houtsma 1995, s. 270).

Den minsta uppfattbara förändringen i tonhöjd kallas differenslimen (DL). Ett sätt att mäta den är att frekvensmodulera en sinuston i låg hastighet och registrera när modulationen blir märkbar. Vid 100 Hz är differenslimen 2.5 - 3.5 %, men för mycket svaga toner är den större. Vid 1000 Hz är DL 0.5 - 1 %. Om man undersöker DL med hjälp av en tonsekvens med näraliggande toner, är det möjligt att urskilja upp till en tiondel så små intervall som vid FM (Sundin 1989 s. 60-61). Tonlängden påverkar också DL, så att den ökar för korta toner. Om man i stället för sinustoner använder komplexa toner, förbättras också urskiljningsförmågan. Sinustoner förekommer ju knappast alls i isolerat tillstånd bland naturliga ljud, och det är viktigt att komma ihåg att resultat baserade på undersökningar med sinustoner inte alltid kan överföras på komplexa toner.

## 2.3 Komplexa toner

I sammansatta toner behöver inte deltonen med grundtonens frekvens vara närvarande i ljudets spektrum för att grundtonen ska kunna uppfattas. En förklaring på detta är att toner med frånvarande grundton fortfarande är periodiska med en period som motsvarar grundtonen, vilket uppfångas av neuronernas faskoppling.

Differenslimen för toner med frånvarande grundton varierar beroende på om alla deltoner ligger så tätt att alltid flera av dem faller i samma kritiska band, eller om örat förmår upplösa vissa av dem. Deltoner med lägre nummer är vidare separerade på en logaritmisk skala, och kan lättare urskiljas. Toner bestående av lägre deltoner ger en liten DL, medan toner med deltoner från nummer 12 och uppåt, som faller innanför samma kritiska band, har en större DL. I det senare fallet har fasförhållandet mellan komponenterna också en klar effekt på DL. Bäst urskiljningsförmåga får man om vågformen är impulsliknande, som när alla deltoner är i sinusfas. Om man justerar fasförhållandena så att man minimerar vågformens impulskaraktär, får man en högre differenslimen (Houtsma 1995 s. 277).

I vissa fall kan tonhöjden av komplexa toner vara tvetydig. Om inte alla deltoner fusionerar till ett holistiskt intryck, en gestalt om man så vill, så kan de höras isolerat som fristående sinustoner. Även om en ton fusionerar kan man med en viss möda skilja ut dess lägsta deltoner. En ton som har konstant stora avstånd, till exempel oktaver mellan deltonerna, tenderar att fusionera. Men om man utsätter hela tonkomplexet för långsam frekvensmodulation, kan man få det att fusionera. Det är ett exempel på *common fate*, dvs. att en synkroniserad variation i någon aspekt i ljudet kan få tidigare separata element att smälta samman till en perception (Se Bregman 1990).

Det är bara syntetiska ljud som kan vara helt periodiska (om än inte i strikt matematisk mening, eftersom detta skulle innebära att de pågick i all evighet), naturliga ljud är antingen kvasiperiodiska eller aperiodiska. En möjlighet att åstadkomma en sådan kvasiperiodicitet är att deltonerna inte står exakt i förhållandet 1 : 2 : 3 : 4 : ... till varandra. Pianotoner har till exempel något sträckta intervall mellan deltonerna. Både sträckning och komprimering påverkar den uppfattade grundtonen. Med sträckta deltonsavstånd höjs grundtonen, men bara till en viss gräns, innan deltonerna upphör att fusionera. Motsvarande effekt uppnås om man bara stämmer om en av deltonerna.

Man kan också filtrera vitt brus eller andra komplexa ljud som saknar tydlig tonhöjd genom ett kamfilter, för att påföra dem en tonhöjd. Frekvensresponsens tydliga toppar motsvarande samtliga harmoniska deltoner innanför nyquistbandet förklarar den tydliga tonhöjden i kamfiltrerade ljud. Mera anmärkningsvärt är kanske att bandpassfiltrerat brus kan få en tonhöjd att framträda i övergången från passband till stoppband, förutsatt att övergången är tillräckligt skarp. Ringmodulerat brus har en ganska tydlig tonhöjd motsvarande modulationsfrekvensen. För amplitudmodulerat brus är möjligheterna att urskilja tonhöjder ganska goda upp till 100 Hz, varifrån de avtar. Enligt Houtsma (1995 s. 284-285) är det amplitudmodulerade brusets långtidsspektrum konstant (utan toppar), så det finns anledning att tro att sådana tonhöjder uppfångas av faskoppling.

För harmoniska toner begränsas antalet deltoner vi kan uppfatta av hur högt grundtonen ligger. I basen kan en ton med grundton 50 Hz ha flera hundra deltoner inom hörbart register. Toner i högt register kan däremot bara ha ett fåtal hörbara deltoner. Det gör att man i princip borde få flera möjliga klangliga variationer av låga toner än av höga. En möjlig invändning mot det resonemanget, är att olika slags mikrovariation i deltonernas frekvens och amplitud kan vara en nog så viktig faktor för att skapa klanglig identitet för ett ljud.

## 2.4 Amplitud och styrkegrad

Vad menas med amplitud? Det finns några möjliga varianter att välja mellan. Man kan tala om vågformens momentana amplitud, alltså amplituden i en given tidpunkt. Men vanligtvis betecknar amplituden en vågforms högsta topp. Ett vanligt mått är RMS-amplitud (*root mean square*) som fås genom att ta kvadratroten av genomsnittsvärdet under en period  $T$  av vågformen i kvadrat:

$$RMS = \sqrt{\frac{1}{T} \int_T (x(t))^2 dt}$$

RMS-amplituden är lämplig om man ska jämföra amplituden av två vågformer. En sinusfunktion med toppamplituden 1 har till exempel RMS-amplituden  $1/\sqrt{2}$  och en rektangulär vågform som antar värdena 1, -1, har RMS-amplitud 1. Ljudstyrka anges vanligen i decibell, som är ett relativt mått på samma sätt som cent är det för tonhöjd. Genom att specificera en referensnivå kan man göra dB till ett absolut mått, som dB SPL (sound pressure level). Om  $p$  är RMS-amplituden för tryckvariationerna i luften, och  $p_0$  motsvarar tryckvariationerna i det svagaste uppfattbara ljudet vid 1 kHz, så kan man beräkna dB enligt sambandet  $\text{dB SPL} = 20 \log(p/p_0)$ .

Den uppfattade hörnivån varierar med frekvensen, på ett sätt som brukar illustreras med Fletcher-Munson-diagram som visar kurvor för lika uppfattad hörnivå som funktion av frekvens. Hörnivån mäts i enheten phon, som visar vilken ljudstyrka som ska till för att få sinustoner av olika frekvenser att låta lika starka. Vid 1 kHz motsvarar phonskalan per definition dB-skalan. Hörtröskeln varierar betraktligt över hela frekvensområdet, känsligast är örat vid 2-4 kHz, och för de lägsta frekvenserna i basen är den uppemot 70 dB högre. Mot högre ljudnivåer flackar kurvorna för lika hörnivå ut något. En viktig konsekvens av de branta kurvorna för lika uppfattad hörnivå under 200 Hz är att andelen bas i en inspelning tycks öka om man spelar upp den på högre volym, och på svag volym kan basen helt försvinna.

Xenakis (1992) har presenterat idén att dela in det hörbara området i ett raster, så att en förflyttning mellan intilliggande rutor längs frekvensaxeln motsvarar nätt och jämt uppfattbara tonhöjdskillnader, och förflyttningar längs amplitudaxeln ger nätt och jämt uppfattbara nivåskillnader. Detta raster transformeras sedan till ett rätlinjigt rutnät, på vilket olika kompositoriska processer definieras. Det är ett intressant sätt att inkorporera en hänsyn till hörselns icke-lineariteter redan i ett prekompositoriskt stadium.

Även om phonskalan visar hur många dB som behövs för att två toner ska låta lika starka, säger den ingenting om vilken ökning som ska till för att en ton ska låta dubbelt så stark. För detta ändamål har man infört sonskalan, som är besläktad med melskalan, i det avseendet att de båda bygger på uppskattningar av fördubblingar och halveringar. Hörnivån 1 son definieras som hörstyrkan av en sinuston vid 1 kHz med nivån 40 dB. En ton som låter dubbelt så stark som denna får värdet 2 son. Över 40 dB behövs det ungefär 10 dB ökning för att fördubbla antalet son, men under 40 dB räcker det med mindre.

## 2.5 Tidsanamorfoser

Urskiljningsförmågan för tonhöjd minskar om man minskar tonlängden. Vid riktigt korta ljud uppstår ingen uppfattbar tonhöjd, utan ljudet blir till ett klick. Men till och med när ljudet består av en enda period av en sinuston, kan man uppfatta en mer eller mindre ljus eller dov klangfärg beroende på frekvensen.

Hörtröskeln för korta ljud beror både på amplituden och ljudets längd. För att förklara tröskelns tidsberoende har man infört en teoretisk modell av hörseln som en ackumulator, eller en konvolution mellan insignalen och en viktfunktion som väger tidigare värden av insignalen. En del olika funktioner har föreslagits, men en vanlig modell är att viktfunktionen är en exponentiell kurva som har sitt maximum i tidpunkten motsvarande nuet, och som avtar mot det förflutna. Om man plottar signalens tröskelvärde i dB som en funktion av dess längd på en logaritmisk tidsskala, får man punkter som ligger på en linje med lutningen ungefär  $-3/4$ . Detta gäller för tonlängder från ca 5 perioder av tonen till 500 ms, och det gäller både för sinustoner över praktiskt taget hela det hörbara registret (i varje fall från 250 till 14000 Hz) och för brus av motsvarande längd (Eddins & Green 1995).

Studier av ackumulationseffekter kan påvisa hur långa tidsintervall örat kan samla eller integrera information över. Å andra sidan kan man fråga sig hur snabbt örat är, vilket kan studeras på flera sätt. Man kan till exempel se på hur bra vi kan avgöra ordningsföljden av två eller flera ljud. Med två toner av olika frekvens, båda ca 500 ms långa, där den ena börjar strax före den andra, har en tröskel på 20 - 30 ms uppmätts för att ge en riktig bedömning av vilken som kommer först (vilket definieras som att försökspersonen ger rätt svar i minst 75 % av fallen; om de bara gissar på måfå kan de inte ha rätt i mer än 50 % av fallen, i genomsnitt). Samma tröskelvärde, 20 - 30 ms, har också uppmätts för att bedöma vilket av två ljud som tänds först, eller ordningsföljden mellan ett ljud och ett ljud (ibid s. 218).

