Akustikk for musikere

Øyvind Hammer
Norsk nettverk for Teknologi, Akustikk og Musikk

Oslo, 10. mai 1995

Innhold

Kapittel 1

Innledning

Hvis man skal ha håp om å kunne produsere lyd med en datamaskin, er det en forutsetning at man har en grunnleggende forståelse for hva lyd er, hvordan den oppstår og hvordan den påvirkes av fysiske hindringer. Læren om disse forholdene heter *akustikk*. Innenfor dette feltet har vi også andre spesielle fagområder som er av interesse for oss: Musikkakustikk, romakustikk, fonetikk, elektroakustikk etc. Det er dessuten nødvendig å vite hvordan forskjellige aspekter ved lyden oppfattes av mennesket (psykoakustikk).

Dette kapittelet vil gi en introduksjon til disse tingene.

Kapittel 2

Hva er lyd?

Lyd er hurtige variasjoner i lufttrykk, som brer seg som bølger med en hastighet på 343 meter i sekundet (1235 km/t) ved romtemperatur. Fra en punktformet lydkilde i rommet brer lyden seg utover i alle retninger, slik at bølgefronten består av et stadig større kuleskall. Arealet av et slikt kuleskall øker med kvadratet av radien i kula. [1] Dette betyr at lyden får en stadig større luftflate å bevege, og siden det ikke tilføres noe energi må det innebære at lydstyrken svekkes tilsvarende. Derfor vil lydstyrken falle med kvadratet av avstanden.

Som vi ser er luft (eller en annen gass) nødvendig for at lyd skal kunne forplante seg. Det nytter lite å rope til hverandre på månen, og drønnene vi kan høre når romskip eksploderer i dårlige science-fiction-filmer er lite realistiske, selv om sjokkbølger fra ekspanderende gasser kanskje kan gi et svakt plopp.

Vanlige overflatebølger på vann er *transversale*, det vil si at bølgebevegelsen (opp og ned) går på tvers av utbredelsesretningen (bortover vannflaten). Lydbølger er derimot *longitudinale*, hvilket betyr at bølgevegelsen (fram og tilbake) går langs samme retning som utbredelsesretningen.

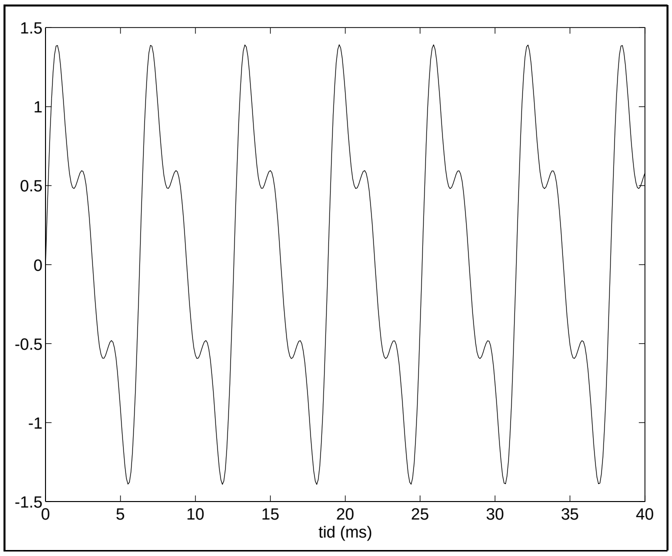
Vanligvis tegner vi likevel lydbølgene som om de var transversale. Vi får da en graf som kalles et *oscillogram*, der tiden går langs x-aksen og lufttrykket langs y-aksen. Grafen er normalisert slik at det "normale" lufttrykket kalles null. Oscillogrammer kan vises på en billedskjerm ved hjelp av et *oscilloskop*.

[1] $A = 4\pi r^2$

Kapittel 3

De tre lyd-parametrene

Ved å betrakte et oscillogram kan vi si endel om lydens egenskaper. Dersom vi har en statisk lyd, dvs. at den ikke endrer seg over måletidsrommet, kan oscillogrammet se ut f.eks. som i figur 3.1.



Figur 3.1: Statisk lyd

Tre grunnleggende parametre bestemmer hvordan en lyd oppfattes av øret: Amplityde, frekvens og klangfarge.

3.1 Amplityde, lydstyrke

Amplityden er graden av utslag, og leses av som maksimalverdier i oscillogrammet. Amplityden til lyden over er ca. 1.4. Lydstyrken vi vil oppfatte avhenger av amplityden: Større amplityde gir kraftigere lyd. Nå har det seg imidlertid slik at øret oppfatter lydstyrke logaritmisk, slik at sammenhengen mellom lydstyrke og amplityde ikke er lineær. Vi bruker derfor gjerne enheten *desi-Bel* (etter Alexander Graham Bell) som mål for amplityde. Når det gjelder akustisk lyd, går vi fram på følgende måte. Først bestemmer vi oss for at den svakeste lyden vi kan høre skal kalles 0 dB (dette er definert som $10^{-16} W/cm^2$ for en sinustone på 2 kHz). Amplityden til denne lyden kan vi kalle a_{ref}, fordi den brukes som en referanse som vi sammenlikner andre amplityder med. En lyd med amplityde a har nå følgende styrke målt i dB [1]:

$$A = 20 \cdot log_{10} \frac{a}{a_{ref}}.$$

Når vi benytter dB-skalaen, kan vi si endel kvantitative ting om hvordan øret oppfatter lydstyrke. Den minste endring i lydstyrke som vi kan merke er ca. 1 dB. Den kraftigste lyd vi kan høre uten at vi får vondt er på ca. 120 dB. Vi sier at ørets *dynamiske område* er 120 dB. Dette tilsvarer et amplityde-område på en million til en! Det er svært vanskelig å lage elektronikk som kan håndtere et så stort spenningsområde. Profesjonelt lydutstyr har oftest et dynamisk område på under 100 dB.

En viktig konsekvens for oss data-musikere er at vi ikke kan øke amplityden lineært og så regne med at vi får en naturlig crescendo. Vi må istedet bruke en eller annen eksponensiell funksjon.

3.2 Frekvens, tonehøyde

I oscillogrammet over kan vi se en periodisitet; den samme kurveformen gjentar seg igjen og igjen. Slike lyder oppfattes som en kontinuerlig tone. Antall gjentakelser pr. sekund gir frekvensen til den grunntonen vi oppfatter, målt i Hertz. Har vi en grunntone på f.eks. 110 Hz, betyr dette at den periodiske kurveformen gjentar seg selv 110 ganger i sekundet. Dersom vi dobler til 220 Hz, hører vi dette som et sprang en oktav opp. Videre dobling til 440 Hz gir enda en oktav. Dette innebærer at heller ikke sammenhengen mellom frekvens og oppfattet tonehøyde er lineær. Jo høyere vi går opp i registret, jo lengre avstand er det mellom halvtonetrinnene, målt i Hz. For å gå opp et halvtonetrinn, kan vi ikke legge til en fast frekvens. Vi må istedet multiplisere frekvensen med en faktor. Når vi har 12 halvtonetrinn i en oktav er denne faktoren 12-roten av 2, eller 1.05946. Musikalske skalaer vil bli omtalt senere.

Vi har sett at både amplityde og frekvens oppfattes logaritmisk av øret. Det viser seg at de fleste sanser virker på samme måte: Multipliserer vi et stimulus med en faktor, oppfattes dette som et fast tillegg. Dette gjelder vår oppfattelse av lys, varme, smak, lukt, tid osv. At sammenhengen mellom stimuli og persepsjon er logaritmisk, (*Fechners lov*) ble først påstått av G. T. Fechner i boka *Elements of Psychophysics* i 1860.