En annan typ av experiment går ut på att höra ordningsföljden i en upprepad sekvens av flera ljud. I sådana försök är prestationerna betydligt sämre: det kan behövas upp till 200 ms långa ljud för att korrekt bedöma ordningsföljden i en upprepad sekvens med fyra ljud, som inte är tal eller musikaliskt material. Eddins & Green (ibid s. 219) förklarar det så här: «It appears that these repeating sequences produce problems for the listener in coding auditory events, rather than taxing the inertial aspects of the auditory system. Listeners can discriminate among the different sequences, but have difficulty in mapping the perceived sequence to the required output code.» Men de nämner också exempel på att en försöksperson efter omfattande övning var i stånd till att ange ordningsföljden i fyratonssekvenser med tonlängder ända ner till 2 ms.

Bregman (1990) refererar många liknande försök, som visar att det är svårare att avgöra ordningsföljden i tonsekvenser som segregeras eller splittras upp i två självständiga skikt. (Ett välkänt exempel på segregering är barockens skenpolyfoni i enstämmig musik.) Varför är det så svårt att korrekt uppfatta ordningsföljden i upprepade tonsekvenser? Jag skulle tro att det beror på att minnet i någon mening överlastas av repetitionen. Det är nämligen ofta mycket lättare att uppfatta ordningsföljden om sekvensen spelas upp endast en gång, eller flera gånger med pauser emellan. En annan indikation på hörselns snabbhet är förmågan att upptäcka avbrott mellan två ljud. I en studie refererad av Eddins & Green (s. 219) har lyssnare klarat av att höra skillnad på två ljud bestående av två rektangulära impulser, med eller utan en lucka emellan. Pulserna var 10  $\mu$ s vardera, och lyssnarna kunde identifiera avbrott så korta som 5 - 10  $\mu$ s (en mikrosekund,  $\mu$ s, motsvarar  $10^{-6}$  sekund). Men även om man kan höra skillnad på extremt korta ljud med eller utan avbrott, betyder inte det nödvändigtvis att det är ljudens temporal utformning det kommer an på. Det är snarare skillnader i frekvensspektrumet som är ledtråden här. Mera pålitliga resultat får man därför om man ställer lyssnaren inför uppgiften att skilja mellan två ljud med olika temporal utformning, men samma frekvensspektrum. Det enklaste är att välja en sekvens som till exempel ett starkt klick åtföljt av ett svagare, och jämföra det med samma sekvens avspelas baklänges. Att vända ljudet baklänges påverkar nämligen inte frekvensspektrumet, sett under hela tidsintervallet,

utan bara fasspektrumet.

När det gäller identifikation av olika instrument, vet man att attacken spelar en stor roll för många instrument. Om man klipper bort den kan det bli näst intill omöjligt att identifiera instrumentet. Det gäller i olika hög grad för olika instrument; spektrala fluktuationer inne i den kvasiperiodiska delen av tonen kan också vara avslöjande. Vissa instrument, som cembalo och säckpipa, har också en karakteristisk tonavslutning. Hos instrument med ljudförloppet attack-resonans (vibrafon, till exempel) tenderar attackens klang att färga av sig på resten av tonen, så att den tycks vara närvarande hela tiden. Klipper man bort attacken och lyssnar på återstoden, är det ett helt nytt ljud som framträder. När Schaeffer höll på att experimentera med sådana klipp, märkte han också att en kort ansats spelad för sig själv kunde uppfattas som lika lång som dess efterklang, även om den var upp till 12 gånger så lång. Det är ett exempel på vad Schaeffer kallar 'anamorphose temps-durée'; ljudets innehåll är helt avgörande för hur långt vi uppfattar det som. «La durée musical est fonction directe de la densité d'information.» (Schaeffer 1977 s. 248.) Det är inte svårt att underbygga det påståendet, och även om Schaeffer inte vill precisera vad som menas med informationstäthet, är det uppenbart att attacken innehåller mycket information, eftersom den ensam ofta är tillräcklig för att identifiera ljudet, medan resonansen är fattigare på information. Ett mera drastiskt exempel på tidsanamorfos är att vända ett attack-resonansljud baklänges. Här är informationstätheten mera spridd över hela ljudet: ett ständigt exponentiellt crescendo och gradvis inträdande av nya deltoner fångar hela tiden intresset. Lyssnandet kan också bli mera abstrakt i det att identifikationen av ljudkällan sätts ur spel. Det är, som Schaeffer uttrycker det, som att dra en akusmatisk slöja över ljudet. Men han varnar också för det naturstridiga med ljud som växer explosionsartat såsom baklänges spelade pianotoner. Efter mer än 50 år med elektroakustisk musik är baklängesljud kanske inte så förvånande längre, men man kommer inte ifrån att det inte existerar någon känd fysisk process som åstadkommer sådana ljud. Möjligtvis med undantag av att det till en viss grad går att imitera sådana baklängesljud med instrument som tillåter kontinuerlig kontroll av styrkegraden.

En annan aspekt av tidsanamorfoser är att man inte utan vidare kan sträcka eller komprimera ett ljuds längd utan att samtidigt påverka vissa andra egenskaper. Det gäller bland annat för transienter, som när de sträcks ut kan förlora sin karaktär fullständigt, speciellt om man använder en fasvocoder med långt fönster. Det gäller likaså för fluktuationer som är avsedda att försiggå i ett visst tempo. Typiska exempel är vibrato, tremolo, portamenti mellan toner och olika slags granularitet som kan vara betecknande och ge ljudet en speciell textur. Problemet är inte begränsat till transformation av ljud, utan är lika aktuellt i syntes. Ofta kan man vilja analysera ett ljud för att skapa en syntesmodell som tillåter oavhängiga förändringar av ljudets aspekter som vibratohastighet och tonlängd.

## 2.6 Spatialt hörande

*"For with what right affirm of a faint sound that it is a less faint made fainter by farness and not a true faint near at hand? Or of a faint fading to fainter that it recedes and not in situ decreases."*  
(Beckett: Company)

Två faktorer är speciellt viktiga för positionsbestämningen av ljud. Den ena är tidsskillnader i signalens ankomsttid till vänster respektive höger öra. Den andra är skillnader i ljudnivå mellan öronen. Med en ljudkälla rakt till höger om lyssnaren har signalen ca 20 cm längre väg till det vänstra örat, vilket fördröjer signalen med ca 0.6 ms (avståndet mellan öronen delat med ljudets hastighet i luft). Samtidigt skuggar huvudet de högre frekvenserna så att de inte tar sig runt, eller försvagas.

Med hörlurar är det lätt att kontrollera signalens tids- och nivåskillnader, och uppmäta effekterna. Men man får också ett ljud som tycks vara placerat inne i huvudet, mer eller mindre

förskjutet i höger-vänsterled. Med sinustoner är tidsskillnader ekvivalenta med fasskillnader. Från 0° till 60° fasskillnad påverkas ljudets position praktiskt taget lineärt i riktning mot det öra som ligger före. Mellan 60° och 120° fortsätter positionen att förflyttas, men med en mindre brant kurva. Omkring 180° är ledtråden tvetydig, och positionen kan vippa mellan höger och vänster. Positionsbestämning på grundval av tidsskillnader fungerar bäst för låga frekvenser, över ca 1.5 kHz försvinner effekten (Grantham 1995 s. 299). Positionsbestämning genom nivåskillnaden mellan öronen fungerar däremot bättre för höga frekvenser. Så små förändringar av balansen som 0.5 - 1 dB kan räcka för att upptäcka en förändring av ljudets position.

Tids- och positionsskillnaderna samverkar vanligtvis i naturliga sammanhang. Men med hörlurar kan man konstruera kombinationer av tids- och nivåskillnader som inte skulle uppkomma naturligt av en ljudkälla i ett rum. Om de samverkar, kan deras effekt summeras, och om de motverkar varann kan de upphäva effekten, eller ge en långt mindre entydig bild av ljudets position (ibid s. 302).

Det som har sagts tidigare om maskering har gällt när samma signal går till båda öronen, men om man låter det maskerande ljudet (brus t.ex.) vara samma i båda öronen, medan man inför en fasskillnad i signalen, kan man sänka dess maskerade tröskelvärde. Störst verkan har det om signalen presenteras i motfas (180°). Vid 200 - 500 Hz kan signalens tröskelvärde i så fall sänkas med upp till 15 dB i förhållande till om signalen är i medfas (ibid s. 304). Det säger något om hur mycket skillnader i spatial position bidrar till förmågan att urskilja samtidiga ljud.

Förutom tids- och nivåskillnader använder vi oss av spektrala ledtrådar som uppkommer genom att ytteröronen filtrerar ljudet olika beroende på dess riktning. (Det är lätt gjort att konstatera att ljud som kommer bakifrån har en dovre klang än de som kommer framifrån.) Denna ledtråd utnyttjas främst för att lokalisera ljud i det vertikala planet, men också i viss mån i det horisontala planet för att skilja mellan bakom och framför. Men det är bara för högfrekventa ljud den spektrala ledtråden fungerar. Positionsbedömningen är mest precis för ljud i det horisontala planet framför lyssnaren. Hittills har vi gjort det tysta antagandet att lyssnaren inte rör på sig. Men det är helt uppenbart att vi vänder oss mot en ljudkälla för att bättre kunna lokalisera den, och även små huvudrörelser torde kunna förbättra lokaliseringsförmågan.

Avståndsbedömningen fungerar lite annorlunda än rikttningsbedömningen. Den tar följande ledtrådar till hjälp:

- Ljudnivå: Amplituden är omvänt proportionell mot kvadraten av avståndet, förutsatt att ljudet inte reflekteras på vägen från ljudkällan till mottagaren.
- Proportionen av direktljud och rumklang: ju större andel direktljud, desto närmare tycks ljudkällan vara.
- Ljudets spektrum: Luftabsorptionen försvagar de högre frekvenserna, så att avlägsna ljud får en mattare klang än ljud i närheten.

I en del fall är absolut avståndsbedömning möjlig, så att man kan peka ut en plats i rummet ljudet tycks komma ifrån. Ibland är bara en relativ avståndsbedömning möjlig, dvs. man kan bara ange vilket av två ljud som verkar vara närmast. Spektrala skillnader, till exempel, duger bara till relativa bedömningar om de andra ledtrådarna är satta ur spel. Avståndsbedömningen är alltid enklare för bekanta ljud än för helt okända. Det är ingen risk för att förväxla en person som ropar på 50 meters avstånd med en som talar lågmält på en meters avstånd, även om den faktiska ljudnivån skulle vara den samma. Här använder vi vår erfarenhet av röstens spektrum i olika röstlägen i kombination med vår erfarenhet av hur ljud dämpas över avstånd. Med obekanta ljud som spelas upp från olika avstånd, och justeras så att ljudnivån alltid blir den samma hos lyssnaren, är det svårare att bedöma avstånd. Men om man går i riktning mot (eller ifrån) ljudkällan, så kan amplitudförändringen ge upplysningar om ljudets position (Grantham 1995 s. 326).

I rum med efterklang är det lättare att bedöma avstånd än i rum med helt torr akustik, även för obekanta ljud, vilket är helt uppenbart eftersom proportionen av direktljud och reflexer används för avståndsbedömning. Om ljudet reflekteras i rummet, når det lyssnaren vid flera tidpunkter: först direktljudet, och därefter primära och sekundära reflexioner. Men ljudkällans uppfattade position

bestäms av direktljudet. Precedenseffekten, som den kallas, innebär att reflexer som är fördröjda mellan 1 och 30 ms i förhållande till direktljudet ignoreras. Med längre fördröjning av den första reflexen kan den bli urskiljbar, men gränsen där precedenseffekten upphör varierar för olika typer av ljud (ibid s. 334).

Det finns en hel del litteratur om hur man kan simulera olika spatiala effekter med högtalare, t.ex. R. F. Moore (1990). De flesta syntes- och signalbehandlingsmodeller som kommer att beskrivas i kapitel 4 och 5 opererar med monofiler. De få undantagen är inte i första hand avsedda att åstadkomma specifika rumsliga placeringar av ljudbilden, utan detta uppkommer i så fall som en sidoeffekt. Men vi behöver ändå förstå hur olikheter i vänster och höger kanal påverkar ljudbilden. En viktig aspekt i elektroakustisk musik är ljudbildens grad av spatial kompaktitet eller öppenhet. Om man bara mixar samman ett antal monoljudfiler och ger dem lite olika panorering, så får man en kompakt ljudbild, och om man hör på det i hörlurar, tycks ljudet befinna sig inne i huvudet. Genom dekorrelation (Kendall 1995) kan man få ljudbilden öppnare och diffusare. Två signaler som är i samma fas har korrelationen 1, och är de i motfas har de korrelationen -1. Dekorrelerade signaler har en korrelation närmare 0. Genom att filtrera en signal genom två filter med flat frekvensrespons, men med olika fasspektrum, får man två signaler som har samma korrelation som de två filtrens impulsrespons. För att få korrelation 0 kan man välja två orelaterade sekvenser av slumpstal till impulsrespons. Kendall konstaterar några effekter dekorrelation har på ljudbilden: Man får diffusa ljudbilder, som från sena reflexioner i en konsertlokal. I hörlurar får man en externaliserad ljudbild, dvs. ljudet tycks komma från någonstans utanför huvudet. Precedenseffekten elimineras, och ljudbilden blir inte beroende av lyssnarens position i förhållande till högtalaren. Dekorrelation är en vanlig effekt, som kan uppstå på många sätt. Det räcker för det mesta med att signalen i vänster och höger kanal är likartade, men inte identiska. Helt orelaterade signaler bör rimligtvis också vara dekorrelerade, eller kanske man hellre borde säga okorrelerade, men det är en annan sak.

Spatialisering av elektroakustisk musik finns det mycket att säga om, men det går utanför ramarna för denna bok. Men Smalleys uppdelning i *composed space* och *listening space* bör nämnas (Smalley 1997). Det handlar om skillnaden mellan den illusoriska rumslighet en komponist kan skapa i ett stycke, och det faktiska rummet där musiken blir framförd. Saker som låter förträffligt i studion behöver inte alltid göra det när stycket spelas upp i ett annat rum.

### 3. Två teorier om lyssning

Både Pierre Schaeffer och Albert S. Bregman har behandlat frågor som har med lyssning och perception att göra; de rör sig delvis inom samma område, men med ganska olika förutsättningar och mål. När Schaeffer skrev sitt *Traité des Objets Musicaux* i 1966 hade forskningen på perceptionspsykologi inte kommit speciellt långt, och till stor del är det en polemik mot andra komponister som i större eller mindre grad ignorerade den mänskliga perceptionen i sin teori eller praktik. Nu, när många av Schaeffers idéer har blivit mera allmänt accepterade, har också perceptionsforskningen kommit längre. Men man kan inte låta bli att undra hur den hade sett ut om de hade haft kännedom om Schaeffers idéer.

En parallell mellan Schaeffers *objet sonore* och Bregmans *auditory stream* är att de båda hänvisar till gestaltpsykologin för att ska beskriva dem. Gestaltpsykologerna studerade främst grupperingsmekanismerna för visuella stimuli, både rörliga och orörliga. Efter att ha råkat i vanrykte några decennier började deras idéer åter dyka upp på olika håll. Man träffar på dem inte bara hos Schaeffer och Bregman, men också hos konstteoretikern Rudolf Arnheim och Lerdahl & Jackendoff, som uppenbart dragit stor nytta av gestaltpsykologin i sin generativa musikteori. Den försöker i några regler sammanfatta hur en införstådd lyssnare grupperar metrisk struktur och fraser, huvudsakligen i wiener-klassicistisk musik. En annan likhet mellan Schaeffer och Bregman är att de båda använder sig av blackboxmodeller av perceptionsprocessen; Schaeffer i störst grad, medan Bregman ibland kan inlåta sig på att diskutera neurologiska förklaringar till vissa fenomen. Medan

Schaeffer förespråkar ett fenomenologiskt anlagt musikaliskt forskningsprogram, lutar Bregman mest åt funktionalistiska förklaringar.

### 3.1 Schaeffer om lyssning

En av de viktigaste sidorna av *Traité des Objets Musicaux* är onekligen Schaeffers skarpsinniga analyser av lyssning. Många av Schaeffers idéer och begrepp som reducerad lyssning och ljudobjekt har uppnått berömdhet ryktesvägen. Men eftersom *Traité* hittills inte har blivit översatt till engelska (en spansk översättning lär visst finnas), har dessa idéer inte fått den spridning de förtjänar. Jag har tidigare berört Schaeffers lyssningsmodell tillsammans med andra betraktelser över lyssning (Holopainen 1998).

Eftersom vi alltid befinner oss i en miljö där det förekommer något slags ljud, finns det alltid något att höra. Det gäller även i fullkomligt ljudisolerade rum, där man, som John Cage gärna påpekade, kan höra ljuden av sin egen blodcirkulation och nervsystemet i arbete. Men plötsligt hör vi någon knacka på dörren, eller en bok falla ner från bokhyllan, och vi är inte sena med att peka ut riktningen ljudet kom ifrån eller vad som förorsakade det. Sådana plötsliga ljud upplyser oss om viktiga händelser i omvärlden, och påkallar genast uppmärksamheten. Men vi kan också lyssna efter särskilda detaljer i samtidigheten av olika ljud från olika ljudkällor. Alltefter vår intention kan vi rikta uppmärksamheten mot vilken aspekt som helst av ljudet. Om vi pratar med någon, eller lyssnar på någon som talar, är det dock helt naturligt att uppmärksamheten är fokuserad på meningen i samtalet snarare än röstklängen. Dessa vardagliga situationer kan stå som exempel på Schaeffers fyra lyssningsfunktioner (se tabell).

4. <i>comprendre</i> (förstå) tecken, begrepp	1. <i>écouter</i> (höra på) indicier på händelser	<b>Objektiv</b>
3. <i>entendre</i> (lyssna) kvalifikationer av perceptionen	2. <i>ouïr</i> (åhöra) rå perception	<b>Subjektiv</b>
<b>Abstrakt</b>	<b>Konkret</b>	

De fyra verben som beskriver de olika lyssningsaktiviteterna används i en speciell mening som kan avvika från deras dagliga betydelse, både på franska och i de översättningar jag har valt.

- 1) *Écouter*, översatt med höra på, innebär att höra efter indicier på händelser, att identifiera och peka ut dem.
- 2) *Ouïr*, att åhöra, betyder att uppfatta med örat, att höra till skillnad från att vara döv.
- 3) *Entendre*, att lyssna, är etymologiskt besläktat med ordet intention, och kan betyda höra, förstå, uppfatta. Det är just aktiviteten att välja ut något för uppmärksamheten att fokusera på.
- 4) *Comprendre*, att förstå. Här är ljudet bärare av en information, inte direkt via indicier, utan genom en kod som vi kan förstå.

Dessa fyra lyssningsaktiviteter ska inte ses som successiva etapper i perceptionsprocessen, och de kan inte heller helt lösrivas från varann. Snarare förekommer de alla samtidigt i en given lyssningssituation, men med olika vikt. Detta schema har, liksom flera andra i *Traité des Objets Musicaux*, en indelning dels i en konkret och en abstrakt pol, till höger resp. vänster, dels i en subjektiv och en objektiv pol, underst resp. överst. Samma indelning i dessa fyra sektorer kommer till användning när det gäller att beskriva ett lingvistiskt system, en akustikers eller en musikers lyssningar och mycket annat. Ett slags ordspråk summerar kärnfullt de fyra lyssningsfunktionerna (Schaeffer 1977 s. 147): «Je vous ai ouï malgré moi, sans que j'aie écouté à la porte, mais je n'ai pas compris ce que j'entendais.» Alltså ungefär: Jag åhördde er ofrivilligt, utan att ha stått vid dörren för



att höra på, men jag förstod inte det jag lyssnade på. Låt oss se närmare på de fyra lyssningsfunktionerna.

1) Man hör efter indicier på händelser, och man frågar sig «vad är det, eller vem är det, vad är det som händer?», utan att stanna upp för att betrakta ljudet som sådant. «Det är utan tvekan det allra vanligaste fallet, eftersom det svarar till vårt mest spontana förhållningssätt och perceptionens mest primitiva roll: att varna för en fara, beledsaga en handling. I allmänhet är denna identifikation av ljudhändelsen [l'événement sonore] i sitt orsakssammanhang ögonblicklig. Men det kan också hända, ifall indicierna är tvetydiga, att den inte uppstår förrän efter en del jämförelser och slutledningar. Den vetenskapliga nyfikenheten strävar mot ett mål som i grunden liknar perceptionens spontana utforskning av en händelse, fastän den använder sig av högt utvecklade kunskaper.» (Schaeffer 1977, s. 114-115) Detta är en bekant situation när man håller på att experimentera med en syntesteknik och resultatet inte blir riktigt som man hade väntat sig. Då frågar man sig kanske inte vad det är som låter, utan varför det låter som det gör, eller i värsta fall varför det är helt tyst, och hur man ska kunna förändra det.

2) Genom att fråga sig «vad är detta jag varseblir?», kan man behandla sin egen perception som ett objekt också. Det är vad Schaeffer kallar det 'råa ljudobjektet' (l'objet sonore brut). «Det är det som förblir identiskt igenom flödet av olika successiva intryck jag har av det, även i förhållande till mina olika intentioner gentemot det. Den andra väsentliga egenskapen hos ett varseblivet objekt är att det inte ges annat än som skisser: i det ljudobjekt jag hör på, finns det alltid mer att lyssna på; det är en källa som aldrig töms på möjligheter. Sålunda hör jag samma objekt vid varje repetition av ett inspelat ljud: fastän jag aldrig lyssnar på samma sätt, fastän det går från att ha varit okänt till att bli bekant, fastän jag succesivt varseblir olika aspekter av det, fastän det alltså aldrig är likadant, identifierar jag det hela tiden som detta bestämda objektet.» (ibid s. 115) Det är uppenbart vilken betydelse inspelningsmedierna med deras möjlighet att upprepa ett ljud har haft, inte bara för att utforska ljud, men på ett mera revolutionerande sätt i synen på ett stycke ljud som ett objekt.