Hvis kurveformen *ikke* er periodisk, kan dette bety to ting. Enten er kurven en sum av to eller flere periodiske kurver med frekvenser som ikke står i enkle forhold til hverandre. Kurven kalles da *kvasiperiodisk*, og vil oppfattes som en dissonant akkord. Dette vil bli omtalt senere. Den andre muligheten er at kurven er "bare rot", og ikke er en enkel sum av

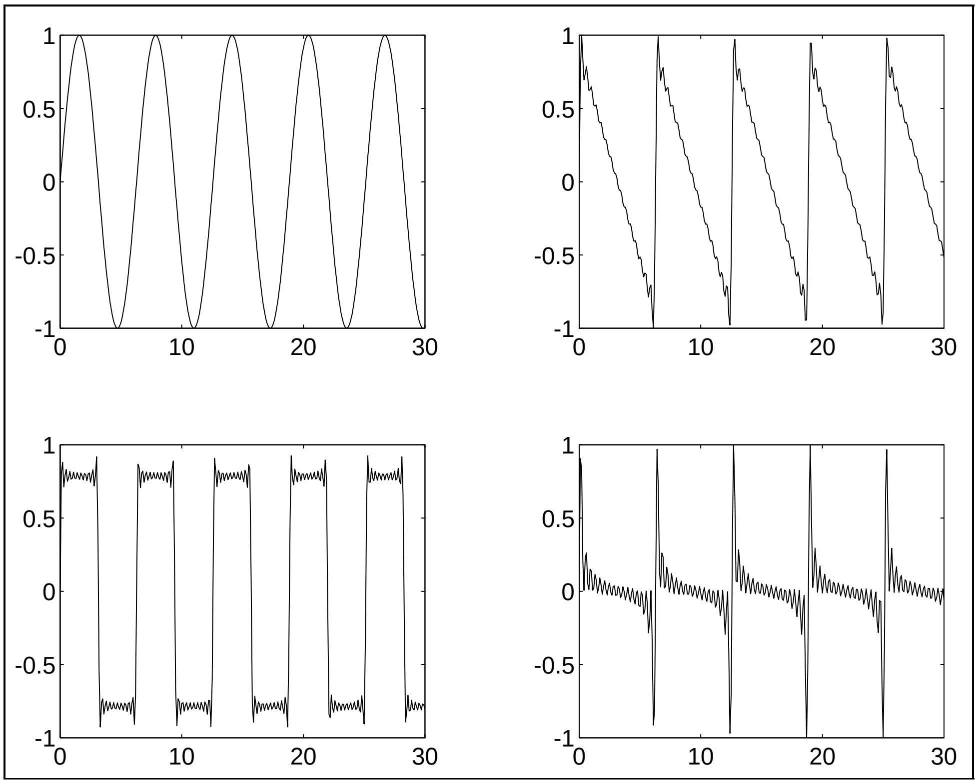
[1] Faktoren 20 ser litt tilfeldig ut. dB-skalaen er egentlig tenkt brukt i forbindelse med *effekter*, ikke amplityder. Effekten er proporsjonal med kvadratet av amplityden. Faktoren blir derfor 10 hvis vi snakker om effekter. Hvis vi i tillegg snakker om Bel istedenfor desi-Bel, blir faktoren 1, i analogi med f.eks. liter og desiliter.

et lite antall periodiske kurver. Dette vil oppfattes som en susing eller støylyd uten noen bestemt tonehøyde (unpitched).

3.3 Klangfarge

Den tredje parameteren, klangfargen, er en svært kompleks ting, og bestemmes av lydkurvens *form.*

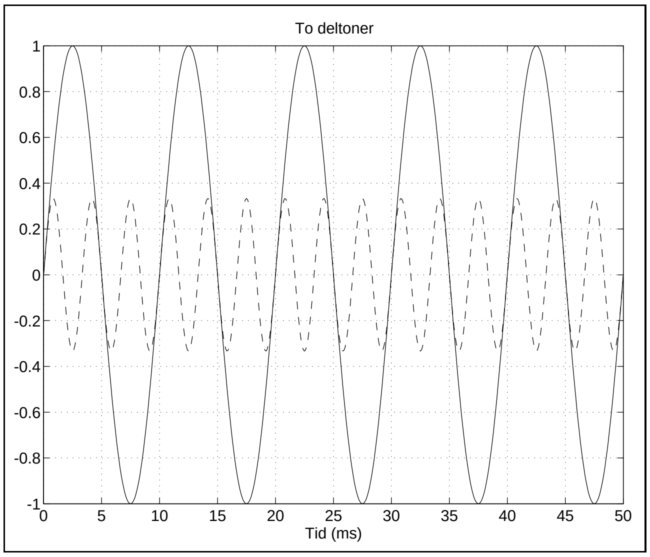
Figur 3.2 viser fire ulike periodiske og statiske lyder. De har samme frekvens og amplityde, og vil alle oppfattes som kontinuerlige toner med samme tonehøyde og omtrent samme lydstyrke. Men kurvene har ulik form. Vi ser en sinustone, en sagtanntone, en firkanttone og et pulstog. Dette gir ulik klangfarge for de forskjellige lydene.



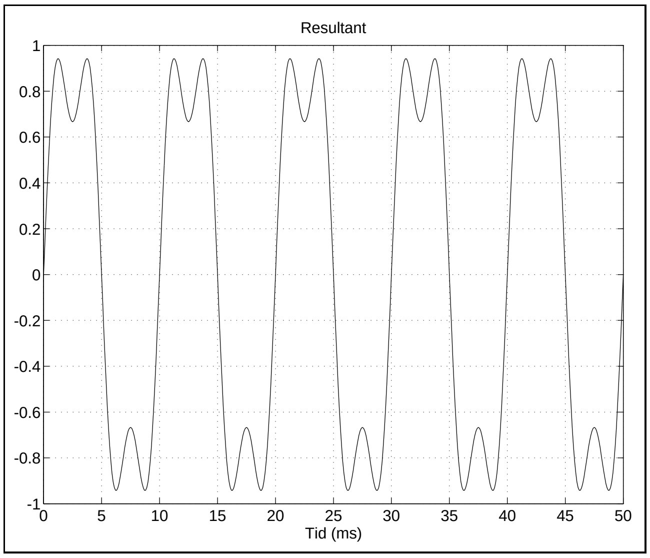
Figur 3.2: Forskjellige kurveformer

Matematikeren Fourier viste at enhver repeterende kurveform kan genereres ved å addere (mikse) et antall sinusoide deltoner, hver med en frekvens som er et heltallsmultiplum av grunnfrekvensen. Figur 3.3 viser prinsippet. Her adderes en sinusformet grunntone med frekvens 100 Hz til en overtone med frekvensen 300 Hz, dvs. en oktav og en kvint over. Resultatet vises i figur 3.4.

De forskjellige deltonene i en tone kalles ofte de *harmoniske*. Grunntonen heter da den første harmoniske, første overtone på to ganger grunnfrekvensen heter andre harmoniske, andre overtone på tre ganger grunnfrekvensen heter tredje harmoniske etc. Som man ser er terminologien litt forvirrende.



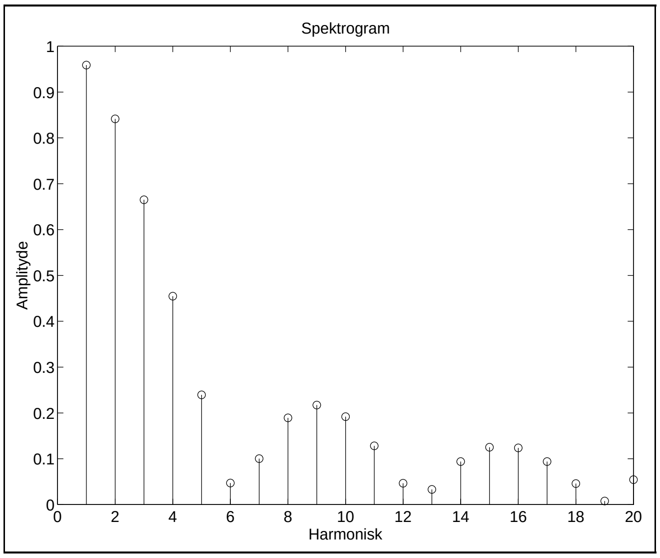
Figur 3.3: Grunntone og andre overtone



Figur 3.4: Grunntone og andre overtone mikset sammen

I figur 3.3 begynte begge deltonene på samme sted i sinuskurven, dvs. at deltonene har samme fase. Hvis deltonene har ulike faser, vil den resulterende kurveformen naturligvis se annerledes ut. Men øret oppfatter i utgangspunktet ikke faseinformasjon, det oppfatter bare amplityden til de forskjellige deltonene. Forholdet mellom disse amplitydene bestemmer klangfargen. Vi kan med andre ord ha to lyder som ser radikalt ulike ut på et oscilloskop, fordi de har ulik faseinformasjon, men som låter nøyaktig likt, fordi deltonenes amplityder er de samme.

Dersom vi er interessert i klangfarge, er det derfor mest praktisk å beskrive lyden i et stolpediagram der deltonenes amplityder er vist som funksjon av frekvensen. En slik figur kalles et spektrogram, og kan avbildes ved hjelp av et måleinstrument som kalles en spektralanalysator eller *spektrograf*. Et spektrogram kan f.eks. se ut som i figur 3.5. Overtonene er vanligvis svakere enn grunntonen. Legg også merke til de to "toppene" rundt den 9. og den 15. harmoniske. Slike topper er typiske i naturlig lyd, og kalles *resonanser* (fonetikerne bruker begrepet *formanter*).