3) «Det är likaså samma ljudobjekt som olika lyssnare samlade kring en bandspelare hör. Emellertid lyssnar de inte alla på samma sak, de väljer inte ut och uppskattar det samma, och i den grad deras lyssning tar parti för den ena eller andra speciella aspekten av ljudet, ger den upphov till den ena eller andra kvalifikationen av objektet. Dessa kvalifikationer varierar, liksom lyssningen, som en funktion av varje tidigare erfarenhet och varje nyfikenhet.» (ibid s. 115) Men ljudobjektet är inte ett mentalt tillstånd, på så sätt att varje ny lyssning till samma fysiska signal skulle ge upphov till nya ljudobjekt, utan det transcenderar varje enskild upplevelse lyssnaren har av det.

4) «Slutligen kan jag behandla ljudet som ett tecken som introducerar mig i en viss domän av värden [valeurs] och intressera mig för dess mening [sens]. Det mest typiska exemplet är förstås ordet [la parole]. (...) Bland de olika möjliga 'signifierande' lyssningarna, intresserar vi oss naturligtvis i synnerhet för den musikaliska lyssningen, som refererar till musikaliska värden och ger tillgång till en musikalisk mening. Låt oss anmärka att de värden det här är fråga om, till och med kan lösgöras från sin klingande kontext, som sålunda reduceras till en roll som medium [support].» (ibid s. 115-117) De värden det rör sig om är alltså sådana som man lär sig identifiera i gehörsträningen, dvs. skalor, intervall, ackordfunktioner och förhållanden mellan tidslängder. En stor ters förblir en stor ters oavsett om den spelas på piano eller trumpet, och så förhåller det sig med alla abstrakta storheter, att de förblir det samma oavsett hur de realiseras konkret. Timbre är kanske ett mera knepigt fall, och till en början ger Schaeffer något som ser ut som en cirkeldefinition: Ett pianos timbre är det som gör att man kan identifiera det som ett piano. Men samtidigt är det stora klangskillnader mellan registren i ett piano, och dessutom mellan olika instrument. Om man ser på timbre som något som gör att man kan identifiera ljudkällan, så hör det uppenbarligen till sektor 1 och den konkreta sidan i tabellen. Men egentligen innebär referensen till ett instrument i ett partitur en viss abstraktion; om där till exempel står cello, tänker man på en klang som i stor grad är den samma oberoende av specifika realisationer – av olika cellister och olika toner och spelsätt.

I tillägg till de fyra lyssningsfunktionerna kommer två nya par, *écoute banale/praticienne*

och *écoute naturelle/culturelle*. Den naturliga lyssningen är den vi delar med djuren, som använder sig av ljudet för att informera sig om händelser och eventuella hotande faror. Den hör alltså hemma i sektor 1. Den kulturella lyssningen inriktar sig på att förstå något meddelande genom ljudet, och motsvarar sektor 4. Detta par representerar således den konkret-abstrakta axeln i schemat. *Écoute banal*, som kanske bättre kunde översättas med naiv lyssning, vänder sig hela tiden mot ljudets mest markanta drag, om det så är dess kausalitet eller signifikation, men intresserar sig föga för ljudet som sådant. Emot det står *écoute praticienne*, det specialiserade lyssnandet. Olika lyssnare uppövar olika specialiteter: pianostämmaren är bra på att höra efter intonation och svävningar i ett piano, pianisten förmår uppskatta fina nyanser i klang mellan olika instrument eller olika anslag på samma instrument, en ljudtekniker som ska ställa in en nivå lyssnar på den dynamiska konturen och en musikintresserad lyssnar efter intervall och rytmer. Den naiva lyssningen leder till en automatisk respons på ljudet, så som är fallet i vardagliga situationer, medan den specialiserade lyssningen klamrar sig fast vid vissa drag eller inövade sätt att kvalificera ljudobjektet. Om den naiva lyssningen är oförfinad, så är den i alla fall öppen för allt som händer, till skillnad från den specialiserade, som riskerar att göra sig döv för allt annat än det den råkar fokusera på.

### 3.2 Reducerad lyssning

Ett viktigt begrepp som har förekommit några gånger utan att definieras är ljudobjektet, *l'objet sonore*. Det kan beskrivas som en ljudhändelse upplevd som en enhet, som en gestalt. Men ljudobjektet existerar inte i sig, utan det är förbundet med en speciell lyssnande aktivitet, nämligen *l'écoute réduite*, den reducerade lyssningen. Den är reducerad i den meningen att man skalar av allt från ljudet som inte är enbart ljud, dvs. de upplysningar om händelser i omvärlden ljudet förmedlar i sektor 1, eller den mening det förmedlar i sektor 4. Man företar en fenomenologisk reduktion eller *epoché*, ett sättande inom parentes av problemet med omvärldens existens, och vänder sig till det omedelbart och obestridligen givna: sina egna upplevelser. Det är inte ett "naturligt" förhållningssätt, eftersom det strider mot våra mest spontana reaktioner på sinnesintryck, och det kan kräva en del övning för att avbeta lyssnandets vanor. Men samtidigt består reduktionen inte i att förneka dessa naturliga perceptioner och få dem att försvinna, utan snarare sätter den dem i ett nytt perspektiv: «De même l'écoute réduite définie par Pierre Schaeffer ne consiste pas à nier les écoutes 'naturelles' (de son comme indice ou comme signe), ou à y dénoncer une 'illusion', mais à en démêler les différentes intentions constitutives, et à retourner ces intentions sur l'objet sonore, en tant que support des perceptions qui le prennent comme véhicule, pour le définir lui-même par une nouvelle intentionalité spécifique, celle de l'écoute réduite.» (Chion 1983 s. 32) I stället för att uppfatta en händelse genom ljudet som mellanled, hör vi på ljudet självt som en ljud-händelse. «Plus précisément, elle [l'écoute réduite] consiste à inverser cette double curiosité pour les causes et le sens (qui traite le son comme un intermédiaire vers d'autres objets visés à travers lui) pour la retourner sur le son lui-même. C'est l'événement que l'objet sonore est en lui-même (et non auquel il renvoie), ce sont les valeurs qu'il porte en lui-même (et non dont il est le support) que vise, dans l'écoute réduite, notre intention d'écoute.» (ibid s. 33) Uppmärksamheten riktad mot objektet i ett reducerat lyssnande kan använda sig av det den vet om händelsen, till och med om meningen, för att bättre förstå hur objektet är beskaffat (Schaeffer 1977 s. 293).

Ljudobjektet och den reducerade lyssningen är varandras korrelat; de definieras ömsesidigt i förhållande till varann som objektet för perceptionen respektive den perceptiva aktiviteten som konstituerar objektet. Eftersom ljudobjektet inte existerar för sig, kan det naturligtvis inte vara det samma som ett stycke inspelat ljud på en skiva eller ett band, och dessutom ger avspelningar på olika hastighet, eller baklänges, upphov till nya ljudobjekt. Men det är inte heller enbart ett sinnestillstånd hos lyssnaren (l'objet sonore n'est pas un état d'âme) utan är det samma från gång till gång, och har en objektiv existens på så sätt att flera lyssnare till slut förväntas komma till en gemensam uppfattning av det om de skulle försöka beskriva det. Men det är möjligt att fokusera på

olika nivåer i ljudflödet och avgränsa olika stora bitar som ljudobjekt. Varje objekt är del av en större struktur, en *contexte* kallar Schaeffer det, men denna struktur kan också behandlas som objekt, som ingår i en ännu större struktur, och så vidare. Omvänt har objektet sin struktur eller *contexture*, som återigen består av mindre objekt. Det är alltså inte givet från början hur pass utsträckt ett ljudobjekt är. Ibland är en not i traditionell musik en lämplig enhet, men i vissa sammanhang, som ett snabbt arpeggio på en harpa eller en virtuos kromatisk löpning, är hela gesten en bättre kandidat. Enligt Schaeffer är det vissa typer av ljudobjekt som bättre lämpar sig för musik än andra, vilka han kallar *objets convenables* (se avsnitt 4.3). De utmärker sig bland annat genom att vara lätta att förhålla sig till med en reducerad lyssning. Om de inte vore det skulle man kanske grubbla över ljudproduktionsmekanismerna i instrumenten när man gick på konsert. Det är intressant att se hur ovana lyssnare reagerar på spel på preparerat piano, om de inte är förberedda på prepareringen: de försöker titta efter för att se om de kan identifiera vad som orsakade de ovana ljuden. I den elektroakustiska musikens barndom användes ett utmärkt verktyg att öva sitt reducerade lyssnande med i Schaeffers studio, nämligen grammofonskivor graverade med ett slutet spår, en «*sillon fermé*». Genom att spela in ett ljud på en skiva med ett sådant slutet spår, kunde man upprepa det hur länge som helst. Dessa upprepningar gör att man fort tappar intresset för ljudets anekdotiska ursprung eller dess mening. I stället framträder ljudfragmentet för lyssnaren som ett objekt, lika vid varje återkomst, men som ändå ständigt kan avslöja nya drag. Men denna situation är inte alls någon garanti för ett reducerat lyssnande, för det är fortfarande möjligt att vända sig mot ljudets indicier: «det är ett hack i skivan, gå och lyft nålen!» eller meningen: «det är någon som upprepar en fras gång på gång, varför det?» Sammanfattningsvis tror jag att Schaeffers beskrivning av de olika lyssningsaktiviteterna kan vara en användbar referensram att beskriva många situationer med, inte minst i signalbehandling och elektroakustisk komposition, där man alltid löper en risk att specialisera sin lyssning så till den grad att man blir döv för de mest uppenbara aspekterna av ljudet.

### 3.3 Auditory Scene Analysis

«Låt mig klargöra vad jag menar med auditiv scenanalys. Det bästa sättet att börja är att fråga oss vad perception är till för. Allt sedan Aristoteles har många filosofer och psykologer ansett att perception är en process där vi använder informationen våra sinnen förser oss med för att bilda mentala representationer av vår omvärld. Genom att använda ordet representationer, implicerar vi att det finns ett tvådelat system: en del bildar representationerna och en annan använder dem för att göra sådant som att beräkna lämpliga planer och handlingar. Perceptionens uppgift är då att ta emot sinnesdata och härleda en användbar representation av verkligheten från dem.» (Bregman 1990, s. 3)

En viktig del i uppbyggandet av en sådan representation är att avgöra vilka sinnesdata som härrör från samma objekt eller händelse i omgivningen. En fördel med att studera perception på detta sätt, utifrån representationer, är att det blir möjligt att ta hänsyn till samspelet av olika sinnesmodaliteter i det att flera sinnen kan ge upplysningar om samma händelse eller objekt. Men det Bregman koncentrerar sig om, är hur sinnesdata grupperar ihop sig, främst auditivt, men också med några exempel på visuell gruppering. En sådan grupperingsmekanism är nödvändig för att bilda en god eller användbar representation av tilldragelserna eller föremålen i omgivningen, för om helt disparata objekt skulle grupperas ihop, skulle det ge upphov till «schimär-objekt», som inte svarar till några verkliga objekt. Och omvänt måste de sinnesdata som härrör från ett och samma objekt grupperas ihop, även om de skulle vara av inbördes varierande art, som till exempel de språkljuden som kommer från en person. Man kan spekulera i om inte perceptionspsykologiska försök med hallucinogena droger skulle kunna avslöja något mer om de mekanismer som ligger bakom den vanligtvis koherenta mentala representationen av verkligheten. Annars har mycken insikt om hjärnans mekanismer utvunnits ur fallbeskrivningar av hjärnskador. Till vardags lägger vi knappast märke till de problem hörseln stöter på när det gäller att gruppera och identifiera

ljudkällor. Om det är så som Bregman argumenterar för, att en stor del av grupperingen sker automatiskt och på en förmedveten nivå, så är det inte så konstigt om vi inte observerar denna process medvetet. Bregman (1990 s. 5-6) föreslår följande tankeexperiment för att erinra om de svårigheter hörseln har att övervinna: Föreställ dig att vi befinner oss vid en sjö, där vi gräver två smala kanaler från strandkanten en bit från varann. I vardera kanal fäster vi en utspänd näsduk, så att vågor från sjön sätter dem i rörelse. Vi tillåts bara titta på näsdukarna, och utifrån deras rörelser ska vi besvara ett antal frågor, som: Hur många båtar är det på sjön, och var är de? Vilken är närmast? Blåser det? Har man plötsligen släppt ett stort föremål i sjön?

Bregman inför en del terminologi som behöver förklaras närmare. Auditory scene analysis (ASA) går alltså ut på att det föregår en tolkningsprocess där sinnesdata väljs ut till att representera olika objekt i omvärlden. Ett annat centralt begrepp är auditory stream. «An auditory stream is our perceptual grouping of the parts of the neural spectrogram that go together.» (Bregman 1990, s. 9.) Den som har försökt identifiera ett ljud enbart genom att avläsa ett spektrogram eller cochleagram vet hur svårt det är. Men denna svårighet blir genast oerhört mycket större om spektrogrammet innehåller en överlagring av flera olika ljud. Bregman (1990, s 10) fortsätter: «Many happenings go on at the same time in the world, each one a distinct event. If we are to react to them as distinct, there has to be a level of mental description in which there are separate representations of the individual ones. I refer to the perceptual unit that represents a single happening as an auditory stream.» Men denna perceptuella enhet som ska representera en enskild händelse kan mycket väl vara orsakad av flera händelser. Man finner lätt exempel i musik där unisoner och parallellförda mixturklanger smälter samman trots att ljuden kommer från helt olika ljudkällor och riktningar. Varför denna neologism, *auditory stream*, varför inte bara kalla det ett ljud? Det finns två skäl till det, säger Bregman. Det ena är för att ordet ljud utan någon åtskillnad refererar både till fysiska vibrationer i luft och till vår motsvarande subjektiva upplevelse av det. Det andra är att en fysisk händelse och motsvarande mentala representation kan innehålla flera ljud, som exempelvis en serie fotsteg eller ett utsnitt av ett stycke musik. Det är klart att vare sig man hör en person som går eller en ensemble som musicerar, så kan man gruppera in det hierarkiskt i flera nivåer på olika tidsskalor, men detta säger Bregman inte mycket om. Hans undersökningar, liksom de många andras han citerar, rör sig mest om grupperingsmekanismer i ett tämligen lokalt tidsperspektiv.

På många sätt påminner *auditory stream* om *l'objet sonore* hos Schaeffer. Är det rent av samma sak? Det väsentliga i en *auditory stream* är att den är den subjektiva upplevelsen av ett ljud, och den grupperar ihop ljud som är tillräckligt likartade eller av någon annan anledning kan fås att gruppera sig, och som ofta kan förstås vara föranledda av samma fysiska tilldragelser. Ett *objet sonore* är också det den subjektiva sidan av ljudet, och här är Schaeffer ännu mer specifik: det är frågan om ett ljud så som det är hört under intentionen att bedriva *écoute réduite*. Ett *objet sonore* kan ingå i en struktur som igen kan höras som ett objekt, som kan ingå i en större struktur, osv., samtidigt som objektet har sin struktur, som igen består av mindre objekt, och så vidare i all evighet. Men någonstans tar naturligtvis tidsupplösningen stopp, och likaså minnets förmåga att binda ihop långa tidssträckor till en helhet. Denna flexibilitet i utsträckning är den andra stora likheten mellan *auditory stream* och *l'objet sonore*. Men det finns också några viktiga skillnader. Medan det reducerade lyssnandet som ska konstituera *l'objet sonore* kräver ett sättande inom parentes av alla slags indicier och mening (sektor 1 och 4 i Schaeffers lyssningsmodell), finns det ingen liknande inskränkning för en *auditory stream*; den säger inget om lyssnarens intention. Det är mindre klart hur Schaeffer tänker sig samtida ljudobjekt; är det totaliteten av ljud uppfattade över ett tidsrum som utgör *l'objet sonore*, eller kan man tänka sig flera samtida *objets sonores*? Däremot är det helt klart att *auditory streaming* ger utrymme för flera samtida «strömmar». Här kommer ytterligare två viktiga begrepp in, nämligen *stream segregation* och *stream integration*. De faktorer som påverkar vad som integreras i en ström och vad som stöts bort från den är egentligen själva huvudfokus för Bregman.

### 3.4 Principer för gruppering

Grupperingen i form av integration och segregation studeras på två sätt, sekventiellt och simultant. Typiska laborieförsök som försöker studera de sekventiella grupperingsmekanismerna går ut på att spela upp tonsekvenser med två eller flera toner som upprepas i snabb följd för lyssnare som ska ange om de hör en eller två strömmar. En viktig princip för sekventiell organisering är effekten av samspelet mellan hastigheten och tonavståndet i en tonsekvens. En sekvens med fyra olika, ömsom höga och låga toner som upprepas en lång stund, kan antingen gruppera ihop sig till en sammanhängande sekvens, eller till två simultana sekvenser med höga respektive låga toner (Bregman 1990 refererar flera sådana försök). Vilken av dessa grupperingar perceptionen organiseras i beror både på tonavståndet på en logaritmisk frekvensskala och tidsavståndet mellan tonansatserna. Om tonavståndet är stort så tenderar sekvensen lätt att splittras i två skikt om tidsavståndet samtidigt blir tillräckligt litet. Men om samma sekvens spelas i ett lägre tempo kan den integreras till en helhet. Med mindre melodiska intervall tenderar sekvensen att sitta ihop så att man hör en växling mellan höga och låga toner.

Sekventiell organisation är nog mest intressant att studera just i gränfallen, där den perceptuella grupperingen tippar över från en organisering till en annan. Lagen om det uteslutna tredje säger att ett faktum inte kan både vara fallet och inte vara fallet, det måste vara endera sant eller falskt. Och den perceptuella organisationen tycks också ofta åtlyda en sådan antingen-eller-princip. Den välkända bilden av en vas som också kan ses som två profiler vända mot varann, men aldrig samtidigt, är ett bra exempel på det. Det är karakteristiskt för alla sådana tvetydiga bilder att man fritt kan växla mellan att se dem på det ena eller andra viset. På samma sätt finns det tvetydiga tonsekvenser som låter lyssnaren gruppera tonerna efter eget förgottfinnande. Man finner det inte bara i laborieförsök, men också i en hel del minimalistisk musik och monoton elektronisk dansmusik. I den mån man kan tala om någon komplexitet i sådana musikformer, är det väl mest på grund av den frihet de ger lyssnaren att själv organisera ljudmönstren. Men det är bara i ett begränsat område av tidslängder som denna effekt av tvetydighet brukar uppstå. Vid tonlängder på 500 ms eller mer börjar det bli svårt att uppnå någon spontan uppdelning av tonsekvensen. Dessutom kräver tonsekvenser i lägre tempo en längre tid innan den börjar splittras i två delar; under de första tonerna har den speciella grupperingen av sekvensen ännu inte hunnit etablera sig. En motsvarande effekt uppstår i slutet av sekvensen. En effekt av att en tonsekvens splittras upp i två delar är att det blir svårt eller omöjligt att bedöma ordningsföljden mellan toner i de olika strömmarna. Men för de första och sista tonerna i sekvensen kan det ändå vara möjligt att uppfatta ordningsföljden, eftersom grupperingen inte hinner etablera sig på en gång i början.

Det finns en nedre tidsgräns för var det är meningsfullt att studera sekventiell organisation, och den har att göra med hörselns tidsupplösning och integrationströskeln vid 20 Hz där enskilda impulsionsövergångar övergår till tonhöjdsförnimmelser. Omräknat till periodlängd motsvaras 20 Hz av 50 ms. Så om man presenterar en sekvens med tonlängd 50 ms eller kortare, borde man höra en låg tonhöjd som svarar mot perioden. Men en tvåtonsssekvens av 50 ms-toner är periodisk med 100 ms, vilket ger frekvensen 10 Hz. Därför måste man ner till tonlängder under 25 ms innan sekvensen ska integreras fullkomligt och bara låta som en låg ton<sup>3</sup> vars form bestäms av de två frekvenserna i sekvensen. Detta är ett exempel på att tonhöjdsupplevelsen kan basera sig på ett ljuds period. Grundtonen kan mycket väl vara frånvarande i spektrumet i det fallet, och det är den sannolikt också i den här varianten av granulärsyntes.

Även om sekvenser av toner som växlar så snabbt att de ger upphov till tonhöjdsförnimmelser är ett exempel på simultan gruppering av olika spektrala komponenter, är det inte den vägen man vanligen har gått för att studera villkoren för simultan gruppering. I stället har man bland annat gjort försök med att stämna om en av deltonerna i ett i övrigt harmoniskt spektrum. Med en liten grad av omstämning, några få procent av frekvensen, förändras den

---

3 Denna iakttagelse behöver inte vara helt korrekt, eftersom tonens amplitudkurva (envelope) upprepar sig för varje ny ton, oavsett dess frekvens.

uppfattade tonhöjden i samma riktning som deltonens frekvensförändring, men med ytterligare några procents avvikelse lösgör sig den omstämda deltonen, och i det samma upphör den att förändra tonhöjden i det resterande tonkomplexet. I ett försök jag gjorde med grundfrekvens 440 Hz, åtta deltoner, och den tredje varierande stämd, kunde jag bekräfta effekten på tonhöjden för en stämning 1% högre, och från ca 2% började duodeciman lösgöra sig. (Det kan vara förvirrande att tala om frekvensavvikelser i procent, såsom psykoakustiker gärna gör, så det kan löna sig att räkna om det till cent. I så fall ser man att 1% motsvarar ca 17 cent, och en kvartston motsvarar ca 2.9%. Det rör sig alltså om några av de minsta uppfattbara mikrintervallen.) Men för att höra en omstämd delton i en sammansatt ton kan man behöva växla mellan att höra den omstämda varianten och den ursprungliga. I så fall hör man två sammansatta toner som är helt lika med undantag för den omstämda tonen, och det blir möjligt att rikta uppmärksamheten mot den delen av spektrumet. Denna typ av spektralanalys på gehör har gamla anor, Helmholtz sägs ha varit bra på att urskilja de lägsta deltonerna i ett sammansatt ljud. En mycket vanlig experimentsituation går ut på att ställa upp konkurrerande ledtrådar om hur olika ljud ska grupperas. Kan man till exempel bryta ut en harmonisk delton ur en sammansatt ton genom att först spela upp samma ton eller en annan tillräckligt nära?