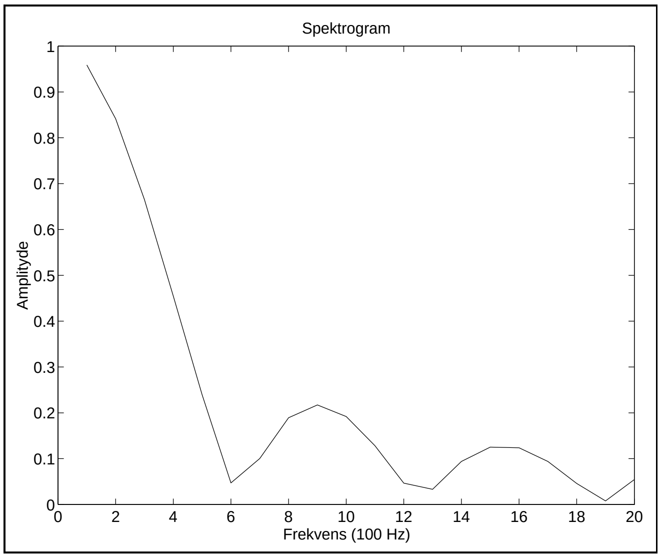
Figur 3.5: Spektrogram

Dette er altså en enkel tone med en veldefinert tonehøyde. Kurveformen er da som vi har sett periodisk, og spektrogrammet består av adskilte stolper med fast avstand lik grunntonens frekvens. Vi sier at spekteret er *diskret*.

Hvis lyden ikke er periodisk, og dermed uten tonehøyde, vil vi kunne få et *kontinuerlig* spektrum der alle mulige frekvenser er tilstede. Dette er typisk for støylyder, og spektrogrammet kan da se ut f.eks. som i figur 3.6.

Dersom en lyd vises som et oscillogram, sier vi at lyden er representert i *tidsdomenet*. En lyd vist som et spektrogram er representert i *frekvensdomenet*. Det er viktig å forstå at disse to representasjonsformene inneholder den samme informasjonen, dersom vi ser bort fra faseforhold. [2] Omregningen mellom tidsdomenet og frekvensdomenet foretas ved hjelp av en spesiell regnemetode som kalles *Fouriertransformasjon*. Fouriertransformasjon er

[2] Det er imidlertid ikke uvanlig at spektrografer også kan vise "fasespekteret"



Figur 3.6: Kontinuerlig spektrum

sentral i digital lydbehandling. Omregningen fra tidsdomenet til frekvensdomenet kan vi også kalle *spektralanalyse*. Omregningen fra frekvensdomenet til tidsdomenet kan vi kalle *additiv syntese*, fordi vi adderer sammen alle de sinusoide deltonene i spekteret for å lage en lyd i tidsdomenet.

3.4 Dynamikk

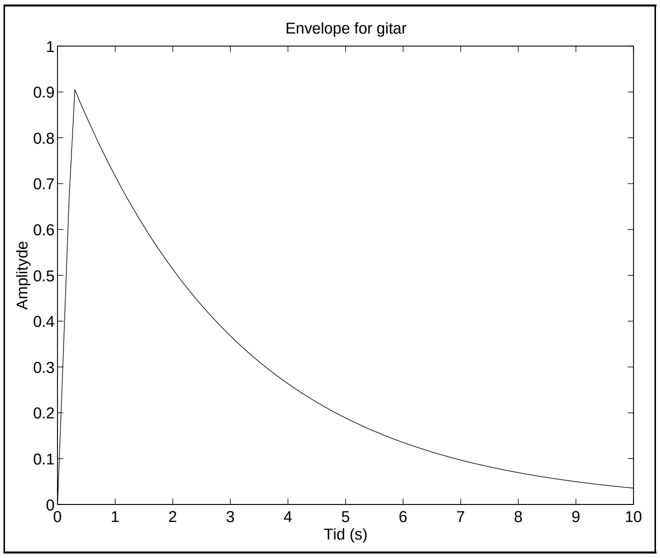
Hittil har vi betraktet lyd som ikke endrer seg over tid. Dette gir kontinuerlige toner. Skal vi lage musikk, må selvsagt de tre parametrene i lyden forandre seg hele tiden.

Dynamikk i amplityden

Dynamikk i amplityden er det første vi må få til. Alle musikkinstrumenter fungerer slik at amplityden forandrer seg i løpet av hver tone. En note på en gitar f.eks. starter med en svært hurtig øking av amplityden (attack) fulgt av en langsom utdøing (decay). Dette kan vises i en graf som figur 3.7.

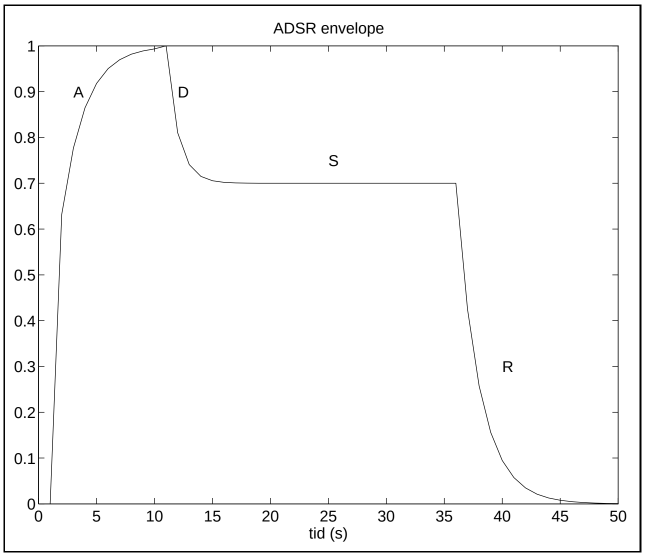
En slik graf kalles en *envelope*. [3] Når vi skal syntetisere lyd er det selvsagt ønskelig å kunne tegne enhver envelope, og dette er mulig med datamaskiner. Mange synthesizere begrenser seg imidlertid til en generalisert envelope-type med fire linjesegmenter. Dette heter *ADSR*, for Attack-Decay-Sustain-Release (figur 3.8). Knekkpunktene mellom segmentene kan plasseres fritt av brukeren. Bemerk at linjene ikke er rette, dette kommer av den ulineære sammenhengen mellom amplityde og oppfattet lydstyrke. Attack-biten er litt

[3] Det finnes et godt norsk ord, nemlig *omhyllingskurve*. Dette er imidlertid ikke mye brukt i musikkbransjen



Figur 3.7: Gitar-envelope

rar, og gir ikke en naturlig crescendo. Det viser seg imidlertid at naturlige instrumenter ofte oppfører seg på en liknende måte.



Figur 3.8: ADSR-envelope

Vi kan også la amplityden kontrolleres delvis av en lavfrekvent sinusoid. Dette lager en effekt som heter *tremolo*.

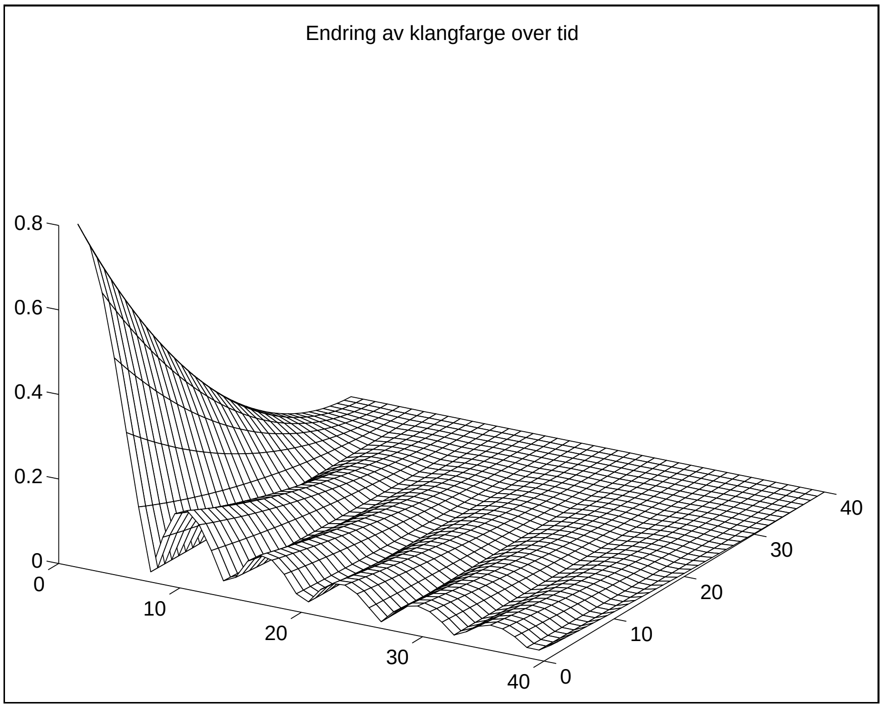
Dynamikk i frekvens

Dynamikk i frekvens er selvsagt nødvendig for å få fram forskjellige noter, men det er også vanlig å variere frekvensen kontinuerlig, enten fra en note til en annen (glissando), eller ved hjelp av en lavfrekvent sinusoid (vibrato).

Dynamikk i klangfargen

For at lyden ikke skal bli helt død, må klangen endres hele tiden. Denne kompliserte oppgaven forsøkes løst på ulike måter ved hjelp av forskjellige syntesemetoder.

Spektrogrammene vi har sett på viser klangen bare på ett bestemt tidspunkt. Når klangen varierer over tid, må vi bruke diagrammer som forsøker å vise lyden i tre dimensjoner (vi får en tidsakse i tillegg til frekvens- og amplitydeaksene). Dette kan gjøres i et "landskapsdiagram" der frekvensen går mot høyre og tiden går innover i bildet (figur 3.9).



Figur 3.9: Tredimensjonalt spektrogram

Slike diagrammer kan være vanskelige å lese. En representasjonsform som ofte er mer praktisk er *sonogrammet*, der tiden løper langs x-aksen, frekvensen langs y-aksen og amplityden vises som gråtoner.

Kapittel 4

Beats, kritiske bånd

I dette avsnittet skal vi se på et psykoakustisk fenomen som kalles *beats*. Dette er interessant i seg selv, men først og fremst er det meningen at dette stoffet skal gi en bedre forståelse for dualiteten mellom tids- og frekvensdomenet.

To samtidige sinusoider

Dersom vi spiller av to sinusoider samtidig, vil de mikses sammen. Miksing er det samme som at utslagene *summeres*. Denne summeringen (eller *superposisjonen*) kan skje internt i en miksepult eller en datamaskin, eller dersom sinusoidene avspilles fra separate høyttalere, vil summeringen først skje når lydbølgene blandes (*interfererer*) i luften. Vi har allerede sett et eksempel på miksing av sinusoider, i avsnittet om klangfarge.

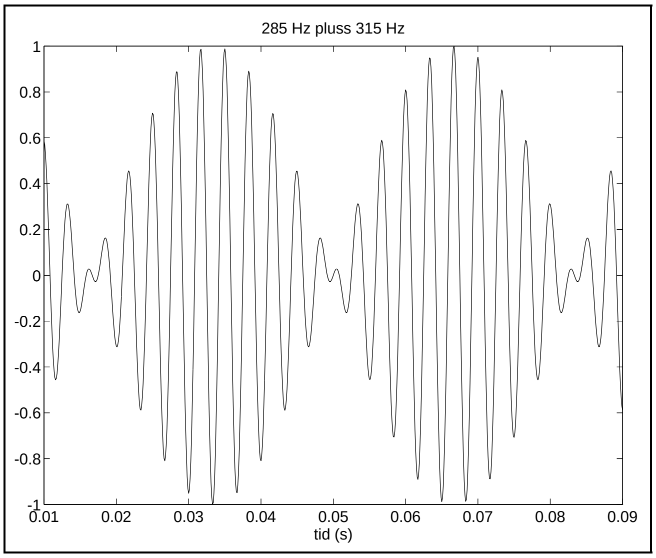
Hvis disse sinusoidene ligger langt fra hverandre i frekvens, vil vi høre dem som to separate toner. Siden øret analyserer lyden opp i sine enkelte frekvenskomponenter, kan vi si at vi her oppfatter lyden i frekvensdomenet.

Hvis vi nå skrur frekvensene nærmere hverandre, f.eks. til en forskjell på 30 Hz, vil vi etterhvert få en "sur" og "grumsete" lyd. Vi er kommet inn i det *dissonante overgang-området*. Oscillogrammet i figur 4.1 viser den resulterende lydbølgen.

La oss nå skru frekvensene enda nærmere, f.eks. slik at de ligger bare 10 Hz fra hverandre. Vi vil da ikke lenger oppfatte de to sinusoidene separat i frekvensdomenet. I stedet vil vi høre en enkel tone med frekvens som ligger midt imellom de to komponentene. Denne tonen vil pulsere med en frekvens på 5 Hz, som om vi hadde en enkelt tone hvis amplityde ble styrt av en lavfrekvent sinusoid med frekvens 5 Hz (tremolo eller ringmodulasjon). [1] Denne pulseringen kan tydelig sees på grafen over, men der var pulsfrekvensen så høy (15 Hz) at øret ikke fikk den med seg.

Det som nå er skjedd, er at vi begynner å oppfatte lyden mer i tidsdomenet, dvs. at vi kan høre pulseringen utstrakt i tid, som *beats*. Det er viktig å forstå at dette er en egenskap ved *øret*; det er ikke skjedd noen kvalitativ endring av selve lyden i og med at vi skrudde

[1] Vi vil faktisk høre 10 pulser i sekundet, fordi vi får en amplitydetopp både for den positive og den negative halvperioden til den modulerende sinusoiden. Den sanne periodisiteten er imidlertid 5 Hz, fordi annenhver topp er "snudd opp ned" (denne fotnoten er sikkert bare forvirrende).



Figur 4.1: Beats

sinusoidene stadig nærmere hverandre. På et spektrogram vil den pulserende sinusoiden vises som de to tette frekvenskomponentene som lyden vitterlig består av.

Litt tallmagi

Hvis vi vil lage en sinusoid med frekvens f_s som pulserer med frekvensen f_p, kan vi altså *enten* tenke i frekvensdomenet og addere to sinusoider, slik:

$$f(t) = \frac{1}{2}[cos(2\pi(f_s + f_p)t) + cos(2\pi(f_s - f_p)t)],$$

(sinusoidene ligger da $2f_p$ Hz fra hverandre), *eller* vi kan tenke i tidsdomenet og modulere en sinusoid med en annen, lavfrekvent sinusoid, slik:

$$g(t) = cos(2\pi f_p t)cos(2\pi f_s t).$$

Hvilken måte vi tenker på er ett fett, $f(t) = g(t)$. For å vise dette gjør vi som alltid når vi trenger et trøstens ord i hverdagen: Slår opp i Rottmanns "Mathematische Formelsammlung". Der finner vi at

$$\frac{1}{2}[cos(a + b) + cos(a - b)] = cos(a)cos(b).$$

Dette er forøvrig nyttig kunnskap for elektronmusikeren, fordi det forteller hvordan et spektrum forandres når vi utsetter lyden for ringmodulasjon. Dette kommer vi til i den digitale lydbehandlingen.

4.1 Kritiske bånd

Vi har sett at når to sinusoider nærmer seg hverandre, vil de bli vanskeligere å skille fra hverandre. Til slutt vil de smelte helt sammen til en enkelt, pulserende sinusoid. Eksakt når slutter vi å oppfatte de separate deltonene? Målinger viser at dette inntrer når sinusoidene kommer innenfor samme *kritiske bånd*. Et kritisk bånd er et frekvensområde som mer eller mindre korresponderer med et avgrenset område på ørets *basiliærmembran*.

Basiliærmembranet er en lang "tunge" som er kveilet opp inne i sneglehuset (cochlea) i det indre øret. Forskjellige områder på dette membranet reagerer på forskjellige frekvenser, og membranet fungerer derfor som første ledd i den kompliserte spektralanalysen som øret foretar. Sansehårene for de laveste frekvensene ligger lengst inne, mens sansehårene for de høyeste frekvensene ligger ytterst ut mot mellomøret. [2]

De kritiske båndene ligger spredt oppover med ca. 1/3 oktavs mellomrom, og overlapper kraftig slik at det ikke blir noen hull i det spekteret vi kan høre.

[2] At de "lyse" sansehårene ligger mest ubeskyttet er nok noe av årsaken til at vi mister de høye frekvensene først når øret skades av kraftige lyder.

Kapittel 5

Spektre og tonalitet

I dette avsnittet skal vi se mer på komplekse lyder og kritiske bånd. Vi skal se at dette gir oss en rimelig god forklaring på fenomenet tonalitet.