Det är förstås inte bara tonhöjd som styr grupperingen av ljud, många andra egenskaper kan också göra det, och ofta samverkar de och förstärker effekten. Variationer i spatial placering, spektralkontur, formanter, amplitud- och frekvensmodulation är några saker som kan användas till att separera eller gruppera ljudkällor. Amplitudskillnader verkar inte kunna orsaka någon segregering på egen hand, men kan förstärka effekten om andra egenskaper varierar tillsammans med amplituden. Att olika komponenter av ett ljud varierar på samma sätt (som i ett synkroniserat vibrato) är ett exempel på common fate, som är en vanlig grupperingsprincip som redan gestaltpsykologerna kände till.

Bregman skiljer på primitiv och schema-baserad auditiv organisering, och det är huvudsakligen de primitiva grupperingsmekanismerna han beskriver. Den primitiva grupperingen antas ske automatiskt och på ett förmedvetet plan och är inte beroende av tidigare inläring. Ett schema eller en kognitiv struktur är däremot ett inlärt mönster, eller möjligen någon medfödd kunskap. Typiska exempel är ord, grammatiska regler (som yttrar sig som en naturlig språkkänsla) och melodier. Det har gjorts försök där tonerna från två melodier i ungefär samma register spelas växelvis, och det gäller att känna igen en av melodierna (Bregman 1990 s. 407, 411). Det är till en början inte möjligt, eftersom successiva toner grupperar sig med varandra och bildar en helt ny melodi, men försökspersoner har klarat av att identifiera en av de ursprungliga melodierna under förutsättning att den var känd sedan tidigare, och att man talade om vilken det var. Men inte ens då förmådde de uppfatta den förrän efter flera lyssningar. Det är generellt sett lättare att få en del av en sammanhängande sekvens att segregera genom att koncentrera sig på delar av den än det är att gruppera ihop de delar som tenderar att segregeras av den primitiva grupperingsmekanismen. Med en funktionalistisk förklaring konstaterar Bregman (1990 s. 452): «Attention does not operate with equal facility on any combination of sensory elements. Otherwise it would not be hard to select anything at all for attention.» Det kan vara värt att tänka på när det gäller Schaeffers objekt sonore, som inte enbart är en viljemässig konstruktion, utan vars utsträckning och avgränsning faktiskt bestäms till stor grad av ljudets egenskaper. Det finns alltså naturliga gränser i strömmen av ljud, som den mellan tystnad och ljud, eller mellan en ton på ett instrument och en hostning.

### 3.5 Tillämpningar av ASA

De kunskaper man börjar samla på sig om hörselns grupperingsmekanismer är av stort intresse i tillämpningar som automatisk igenkänning av ljud, som försvåras drastiskt om störande ljud är närvarande; transkriptioner av musik direkt från en ljudsignal, vilket är ett avancerat problem om musiken är polyfon, och datormodeller av perception. Bregman (1998) varnar för datormodeller

som inte tar hänsyn till mängden av psykologiska data om perceptionsprocessen, sådana modeller riskerar att bara duga till att lösa en begränsad klass av uppgifter. Enligt Slaney (1998) har datormodeller av perception hittills varit nedifrån-och-upp, där all information kommer från sensorer, varefter bitar av den kombineras tills slutligen ett objekt känns igen. Motsatsen, en uppifrån-och-ned-modell, utgår ifrån att man redan vet vad som finns där ute, och perceptionens uppgift är bara att samla ihop bevis som stödjer antagelsen. Det mesta tyder på att perceptionen är mera interaktiv, den arbetar med modeller av världen, och de olika sinnena samverkar. Om vi hör ett plötsligt ljud vänder vi oss mot det för att se vad som orsakade det. Signaler i nervsystemet är både efferenta och afferenta, de går åt båda hållen. Antagligen måste man ta hänsyn till det om man ska lyckas bygga upp en god perceptionsmodell. Kan man använda ASA till något nyttigt i musiken, i komposition eller signalbehandling? En möjlighet som Bregman är inne på, är en vetenskap om orkestrering, och inte bara en lära. Den skulle i så fall kunna förutsäga resultaten av olika klangkombinationer och relatera det till dissonansgrad och andra uppfattade egenskaper. De försök som har gjorts med multidimensionell skalering, som vi ska se lite på nedan, skulle kunna bidra till att kartlägga detta område, men det tycks ännu finnas mycket kvar att göra.

## 4. Ljudmodeller

I många vitt skilda sammanhang finns det behov för ljudmodeller i form av klassifikationer av ljud eller utpekande av olika väsentliga klangliga egenskaper. Den elektroakustiska musikens öppenhet inför alla slags ljud gör det önskvärt att ha någon slags begreppsapparat för att beskriva och sortera genom att iaktta skillnader i flödet av ljud. Vi kommer i korthet att beskriva Schaeffers typomorfologi. Schaeffers *Traité* måste ses som ett av de mest ambitiösa projekten när det gäller att skapa ljudmodeller. Från ett helt annat perspektiv kan man konstruera ljudmodeller för automatisk igenkänning av ljud inom artificiell intelligens. Dessa behöver inte använda sig av samma strategier som den mänskliga perceptionen vid identifikation av ljudkällor, även om det troligen skulle förbättra modellerna. Fonetikens klassifikation av språkljud i fonem visar på rikedom av möjligheter. En viktig poäng är den kategoriska perceptionen av fonem, som yttrar sig i att vissa olika artikulerade ljud i vissa språk tolkas som samma (allofoner), medan andra språk särskiljer dem (Lundström-Holmberg & af Trampe 1987). I musiken är kategorisk perception välkänd när det gäller skalor. Man kan säga att det är den som gör att vi kan höra en avvikande ton såsom falsk.

Klang eller timbre har till stor del varit ett negligerat område i musikteorin. Tonhöjd, tidsvärden och dynamik har, som grundläggande musikaliska dimensioner, visat sig vara lättare att behandla. Klangfärgen däremot, kan inte sägas vara en dimension, utan flera. Men hur många?

### 4.1 Multidimensionell skalering

Det har på senare tid gjorts en del försök med uppfattningen av klangfärgen av olika orkesterinstrument och kombinationer av instrument. Kendall & Carterette (1993) studerade hur bra två blåsinstrument smälte samman i olika kontexter som unison ton, separerade med en stor ters eller i melodiska rörelser. Instrumenten de använde var oboe, flöjt, trumpet, sax och klarinett. Ett antal försökspersoner fick ange graden av likhet mellan olika par av instrumentkombinationer. Dessa data tolkades genom multidimensionell skalering: uppskattningen av likhet mellan två ljud betraktas som ett avstånd mellan punkter i något geometriskt rum. Problemet man försöker lösa med multidimensionell skalering är att finna ett rum och en konfiguration av punkterna som är förenlig med data om deras inbördes avstånd. Kendall & Carterette fann ett tredimensionellt rum där punkterna är ansamlade i klasar. För att tolka innebörden av de olika dimensionerna lät de försökspersonerna välja bland ett urval adjektiv för att beskriva skillnader längs de olika axlarna. På den första axeln visade instrumentkombinationerna sig vara ordnade efter graden av nasalitet, med

oboe/flöjt som den mest nasala kombinationen, och sax/klarinetten den minst nasala. Den andra axeln representerar 'brilljanta' klanger i den ena riktningen och 'rika' klanger i den andra. Kombinationer av trumpet och ett annat instrument tenderar att vara 'brilljanta', medan kombinationer med sax är 'rika'. Den tredje axeln representerar 'starka' och 'komplexa' mot 'svaga' och 'enkla' klanger.

Genom att låta försökspersonerna bedöma i vilken grad instrumentkombinationerna lät som ett eller två instrument fann de föga förvånande sambandet, att ju större grad av sammansmältning, desto svårare var det att identifiera båda instrumenten. Det de säger om att para ihop verbala etiketter med klangliga karaktärer är belysande:

«We can conceive of the task of assigning a verbal attribute, such as "blend", to a sound event as superimposing two multidimensional configurations, one sonic, the other verbal. The verbal configuration relates the target word, "blend", to the multiplicity of meanings about itself in relation to all other words. The "meaning" of the target timbre is defined by its position relative to mental structures of timbre.» (Kendall & Carterette 1993 s. 58.)

Men förhållandet mellan ord och klang är dessvärre inte alltid så enkelt, vilket Handel (1995 s. 444) påpekar: «It should be noted that a similarity space based on direct sound judgments is not simply related to a similarity space based on adjective judgments. (...) If there was an equivalence between sound similarity and adjective similarity, the configurations ought to map onto each other. But, this is not usually the case. The adjectives are useful for conveying sound quality, but they are only tentative metaphors about timbre. The adjectives are easy to interpret in different ways, and the variability across subjects is often large; this may explain the lack of convergence.»

Som ett nästa steg kan man försöka ställa upp korrelationer mellan de verbala etiketterna och akustiska egenskaper som styrkegradsförhållanden mellan deltonerna, spektral centroid, deltonernas grad av fluktuation eller stabilitet, med mera. Som ett verkligt ambitiöst projekt skulle man slutligen kunna designa kontrollmöjligheter för en syntesmodell så att de direkt relaterade till de funna perceptuellt relevanta egenskaperna.

## 4.2 Identifikation av ljudkällor

Med tanke på vår goda förmåga att identifiera olika objekt och händelser genom ljud, vilka problem är det processeringen av sinnesdata måste lösa? Som Handel (1995) påpekar, måste vi kunna upptäcka en enhet i mångfalden av ljud som har uppstått på samma eller liknande sätt. Om ljudvågorna från ett visst objekt alltid var identiska, skulle vi bara behöva memorera dem, men vi vet att de varierar (och även deras korttidsspektrum, vilket är mera relevant för perceptionen), och om flera ljudkällor ljuder samtidigt måste vi också kunna separera dem. På vilken nivå ska vi beskriva den invarians som möjliggör vår identifikation av föremål? Å ena sidan kan timbre (uppfattat som den perceptuella kvaliteten som vidlåder objekt) uppfattas i termer av fysisk rörelse (studsande, strykande, knäppande) och egenskaper hos den vibrerande strukturen (material, storlek). En perception omedelbart knuten till ljudets tillblivelseprocess skulle möjliggöra identifikation av ljud med starkt varierande vågformer på grundval av ljudproduktionens invarians. Å andra sidan är det tänkbart att timbre uppfattas i termer av akustiska egenskaper, och att sambandet mellan dessa och föremål/händelser är inlärd. Det är rimligt att tänka sig en kombination av båda dessa strategier (Handel 1995). Utan möjligheten att lära in nya kopplingar från ljudkälla till ljudmönster hamnar man fort i orimligheter. Men det kan också argumenteras för att vår art har genomgått en ekologisk anpassning för att utvinna livsviktig information ur akustiska signaler, och att vår hörsel därför är anpassad till ljudkällor i vår omgivning. Därför skulle man redan på förhand kunna anta att fysiska ljudmodeller var speciellt relevanta.