5.1 Overtoner og kritiske bånd

Vi har sett at vi får en "grumset" lyd når to sinusoider ligger innenfor samme kritiske bånd. Dette forklarer hvorfor f.eks. C og Ciss (halvtonetrinn) gir dissonans. Tersen er såvidt et stort nok intervall til at vi får konsonans. Men hvorfor er da f.eks. C4 og Ciss5 dissonant? Dette intervallet er jo større enn en ters?

Vi kan forklare dette ved at vi får kollisjon mellom *overtonene*. Grunntonene C4 og Ciss5 ligger langt fra hverandre, og gir ikke dissonans. Første overtone til grunntonen C4 er imidlertid C5, og den ligger innenfor samme kritiske bånd som grunntonen Ciss5.

En naturlig følge av denne diskusjonen skulle være at to sinustoner alltid vil låte konsonant, unntatt hvis de kommer for nær hverandre. Sinustoner, som bare har en grunntone, kan jo ikke gi kollisjon i overtonene. I litteraturen påstås det at dette faktisk er tilfelle, men det er nok vanskelig å verifisere dette. For det første vil ulineæriteter alltid føre til at vi får enkelte svake overtoner, og for det andre vil enhver musikalsk trenet person umiddelbart gjenkjenne et dissonant intervall, og dermed oppfatte lyden som dissonant.

5.2 Musikalske skalaer

Dersom vi ønsker konsonans, bør frekvensene til grunntonene stå i *enkle tallmessige forhold til hverandre.* På den måten unngår vi kollisjoner i overtonene. Forholdet 2:1 (f.eks. 500 Hz og 250 Hz) er veldig konsonant (oktav), mens forholdet 200:141 er dissonant. Andre konsonante intervaller er 3:2 (kvint), 4:3 (kvart), 5:3 (sekst), 5:4 (ters), 6:5 (liten ters) etc.

Hvis vi vil ha en skala der frekvensene står i enkle forhold til hverandre, kan vi gå fram f.eks. slik for C-dur: Vi kaller frekvensen til C for 1 (dette kan skaleres senere). Vi ønsker at C, E og G skal stå i forholdet 4:5:6, og setter derfor opp E og G som $\frac{5}{4}$ og $\frac{3}{2}$ ($1:\frac{5}{4}:\frac{3}{2}=4:5:6$). Deretter vil vi sette en tilsvarende akkord G,H,D oppå G (dominanten), fremdeles i forholdet 4:5:6. Hvis vi sender D ned i samme oktav får vi etter litt regning at H havner på $\frac{15}{8}$ og D på $\frac{9}{8}$. Til slutt lager vi en treklang på subdominanten (F,A,C) i forholdet 4:5:6, og vi ender opp med den *rettstemte* skalaen (just intonation):

$C = 1, D = \frac{9}{8}, E = \frac{5}{4}, F = \frac{4}{3}, G = \frac{3}{2}, A = \frac{5}{3}, H = \frac{15}{8}, C = 2.$

Hvis vi nå sier at C skal være 100 Hz, vil f.eks. G havne på 150 Hz. Det leie er imidlertid at dersom vi bruker en annen utgangsnote (f.eks. D) når vi konstruerer skalaen vår, vil notene havne på litt andre frekvenser. Kvinten over D, altså A, vil f.eks. plasseres på $100 \cdot \frac{27}{16} = 168.75$ Hz, og det er ikke det samme som A'en vi hadde i C-dur, nemlig $100 \cdot \frac{5}{3} = 166.67$ Hz. Dette betyr at vi må stemme om instrumentene for hver gang vi skifter toneart. Meget upraktisk.

Den mest brukte skalaen er den *tempererte* ("equal temperament"), som lar alle halvtonetrinn være like store. Man passer på at oktaven har forholdet 2:1, men ellers legger man ikke vekt på enkle tallforhold. Heldigvis treffer frekvensene noenlunde riktig likevel. Hvis vi starter med C på 100 Hz, går vi til Ciss ved å multiplisere med 12-roten av 2, eller 1.05946. Vi havner da på 105.946 Hz. Vi lager D ved å multiplisere med 1.05946 enda en gang, og får 112.246 Hz. Kvinten G blir da 149.8 Hz, ikke langt fra 150 Hz som vi hadde over. Etter å ha multiplisert med 12-roten av 2 tolv ganger, har vi gått opp en oktav og havner på eksakt 200 Hz.

Det finnes som kjent mange andre skalaer (pythagoreisk, meantone etc.), og dette er i det hele tatt et stort felt. Vi skal ikke gå mer i detalj om skalaer her.

5.3 Harmoniske spektre, cocktailparty-effekten

Dersom et spekter bare inneholder en grunnfrekvens og heltallsmultipla av denne (100 Hz, 200 Hz, 300 Hz osv.), vil øret oppfatte dette som en enkelt lyd med grunntone og sine overtoner, som kommer fra en enkelt lydkilde. Dette kalles et *harmonisk spektrum.*

Dersom et spekter inneholder frekvenser som ikke passer inn i dette, vil hjernen forsøke å finne andre grunnfrekvenser og overtoner som kan være tilstede, og gruppere frekvensene sammen i harmoniske klasser. Hver slik harmonisk klasse vil tolkes som en separat lydkilde.

Det er utrolig hvordan vi på denne måten kan splitte et komplisert spektrum opp i enkelte lydkilder. Fra et hav av skravling i et selskap kan vi skille ut og konsentrere oss om en enkelt stemme. Dette kalles *cocktailparty-effekten.* Denne effekten kan observeres i ekstrem grad på Sør-Georgia i Sør-Atlanteren. Når en kongepingvin kommer inn fra havet, kan den raskt lokalisere ungen i en koloni på flere hundre tusen individer, selv om alle ungene er i konstant bevegelse rundt i kolonien. Forsøk har vist at foreldrene kjenner igjen ungen på skrikene, og dette skjer altså i en øredøvende kakafoni der kanskje hundre tusen unger skriker samtidig. En slik utrolig signalbehandling vil vi neppe klare å etterlikne på våre datamaskiner på en god stund ennå.

5.4 Virtuelle grunntoner

Hvis hjernen synes at det "burde vært" en grunntone på en viss frekvens, vil vi kunne oppfatte en grunntone selv om det ikke finnes noen. Dersom vi f.eks. lager oss et spektrum med frekvensene 200 Hz, 300 Hz, 400 Hz etc., vil vi oppfatte en *virtuell grunntone* på 100 Hz, og dette gir den pitch vi vil høre.

Bjeller og klokker har ofte inharmoniske spektre der frekvensene står i svært enkle forhold til hverandre. Dette produserer gjerne virtuelle grunntoner.

En annen illustrasjon av dette fenomenet har vi for de dypeste tangentene på et piano. Grunntonen til disse tangentene ligger helt nede under 30 Hz, og dette er helt på grensen av det vi kan høre. Vi kan imidlertid høre overtonene helt utmerket, og ut fra dette konstruerer vi oss en grunntone.

Kapittel 6

Musikkinstrumentenes fysikk

Vi kan ikke gå i noen detalj om hvordan musikkinstrumentene fungerer her, men det er viktig å forstå *eksitator-resonator-modellen* som gir en viss forståelse av mange forskjellige instrumenter.

6.1 Gitar

Når vi klimprer en gitarstreng med lengde L meter, vil det forplante seg en bølge til begge sider i strengen. Denne bølgen reflekteres i endene, og farer fram og tilbake. Vi får dermed en svingning som kunne ha vart til evig tid, hvis det ikke var for at friksjonen demper bevegelsene etterhvert.

Hva slags svingninger kan vi så få? Siden strengen er spikret fast i endene, må det kreves at posisjonen her er null, og dessuten at strengen står stille her. [1] Dette gir såkalte *stående bølger*. Strengens form på ethvert tidspunkt kan beskrives som en sum av sinusoider. [2] Den første mulige sinusoiden som er null i begge ender, har bølgelengde lik to ganger strengens lengde ($\lambda = 2L$), og vi får en bevegelse opp og ned overalt på strengen, men mest på midten (figur 6.1).

Dette er en halv sinusbølge, og denne halvdelen har altså en lengde $\frac{\lambda}{2} = L$. Det er derfor vi har at $\lambda = 2L$.

Hvilken frekvens vi får, vil foruten bølgelengden avhenge av lydhastigheten gjennom strengen, som igjen avhenger av strengens tykkelse, masse og stivhet. Sammenhengen mellom frekvens, fart og bølgelengde er $f = \frac{v}{\lambda}$.