En vanlig fysisk modell är uppdelningen i excitation och resonans, eller ljudkälla/filter. Trögheten i vibrerande strukturer gör att de fortsätter att vibrera en stund efter att de har satts i svängning, vilket kan modelleras med ett filter. Den vibrerande strukturen kännetecknas av mer eller mindre dämpade resonanser i harmoniska eller inharmoniska förhållanden. I många fall har den



en återkoppling mellan excitation och resonans, till exempel i blås- och stråkinstrument. Även anslagna och knäppta stränginstrument som gitarr och piano har en återkoppling mellan strängen och resonanslådan.

Vilka akustiska egenskaper kan vi tänkas använda vid identifikation av ljud? Enligt ett gammalt antagande är spektrumet avgörande. Men spektrumet har flera egenskaper som kan spela in, som centroid, spektralprofil (ett mått på hur amplituden avtar som funktion av frekvensen), varians, som ger ett mått på spektrumets jämnhet eller kantighet, och formantfrekvenser. Musikinstrument och många andra föremål ger ifrån sig ett stort register av olika ljud med ytterst varierande spektral karakteristik. Att vi ändå kan identifiera sådana ljud som härrörande från samma objekt tyder på att spektrumet i sig inte är en tillräcklig ledtråd.

Vi har tidigare nämnt att transienter i början och slutet av ljud kan vara mycket avslöjande, och att identifikationen av vissa instrument drastiskt försvåras om attacken klipps bort. I akustiska termer karakteriseras attacktransienter av sådant som synkroniseringen av deltoner eller deras fasförhållande, och i vilken ordning olika frekvensområden inträder. Men olika toner på samma instrument kan uppvisa varierande transientmönster, så inte heller det kan vara en alena fungerande ledtråd.

Många andra akustiska egenskaper är säkerligen viktiga för ljudidentifieringen, som rytmiska aspekter (kontinuerliga eller upphackade ljud) och övergångar, till exempel mellan olika fonem. Släktskapet mellan en ytas textur — mönster, gräng — och de ljud man får genom att skrapa eller stryka över ytan är uppenbar.

Ett fåtal studier har gjorts med andra ljudkällor än traditionella musikinstrument eller röster; bland annat handklapp, fotsteg och saker som krossas eller studsar har studerats. I ett försök (refererat av Handel 1995 s. 451) med stekpannor i olika storlekar som anslogs med klubbor i varierande hårdhetsgrad (metall, trä, gummi, tygklätt trä, filt, filtklätt gummi) kunde lyssnarna korrekt inbördes rangordna klubbornas hårdhetsgrad. De hårdaste klubborna ger ljud med mycket högfrekvent energi under ansatsen, varefter den spektrala centroiden sjunker. Mjuka klubbor ger generellt svagare ljud och en mindre andel höga frekvenskomponenter, och centroiden förblir någorlunda konstant.

Föremål som faller och krossas beter sig ungefär likadant som föremål som faller och studsar. Om det splittras, så bildas skärvor som studsar. Hur hör vi skillnad på splittring och studs? Vi använder flera ledtrådar: något som studsar ger ifrån sig en accelererande sekvens ljud som blir svagare, medan något som krossas har flera överlappande sekvenser av studsande ljud med olika spektrum. Genom att variera betingelserna, kan man få syn på vad ljudets krossadhet eller studsadhet består i.

- En sekvens ljud som från studsande föremål, men påfört olika spektrum för varje studs, förstör egenskapen 'studsande föremål'.
- Ljuden från ett splittrande föremål påförs samma spektrum. Det förstör intrycket av splittring.
- Att byta ut den accelererande sekvensen i ett studsande ljud mot en sekvens med lika tidsavstånd förstör intrycket av studsande. (Handel 1995 s. 455)

I ljudsyntes och understundom i ljudprocessering händer det att ett utmärkande drag är att ljudet låter syntetiskt i tillägg till alla andra kvaliteter det kan ha. Det vore därför på sin plats att utvärdera vad det är som gör att vi identifierar ett ljud som syntetiskt eller naturligt.

### 4.3 Schaeffers typomorfologi

Det är ett formidabelt problem man ställs inför om man försöker klassificera ljud: var ska man börja, hur ska man avgöra vilka ljud som hör ihop? Fonetikens indelningar av språkljud är inte direkt användbar på ljud i allmänhet, annat än som ett exempel på hur en sådan sortering kan gå till. En indelning som låter sig tänkas, är den som sorterar ljuden efter ljudkällorna. Det som gör det till en opraktisk strategi, är att samma ljudkälla (eller syntesmetod, vilket är mera relevant i vårt fall)

ofta kan producera vitt skilda ljud, varav vissa kan vara mera besläktade med ljud producerade på helt andra sätt. Man kan därför som Schaeffer utgå från den reducerade lyssningen som metod att urskilja ljudobjekten. Men den reducerade lyssningen i sig ger inga ledtrådar om hur man ska klassificera ljudobjekten. Den säger heller inget om vilka kriterier man ska lyssna efter för att med hjälp av dem klassificera ljudobjekten. En första början blir att finna diskontinuiteter i ljudet, för att på så sätt isolera ljudobjekten ur strömmen av ljud.

Men man kommer inte så mycket längre utan att ha en intention med klassificeringen. När nu Schaeffers intention är en klassifikation av ljudobjekt i all generalitet, men med tanke på potentiellt musikaliskt bruk, lägger det en del premisser. Det är ingen hemlighet att ett estetiskt ställningstagande ligger till grund för typologin — det kommer till syne i uppdelningen i 'objets convenables', objekt lämpade för musik, och andra slags objekt som utmärker sig genom att vara alltför enkla eller alltför originella.

Typologins uppgift är att identifiera och klassificera ljudobjekten; identifikationen innebär här att urskilja eller isolera objekten ur ett sammanhang de ingår i. Till sin hjälp behöver typologin en morfologi för en preliminär beskrivning av ljudobjekten. Schaeffer utgår från begreppsparet *forme/matière* (form och materia eller innehåll). *Materia* är det som fortbestår mer eller mindre oförändrat genom ljudet, medan *formen* betecknar den dynamiska utvecklingen. Genom att hålla *formen* konstant kan man således studera och klassificera ljud med olika *materia*. De två huvudtyperna är toniska och komplexa ljud. Toniska ljud är allt som har en fast tonhöjd (inte glissandi) medan komplexa ljud exemplifieras av olika slags brus och slagverk utan bestämd tonhöjd. Tre par av kriterier ligger till grund för typologin. Det första är *Masse/Facture* (bokstavligen: massa/utförande). Massan kan vara fixerad, och i så fall antingen tonisk eller komplex, eller den kan variera på ett överskådligt sätt över tid, som i ett glissando. Den kan också variera på ett oförutsägbart sätt över tid. Det är de fyra typerna av massa. Tre typer av faktur urskiljs: utsträckt, impulsiv eller punktuell, och iterativ. Vanliga exempel är uthållna toner, korta slagverksljud, tremolo. Det är en ganska originell idé att klassificera ljud efter längd, det medför att en lång uthållen ton betraktas som av en annan typ än 'samma ton' spelad staccato.

Det andra paret kriterier *Durée/Variation* (varaktighet/variation) tar för sig temporala egenskaper, och tänker samman två kriterier som inte är fullständigt oberoende av varandra (det gäller för övrigt de två andra paren också). Via det fenomen Schaeffer kallar tidsanamorfoser är den upplevda tidslängden beroende av informationstätheten. En annan sida av saken är att korta ljud helt enkelt inte hinner uppvisa samma grad av intern variation som långa.

Det tredje och sista paret kriterier är *Équilibre/Originalité* (balans eller jämvikt och originalitet). Balansen är att förstå som kompromissen mellan det alltför strukturerade och det alltför enkla, och originaliteten är ljudobjektets förmåga att överraska, att ta oväntade vändningar. I typologin klassificeras sålunda ljudobjekten i redundanta, excentriska (*trop originaux*) och *objets convenables* (passande, lämpliga). Det är alltså här det normativa inslaget inträder: *objets convenables* är de Schaeffer menar ägnar sig väl till musik, för att de per definition är av en sådan beskaffenhet att de är lätta att memorera, ägnar sig till ett reducerat lyssnande och kan användas, tillsammans med andra ljudobjekt av samma slag till att få värden (*valeurs* — typexemplet är tonhöjder) i en musikalisk struktur att framträda.

Dessa sex kriterier kombineras med varandra och sammanfattas i en typologisk tabell med ca 30 typer. Det skulle föra för långt att gå in på dessa typer och beskriva dem var för sig; se Chion (1983) för en överskådlig genomgång. Lasse Thoresen har på senare tid också bidragit med en grafisk notation för de olika typerna. Schaeffers typologi är inte lätt att använda på ljud i allmänhet. Det kan alltid uppstå diskussion om huruvida ett ljud ska placeras i en kategori eller en annan. Det är helt naturligt, dels eftersom kriterierna är diffusa nog att tolkas på flera sätt, dels för att ljuden själva låter sig tolkas på olika sätt alltefter vår lyssningsintention och det sammanhang de står i. Är det att betrakta som ett misslyckande för typologins del? Jag tror inte det. Tvärt om ser jag dess vaghet som en styrka, som gör det möjligt att använda den i en mängd olika sammanhang. Men problemet är att den i praktiken är alltför grov och säger alltför lite om ljuden. Två ljud som vi kan

komma överens om att placera i samma kategori skulle mycket väl kunna vara sinsemellan ytterst olika. Men Schaeffer har också utarbetat en morfologi, där man går vidare utifrån typologins klassificeringar och fortsätter ställa frågor till ljudet — det är i första hand det det handlar om, och inte att placera dem i fack.

Schaeffer skiljer på en extern och en intern morfologi. Den externa är föga utarbetad, och tillför bara två nya begreppspår: *Composé/Composite* och *Accident/Incident*. Den långt mera differentierade interna morfologin behandlar ljudobjektet och dess innanmäte eller *contexture*. Paret *Composé/Composite* (båda översätts dessvärre bäst med sammansatt) betecknar ljudobjekt sammansatta av flera samtidiga element respektive flera element i följd. Man kan urskilja dessa element och behandla dem som underobjekt av olika typer i det stora objektet de utgör tillsammans (ett ackord till exempel). Vi har redan redogjort för de principer för gruppering som har upptäckts inom auditory scene analysis.