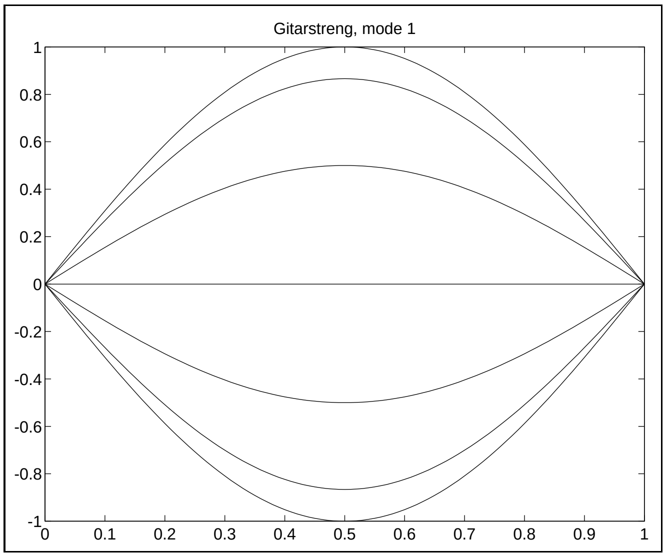
Den andre mulige sinusoiden har bølgelengde lik strengens lengde ($\lambda = L$) og vi får en bevegelse opp og ned med en stillestående *knute* på midten (figur 6.2).

Siden bølgelengden her er halvparten av den over, vil denne svingningen ha dobbelt så høy frekvens.

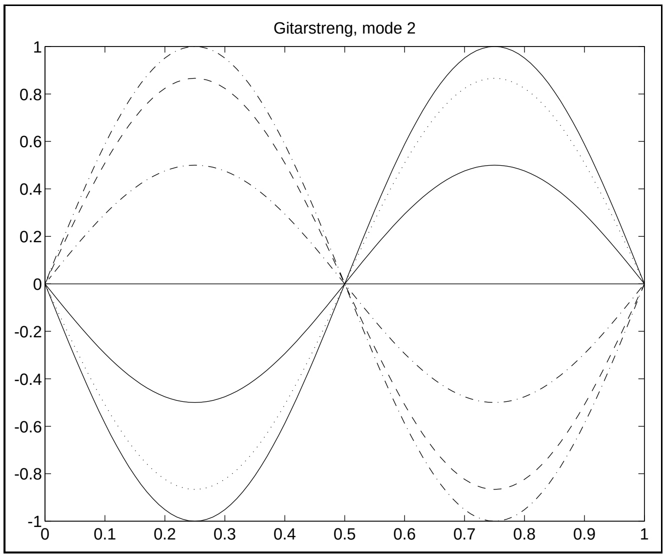
Vi kan også ha svingninger med bølgelengde lik halvparten av strengelengden, en tredel osv. Disse bølgelengdene vil gi en grunnfrekvens (mode 1) med overtoner (mode 2, 3 etc.). Alle disse svingningene vil være tilstede på en gang, og summere seg sammen til et helt

[1] Dette heter *homogene randbetingelser* i matematikken

[2] Hvis vi tenker oss at vi utvider strengen periodisk, følger dette av Fouriers teorem



Figur 6.1: Grunnsvingning på gitarstreng



Figur 6.2: Første overtone på gitarstreng

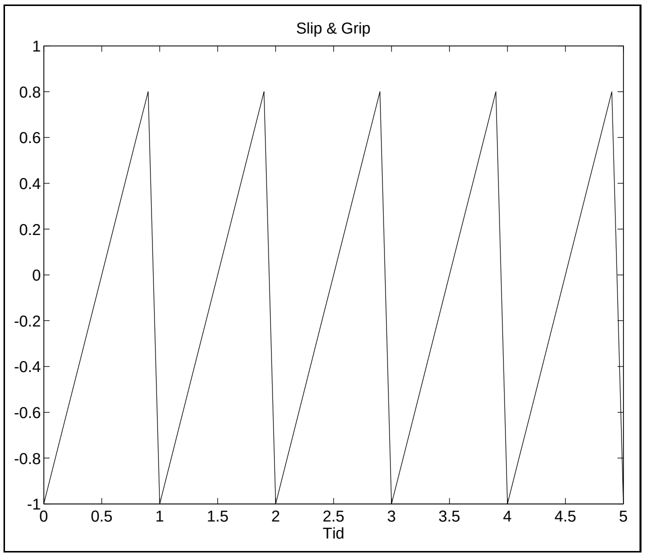
spektrum med grunntone og overtoner. Hvis strengens lengde varierer, vil bølgelengdene variere tilsvarende, og vi får ulike grunnfrekvenser.

Spekteret fra strengen farges videre av gitarkassen, som virker som et filter med forskjellige *resonanser*, dvs. at visse frekvensområder framheves.

I dette tilfellet kan vi kalle strengen for en *eksitator* (lydprodusent), og gitarkassen for en *resonator*.

6.2 Fiolin

Etter at vi har klimpret en gitarstreng svinger den fritt. På en fiolin derimot, opprettholdes svingningen av buens bevegelse. Dette kan vi kalle *tvungen svingning* (forced vibration). Denne mekanismen fungerer ved at friksjonen mellom buen og strengen gjør at strengen dras med buen et lite stykke. På et visst tidspunkt blir imidlertid motkraften for stor, og strengen spretter momentant tilbake. Så tar buen tak i strengen igjen, og prosessen gjentar seg. Fysikere kaller denne prosessen for *slip and grip*. Vi får dermed en slags sagtannkurve (figur 6.3).



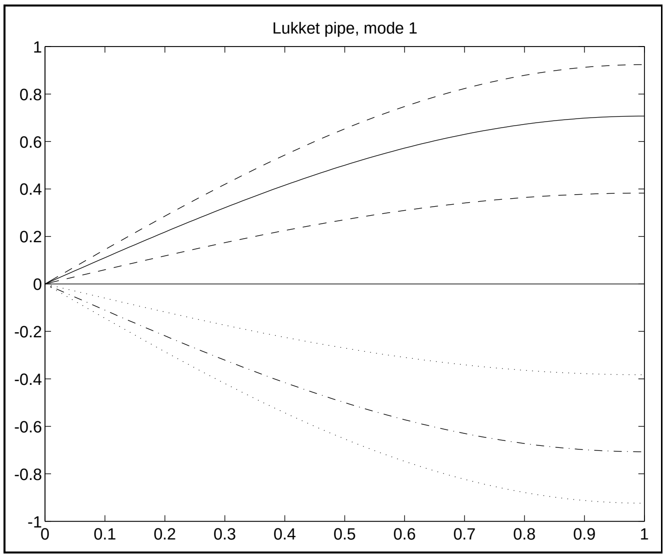
Figur 6.3: Sagtannkurve

Vi får samme type svingning når vi f.eks. drar et bord bortover gulvet. Den fryktede skrikingen av kritt mot tavle er et annet eksempel.

Som for gitaren, vil strengen ha visse frekvenser (grunntone og overtoner) som det er mulig for den å svinge på. Det er viktig at frekvensen til sagtannkurven faller sammen med en slik naturlig frekvens. Det ville vært en håpløs oppgave for fiolinisten å måtte tilpasse frekvensene manuelt. Heldigvis skjer det en tilbakekobling (feedback) fra strengen til buen, som gjør at de to frekvensene låses til hverandre helt av seg selv.

6.3 Piper med en lukket og en åpen ende

Hvis en trykkbølge farer fram og tilbake i et rør som er lukket i den ene enden og åpent i den andre, vil det settes opp en stående bølge. Vi får null utsving (knute) i luftttrykket i den åpne enden, og maksimalt utsving i den lukkede enden. Som vist i figur 6.4, der den åpne enden er til venstre, gir dette modus 1 (grunntonen).



Figur 6.4: Grunnsvingningen i en pipe med en lukket og en åpen ende

Hvis røret er L meter langt, ser vi at bølgelengden til grunntonen er $\lambda = 4L$ meter. Fra sammenhengen mellom frekvens, fart og bølgelengde ($f = \frac{v}{\lambda}$) får vi frekvensen til grunntonen, f.eks. for et 140 cm langt rør:

$$f_1 = \frac{v}{4L} = \frac{343m/s}{4 \cdot 1.4m} = 61Hz.$$

Neste mulige stående bølge er vist i figur 6.5.

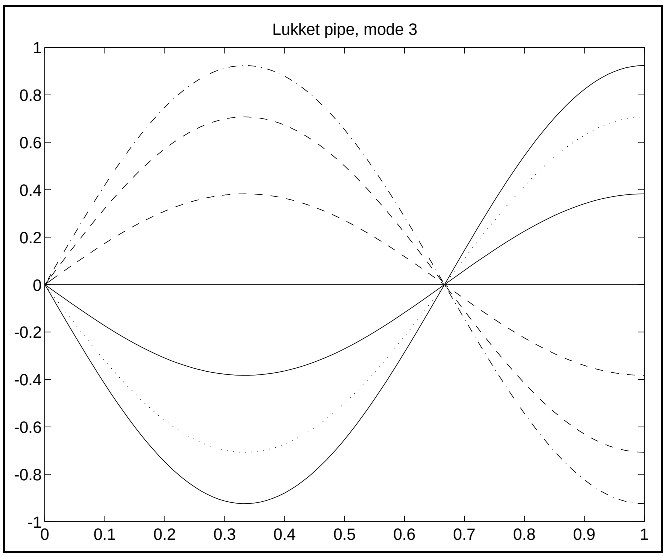
Bølgelengden er her $\lambda = \frac{4L}{3}$, som gir frekvensen

$$f_3 = \frac{v}{\frac{4L}{3}} = 3f_1 = 184Hz.$$

Videre vil vi kunne få svingninger med frekvens $f_5 = 5f_1$, $f_7 = 7f_1$ etc. Vi ser at vi får et spektrum med *bare odde overtoner.*

Parasaurolophus

Andeøglene (hadrosaurene) er en gruppe dinosaurer fra Kritt-tiden. Mange andeøgler har en underlig "hjelm" på hodet, som inneholder lange rør av bein. Mange rare forklaringer



Figur 6.5:

ble gitt på dette tidligere. Det var f.eks. en vanlig misforståelse at disse svære dyrene levde i vann, og at hode-apparatet ble brukt som luft-reservoar eller snorkel.

Det er nå klarlagt at det dreier seg om et akustisk apparat, en resonator som hjalp til i produksjonen av brøl og støt som må ha vært helt fantastiske. [3] Et eksempel er Parasaurolophus, som hadde en lang kam på hodet, hvor to parallelle piper var kveilet dobbelt. Pipene begynner nede i halsen (lukket ende), hvor det formodentlig satt en slags eksitator. Det finnes noen skrå beinflater som minner mistenkelig om dem vi finner i orgelpiper, så kanskje dyret bare blåste luft fra sine digre lunger. Pipene endte i hvert sitt nesebor (åpen ende), men det er ikke funnet noe som likner på sjalstykker.

Lengden på hver pipe er tilsammen ca. 2.7 meter, [4] og ut fra dette kan vi regne ut frekvensen til grunntonen:

$$f_1 = \frac{v}{4L} = \frac{343m/s}{4{\cdot}2.7m} = 32Hz.$$

Dette er virkelig en lav frekvens, og det betyr at lyden bar langt. Dagens elefanter kommuniserer også på svært lave frekvenser, tildels under menneskets hørbare område (subsonisk lyd eller infralyd).

6.4 Trompet

Et trompet er ikke et enkelt rør, og dette gir betydelige endringer sammenliknet med situasjonen over. Et komplisert samspill mellom munnstykke, rør og sjalstykke medfører at vi får resonanser for f_2, f_4, f_6 etc. Vanlig dyp tone (med alle ventiler åpne) for en

[3] Ingen har så vidt vites laget en modell av dette og prøvet det i praksis. Jeg har som prosjekt å gjøre dette en gang!

[4] Målt på et eksemplar som er utstilt på Paleontologisk Museum i Oslo. Gå og ta en titt selv!

messingblåser har grunntonen på f_4, og overtonene på f_8, f_{12} etc. En trompet har et ca. 140 cm langt rør, og dette skulle gi en f_1 på 61 Hz, og dermed grunntone f_4=244 Hz. Dette er omtrent hva man observerer.

Når man spiller en trompet, vil svingningene fra leppene (eksitator) synkroniseres med frekvensen til en av de mulige svingningsmodi, akkurat som vi så for fiolinen. Dette gjør at det er nærmest umulig å spille toner mellom de som er gitt fra rørets lengde. For å utvide antallet tilgjengelige noter, kan man endre rørlengden med ventiler eller glidende rør (trombone).

Sjalstykket ("tuten") fungerer som en pussig resonator, som former spekteret videre.

6.5 Menneskestemmen

Menneskestemmen kan også innpasses i eksitator-resonator-modellen. Eksitator er stemmebåndene, som sitter i strupehodet (glottis). Disse svinger med et meget rikt overtonespektrum. Grunnfrekvensen styres først og fremst av nerveimpulser, men lengden på røret over strupehodet spiller også inn. Hvis man synger en lys tone, vil strupehodet plassere seg høyt oppe i halsen. Når man går ned i tonehøyde, vil man tydelig kunne se at strupehodet faller langt nedover. Hastigheten lyden forplanter seg med i røret er også viktig, fordi $f = \frac{v}{\lambda}$. Hvis man fyller rommet over glottis med helium, vil lyden gå fortere og tonehøyden vil øke.

Når man hvisker, blåser man bare luft istedet for å lage en svingning med en grunnfrekvens. Like over glottis vil vi da ha *hvit støy*, dvs. at alle frekvenser er like sterkt til stede i et kontinuerlig spektrum.

Enten man snakker eller hvisker, vil luftrøret over glottis, samt munnhulen, fungere som et filter (resonator) som farger det rike spekteret fra glottis på ulike måter avhengig av hvordan vi former munnhulen. På den måten oppstår de ulike vokalene.

Kapittel 7

Lokalisering av lydkilder

7.1 Avstanden til lydkilden

Hvordan kan vi oppfatte avstanden til en lydkilde? Den viktigste retningslinjen er nok lydstyrken; amplityden faller proporsjonalt med avstanden. Men dette er ikke så enkelt som det umiddelbart kan virke, for lydstyrken blir den samme enten vi har en sterk lyd langt unna eller en svak lyd som er nær oss. Dette betyr at hjernen er nødt til å ha en ide om hvor sterk lyden var i utgangspunktet. Vi kjenner lydstyrken til de fleste lydkilder (biler, kirkeklokker), og kan derfor bedømme avstanden forholdsvis godt. Riktignok kan man rope både svakt og kraftig, men dette gir endringer i stemmekvalitet slik at vi kan ta dette med i beregningen og få en god avstandsbestemmelse.

Etterhvert som lyden forplanter seg gjennom lufta endrer den karakter. Luft virker som et lavpassfilter, og lyden mister mer diskant jo lenger den går. Dersom vi har en ide om det opprinnelige frekvensinnholdet, hjelper dette oss i avstandsbedømmelsen.

For lyder innendørs er romklangen en viktig ledetråd. Hvis lyden er langt unna, vil etterklangen (den reflekterte lyden) få større lydstyrke i forhold til den direkte lyden.

Alle disse tre effektene bør brukes hvis vi vil gi en realistisk simulering av avstand.

Som sagt er det viktig for avstandsbedømmelsen at vi vet hvor kraftig lyden var i utgangspunktet, og hvilket frekvensinnhold den hadde. Syntetiske lyder, og særlig sinustoner (som bl.a. ikke mister diskant med avstanden), er nesten umulige å avstandsbedømme. *Endring* i avstanden hører vi imidlertid lettere, som vi skal se i neste avsnitt.

Doppler-effekten

Dersom en lydkilde beveger seg mot oss, vil lydkildens egen hastighet komme i tillegg til hastigheten som lyden beveger seg mot oss med. Dermed vil flere bølgetopper enn normalt nå oss pr. sekund, og den høyere frekvensen medfører et tydelig hørbart pitch-skift oppover. Tilsvarende vil vi få et pitch-skift nedover når lydkilden beveger seg fra oss. Dette kalles *Doppler-effekten.* Frekvensene vil multipliseres med en faktor k som er gitt ved

$$k = \frac{v}{v + v_s}$$

der v er lydens hastighet (343 m/s) og v_s er hastigheten lydkilden beveger seg mot (negativ v_s) eller fra (positiv v_s) observatøren med. Når $v_s = 0$ blir $k = 1$, og vi får ikke noe pitch-skift.

Bemerk at når lydkilden beveger seg mot observatøren med lydens hastighet ($v_s = -v$), får vi null i nevner. Dette betyr at frekvensen går mot uendelig; vi har en *singularitet* som medfører at alle bølgetoppene faller sammen og summeres til et kraftig smell. Det er dette som heter å *bryte lydmuren.*

7.2 Retningen til lydkilden

Retningsbedømmelse er en meget komplisert affære, og mye av kunnskapen om dette har vi fått først i senere år. Det viser seg at det er mange forskjellige effekter som utnyttes sammen for å gi en nøyaktig bestemmelse av retningen til lydkilden.