*Accident* är en störning, en liten ovidkommande händelse som vidhäftar en annan, till exempel en sträng som råkar skorra. *Incidenter* är däremot störningar i ljudet på grund av något tekniskt fel, och som inte kan uppfattas som en egenskap vid ljudet. Knaster och sprak från grammofonskivor är ett exempel. Begreppet *incident* är mycket användbart för våra syften, så vi ska återkomma till det senare.

Morfologiska kriterier är egentligen det samma som drag eller egenskaper som kan urskiljas i ljuden, så i princip finns det ett obegränsat antal. I Schaeffers interna (och huvudsakliga) morfologi urskiljs fyra grupper av kriterier.

1) Under kriterier för *materia* räknas massa, som vi har nämnt tidigare, och *timbre harmonique*. Sju typer av massa beskrivs, och de bildar en slags skala från sinustonen via toniska ljud, flera samtidiga toner, ”räfflade” ljud, grupper av noder och enskilda noder (massa utan identifierbar tonhöjd), till vitt brus. *Timbre harmonique* används i det närmaste som ett hjälpkriterium för toniska ljud. Jag tror att sådana egenskaper som bland annat nasalitet, som har identifierats i de ovan beskrivna försöken med multidimensionell skalering, skulle kunna infogas under *timbre harmonique*.

2) Kriteriet för form är den dynamiska profilen. Beroende på den dynamiska utvecklingen av ljudet, klassificeras det bland annat i avsaknad av dynamik (*dynamique nulle*), vek (*faible*) dynamik, cyklisk, itererad eller ackumulerad dynamik (det sista hänför sig till myllerklanger, som består av en ackumulation av mindre ljudhändelser). Det finns också en klassifikation av olika slags attacker.

3) Kriterier för *entretien* (underhåll eller upprätthållande; *sustain* på engelska) är *Grain* (gryn eller korn) och *Allure* (som kan betyda fart, takt, gång; sätt att gå, framträdande eller utseende — det får stå oöversatt!). *Entretien* är den energi som upprätthåller ett ljud över tid, men det ska inte förstås som den fysiska kausaliteten bakom ljudet. Perkussiva ljud saknar ett sådant underhåll, och ljud av typen attack-resonans sägs ha ett passivt upprätthållande. *Allure* är en mer eller mindre regelbunden svängning eller fluktuation, som i ett vibrato. I själva verket är det en generalisering av vibrato, och innefattar variationer av andra aspekter än tonhöjd. Man skiljer mellan tre typer av *allure*: regelbunden, som orsakad av en mekanisk process; levande eller periodisk, men med en viss grad av variation som av en mänsklig utövare; samt naturlig, med en oförutsägbar oregelbundenhet. Grynighet eller kornighet har att göra med materialets mikrostruktur, och upplevelsen av kornighet går igen i de tre sinnesmodaliteterna syn, känsel och hörsel. Iterationer och olika variationer i hög hastighet ger upphov till grynighet i ljudet; liksom en snabb *allure*. Fagottoner har till exempel en kornighet som går från grov i det lägsta registret till fin och knappt märkbar i det höga registret. Det finns tre typer av gryn: *grain de résonance* (resonansgryn) som i det täta myllret i resonansen från en cymbal; *grain de frottement* (friktionsgryn) förekommer i form av luftturbulensen i flöjttoner eller friktionen av en stråke mot en sträng; samt *grain d'itération* (iterationsgryn), bland annat trumvirvlar. I var och en av dessa typer urskiljs vidare tre klasser efter grynets grovhet.

4) Variationen har fått två kriterier: *Profil mélodique* och *Profil de masse*. Den melodiska profilen är precis vad man skulle kunna tro av namnet, konturen av en melodisk rörelse eller av ett glissando. En variation i den melodiska profilen berör hela ljudets massa, medan en förändring i

massans profil bara berör det inre av ljudet, exempelvis formantglissandi eller övergångar mellan vokaler sjungna på samma ton. Men en variation i massa kan också röra sig mellan en tonisk och en komplex klang.

Vi har bara skrapat på ytan av Schaeffers typomorfologi och förhoppningsvis gett en antydan om hur den är uppbyggd (Godøy (1997) ger en översiktstabell med engelska motsvarigheter till Schaeffers begrepp). Det viktigaste är att den kan fungera som ett frågeformulär att ta med sig när man ska förhöra sina ljudobjekt. Det är bara genom att jämföra ljud med varann som man kan urskilja vad som är konstant och vad som varierar dem emellan. Om man väljer en begränsad kollektion ljud att studera, så blir de jämförelsekriterier man använder desto mera finkalibrerade. Det finns alltså alltid plats att hitta på sina egna kriterier och ge dem finurliga namn.

Schaeffers arbete har helt klart inspirerat några efterföljare. Bland annat märker man en influens hos Trevor Wishart. Lasse Thoresens sonologi bygger i allt väsentligt på Schaeffers idéer, men vidareför dem till musikaliska strukturer och mera storskaliga förlopp (Thoresen 1988). Denis Smalleys *spectromorphology* bygger också på Schaeffers tänkande (Smalley 1997). Det den tillför är framför allt en närmare granskning av typer av variation och rörelse samt en spatiomorfologi — en klassifiering av ljudens rörelse i rummet. Smalleys spektromorfologi är inte riktigt lika generell som Schaeffers system, den är nämligen först och främst avsedd för analyser av elektroakustisk musik (man kan till och med säga akusmatisk musik). Schaeffer hade å andra sidan en tanke om att typologin skulle kunna användas på alla slags ljud. En tillämpning han nämner är inom musiketnologin, där ett känt problem är att vår västerländska notskrift inte är avpassad för andra slags tonsystem, och inte nödvändigtvis förmår fånga de pertinenta dragen i en annan musikform.

## Sammanfattning

Vi såg i början av kapitlet att medvetandet är fördröjt med omkring en halv sekund, vilket kan antas ha ett samband med vissa emergenta egenskaper i perceptionen. Under rubriken Anamorfoser beskrevs några av de icke-lineära förhållandena mellan akustisk signal och perception, som de mellan frekvens och tonhöjd, amplitud och styrkegrad, samt tid och varaktighet. Fysiska parametrar samverkar och motsvarar inte helt inbördes oberoende perceptuella dimensioner. Somliga ljud som kan uppfattas isolerat försvinner genom maskering om andra ljud hörs samtidigt.

Uppmärksamheten är utan tvivel viktig för hur och vad vi hör, och har en stor betydelse för urskiljandet av olika aspekter av ljud. Schaeffers ljudobjekt är inte en godtyckligt avgränsad enhet, utan bestäms av grupperingsprinciper utforskade genom ASA. De sinsemellan tämligen olikartade ljudmodellerna som har nämnts svarar mot behov att beskriva och klassifiera ljud. Men poängen med diskussionen om ljudmodeller har varit att visa på metoder att utforska eller utfråga ljud. Det är en central del i arbete med ljudsyntes. Man kan, som Godøy (1997) kalla det en hermeneutisk aktivitet, där man växlar mellan att variera ljudets kontrollparametrar och att lyssna på det för att urskilja nya nyanser. Denna process har inget naturligt slut, utan det finns alltid mer att höra efter.

## Litteratur

- Bregman, A. S. (1990). *Auditory Scene Analysis*. Cambridge, Mass. & London: The MIT Press.
- Bregman, A. S. (1998). Psychological Data and Computational ASA. In Rosenthal, D. F. and Hiroshi G. O. (eds.), *Computational Auditory Scene Analysis* (pp. 1-12). Mahwah, New Jersey & London: Lawrence Erlbaum Associates, Publishers.
- Chion, M. (1983). *Guide des Objets Sonores*. Pierre Schaeffer et la Recherche Musicale. INA-GRM Éditions Buchet/Chastel.
- Eddins, D. A. & Green, D. M. (1995). Temporal Integration and Temporal Resolution. In Moore, B. C. J. (ed): *Hearing* (pp. 207-242). San Diego: Academic Press.
- Godøy, R. I. (1997). *Formalization and Epistemology*. Oslo: Universitetsforlaget (Scandinavian University Press).
- Grantham, D. W. (1995). Spatial Hearing and Related Phenomena. In Moore, B. C. J. (ed), *Hearing* (pp. 297-346). San Diego: Academic Press.
- Handel, S. (1995). Timbre Perception and Auditory Object Identification. In Moore, B. C. J. (ed), *Hearing* (pp. 425-461). San Diego: Academic Press.
- Holopainen, R. (1998). Örats historia. *Parergon* 3, 78-92.
- Houtsma, A. (1995). Pitch Perception. In Moore, B. C. J. (ed), *Hearing* (pp. 267-296). San Diego: Academic Press.
- Kendall, G. S. (1995). The Decorrelation of Audio Signals and Its Impact on Spatial Imagery. *Computer Music Journal* 19 : 4, 71-87.
- Kendall, R. A. & Carterette, E. C. (1993). Identification and blend of timbres as a basis for orchestration. *Contemporary Music Review* 9, Parts 1 & 2, 51-67.
- Lundström-Holmberg, E. & af Trampe, P. (1987). *Elementär fonetik*. Lund: Studentlitteratur.
- Moore, B. C. J. (1995). Frequency Analysis and Masking. In Moore, B. C. J. (ed), *Hearing* (pp. 161-206). San Diego: Academic Press.
- Moore, F. R. (1990). *Elements of Computer Music*. New York: Prentice Hall.
- Schaeffer, P. (1977). *Traité des objets musicaux*. Paris: Éditions du Seuil.
- Slaney, M. (1998). A Critique of Pure Audition. In Rosenthal, D. F. and Hiroshi G. O. (eds.), *Computational Auditory Scene Analysis* (pp. 27-41). Mahwah, New Jersey & London: Lawrence Erlbaum Associates, Publishers.
- Smalley, D. (1997). Spectromorphology: explaining sound-shapes. *Organised Sound* 2 : 2, 107-126.
- Sundberg, J. (1989). *Musikens ljudlära*. Stockholm: Proprius förlag.
- Tervaniemi, M. (1999). Pre-attentive Processing of Musical Information in the Human Brain. *Journal of New Music Research* 28 : 3, 237-245.
- Thoresen, L. (1988). Auditive Analysis of Musical Structures. A summary of analytical terms, graphical signs and definitions. *ICEM Conference on Electro-acoustic Music Stockholm 1985 Proceedings* (pp. 65-90). Royal Swedish Academy of Music.
- Wallin, N. L. (1982). Den musikaliska hjärnan. En kritisk essä om musik och perception i biologisk belysning. *Kungl. Musikaliska Akademiens skriftserie: 34(1982) Skrifter från musikvetenskapliga institutionen, Göteborg: 6 (1982)*
- Xenakis, I. (1992). *Formalized Music. Thought and Mathematics in Music. (Revised Edition.)* Stuyvesant: Pendragon Press.
- Yates, G. (1995). Cochlear Structure and Function. In Moore, B. C. J. (ed), *Hearing* (pp. 41-74). San Diego: Academic Press.