Amplityde som retningsindikator

Hvert av våre to ører har størst følsomhet rett ut til siden. Dette betyr at lyd som kommer rett fra venstre registreres kraftigst i venstre øre, mens lyd som kommer rett fra høyre registreres kraftigst i høyre øre. Dette kan benyttes for å gi en grov retningsbestemmelse. Imidlertid kan vi ikke ut fra dette vite om en lyd kommer forfra eller bakfra, ovenfra eller nedenfra. Vi får bare en viss informasjon om retning venstre-høyre.

Panoreringsknappen som finnes på miksebord styrer retningen til lyden bare ved å regulere lydstyrken i venstre og høyre høyttaler. Dette må betraktes som primitivt.

Mange pattedyr av rovdyrenes orden [1] jakter på små byttedyr som er ute om natten og/eller befinner seg under jorden eller snøen. For slike bytteetere er korrekt retningsbestemmelse av stor betydning, og vi ser ofte at de har parabolformede og meget unidireksjonale ører. Katt og rev er gode eksempler. Hensikten med f.eks. ørkenrevens store ører er kanskje ikke først og fremst å samle mye lyd, men å gi stor retningsselektivitet. Ved å vri ørene i ulike retninger kan dyret finne et maksimum i responsen og derved bestemme retningen til byttet med stor nøyaktighet [2]. Vi er her inne på et spennende felt som vi kan kalle *bioakustikk*, dvs. læren om hvordan og hvorfor dyrene produserer og lytter til lyd.

Fase

Dersom en lydkilde befinner seg til venstre for oss, vil alle "toppene" i trykkvariasjonene komme tidligere til venstre enn til høyre øre. Hvis vi regner at ørene sitter 20 cm fra hverandre (?), vil vi kunne få en maksimal tidsforskjell på $\frac{0.2m}{343m/s} = 0.6ms$. Denne tidsforskjellen faller til null når lydkilden befinner seg rett foran/bak/over/under oss, og dette gir oss ytterligere informasjon om lydkildens retning venstre/høyre.

[1] Carnivora

[2] At ørkenreven har så store ører kan delvis også ha med temperaturkontroll å gjøre

Retningsbestemmelse basert på fase blir mer unøyaktig når vi går oppover i frekvens. Når vi kommer til ca. 2 kHz, vil en tidsforskjell på 0.6 ms tilsvare avstanden mellom bølgetoppene. Begge ører vil derfor oppfatte samme fase, selv om de egentlig registrerer hver sin topp.

Faseinformasjon er den viktigste retningslinjen for å bestemme retningen venstre-høyre for lave frekvenser. For høye frekvenser blir amplityde det viktigste.

Entydig retningsbestemmelse hos dyr

Amplityde og fase kan sammen gi en god retningsbestemmelse høyre-venstre. Det gjenstående problemet er hvordan man finner retningen opp-ned og bak-foran. Som vi har sett benytter katter seg av at de kan snu på ørene i nesten alle retninger og dermed lokalisere lydkilden i alle plan. [3]

Ugler må ha god retningsbestemmelse opp-ned. Ugler i våre strøk er spesielt avhengige av dette, fordi de gjerne sitter lavt over bakken og skal lokalisere mus under snøen. Hvis byttet feilaktig lokaliseres i en høy retning, vil ugla slå ned langt bortenfor musa. Enkelte ugler, som haukugla, bestemmer retningen ved å legge hodet 45 grader på skakke mot venstre og foreta en høyre-venstre-bestemmelse i dette skrå planet. Deretter legger de hodet 45 grader mot høyre og foretar en ny bestemmelse. Etter en rask kalkulasjon beregner de krysningslinjen mellom de to planene og finner en entydig retningsvektor i rommet. Mange ugler finner tydeligvis dette upraktisk, særlig i flukt. De fleste ugler i nordlige strøk har derfor utviklet asymmetriske ører, der bl.a. det ene øret er plassert mye høyere på hodet enn det andre. På den måten kan de gjøre to retningsbestemmelser ved å rotere hodet i stedet for å legge det på skrå.

Ugler har forøvrig mange andre underlige mekanismer for å bedre hørselen. [4]

Frekvensinnhold

Vi mennesker benytter en kanskje enda mer fantastisk, men neppe like nøyaktig, metode for å finne retningen opp-ned og bak-fram. Hele hodet, men særlig ytterøret, virker som et filter som farger lydspekteret ørlite, avhengig av hvilken retning lyden kommer fra. De intrikate brusk-vindingene (pinnae) i det ytre øret bidrar betydelig til dette. Hvis vi så kjenner lydens opprinnelige frekvensinnhold, kan vi anslå lydens retning entydig. Igjen betyr dette at syntetiske lyder, som hjernen ikke har noen modell for, kan være vanskelige å retningsbestemme opp-ned. Sinustoner kan ikke få endret sitt spektrum, fordi de består av bare en deltone. Sinusoider er derfor ekstra vanskelige å retningsbestemme opp-ned og bak-fram.

I de seneste år er det kommet digitale filtre som etterlikner frekvensresponsen til det

[3] Når en katt jakter, peker ørene framover mot byttet som den vil lokalisere. Når den er redd, peker ørene utover, slik at den får en mer omnidireksjonal respons som kan oppfange fiender som kan komme fra alle retninger.

[4] Se Norges Dyr, Fuglene, bind 5

ytre øret for ulike retninger. På den måten kan man panorere en lyd opp-ned og bak-fram i bare to kanaler.

Haas-effekten

Når vi hører en lyd f.eks. i en konsertsal, vil vi først motta den opprinnelige lyden fra den retningen den ble produsert. Deretter vil vi høre et antall refleksjoner fra alle mulige retninger. På mystisk vis klarer hjernen å plukke ut den "kopien" som ankommer først, og anvender utelukkende dette til retningsbestemmelsen, uten å forvirres av retningen til de senere refleksjonene.

Vi kan benytte dette til å simulere lydens retning, enten alene eller i kombonasjon med vanlig panorering. Hvis vi presenterer lyden noe tidligere for venstre øre (v.h.a. en liten digital-delay i høyre kanal) får vi en tydelig følelse av at lyden kommer fra venstre. Dette kalles *Haas-effekten*. Forsinkelsene som eksisterer i MIDI kan gi uønsket panorering pga. Haas-effekten.

Kapittel 8

Morsomme småting

8.1 Cerebral hemisfære-dominans

Hørselsnervene krysses, slik at lyd som kommer inn i venstre øre behandles i høyre hjernehalvdel og omvendt. Det viser seg at ikke- musikere gjenkjenner melodier lettere i venstre øre, det vil si at det meste av behandlingen skjer i høyre hjernehalvdel. Det er høyre hemisfære som tar seg av mønstergjenkjenning, kombinasjonsevne etc. Musikere derimot, gjenkjenner melodier lettere i høyre øre, og det betyr at musikere har lært seg til å oppfatte lyd mest med venstre hjernehalvdel. Venstre hemisfære er sentrum for språk og logikk.

8.2 Shepard-toner

Hvis man lager seg en grunntone med overtoner som langsomt faller i frekvens, får man en fallende glissando. Dersom man da langsomt fader ut de dypeste deltonene ettersom deres frekvenser nærmer seg null og langsomt fader inn nye deltoner med høye frekvenser, vil tonen late til å falle i det uendelige. Dette er en ganske pussig effekt.

8.3 Differanse-toner

Hvis man spiller to sinustoner kraftig, vil ulineæriteter i overføringsmediet og i øret føre til at vi kan høre en tredje tone med frekvens lik differansen av de to. 500 Hz og 400 Hz vil således gi en differanse-tone på 100 Hz. Dette må ikke blandes sammen med beats!

8.4 Sympatiske vibrasjoner

Hvis man setter f.eks. en streng i svingning, vil lydbølgene kunne indusere en svingning i en streng som er plassert et stykke unna. Dette kalles *sympatisk vibrasjon.* Effekten er spesielt sterk dersom den passive strengen har en naturlig svingefrekvens (resonans) som ligger nær frekvensen til den aktive svingeren.

En flott demonstrasjon av sympatisk vibrasjon får vi ved å rope ned i et flygel. Hver av delfrekvensene i ropet vil da indusere en svingning i den strengen som har nettopp denne resonansfrekvensen. Siden tangentene på flygelet dekker det meste av frekvensområdet, vil de induserte svingningene i store trekk gjenoppbygge spekteret i ropet, og vi får en vakker etterklang hvor stemmens klang kan gjenkjennes. Dette er faktisk en akustisk versjon av vocoding.

Hardingfela er et annet eksempel, der bordunstrengene (understrengene) svinger i sympatisk vibrasjon med hovedstrengene.