Den 16 september 2001

Spektrogram

Christer O. Kiselman

Innehåll:

1. Inledning

Under mina föreläsningar hösten 2001 på kursen *Fouriermetoder* för teknologer på utbildningsprogrammen *Miljö- och vattenteknik* och *Kemiteknik* har jag talat litet om spektrogram och deras användning för att studera mänskligt tal och fågelsång. Här beskriver jag idén bakom spektrogram. Boken som vi använder, Sollervall & Styf [1999], behandlar spektrum av signaler, men spektrogram kan ge en mer detaljerad information: spektrum visar vilka frekvenser som förekommer, medan spektrogram avslöjar både vilka frekvenser som förekommer och när de förekommer – trots att detta i princip är omöjligt.

Ett varmt tack till Sven Öhman för tips om Ladefogeds böcker om akustisk fonetik!

2. Spektrum

2.1. Allmänt om spektra

En funktion $f\colon \mathbf{R} \to \mathbf{C}$ har Fouriertransformen $\hat{f}$, som definieras av

$$\hat{f}(\omega) = \int_{\mathbf{R}} f(t)e^{-i\omega t}dt, \qquad \omega \in \mathbf{R}.$$

Här tänker vi oss t som tiden (mätt i till exempel sekunder) och ω som vinkelfrekvensen eller vågtalet (mätt i radianer per sekund). Frekvensen hos svängningen $t \mapsto e^{-i\omega t}$ är $\omega/2\pi$ perioder per sekund, dvs. $\omega/2\pi$ Hz; Hz är en förkortning av hertz.[1]

[1] Heinrich Rudolf Hertz, 1857–1894.

Eftersom $\hat{f}$ oftast är komplexvärd, så behöver man ett tre-dimensionellt diagram: man kan avsätta tiden åt höger, $\operatorname{Re}\hat{f}$ rakt fram och $\operatorname{Im}\hat{f}$ uppåt. Det är förstås ekvivalent att rita två två-dimensionella diagram, ett med alla par $(\omega, \operatorname{Re}\hat{f}(\omega))$ och ett med alla par $(\omega, \operatorname{Im}\hat{f}(\omega))$; de uppstår ju från det förra genom projektion (Sollervall & Styf [1999:84–88]). Vi kan också göra ett diagram med $|\hat{f}|$ (absolutbeloppet eller amplituden) och ett med $\arg\hat{f}$ (argumentet eller fasvinkeln). De kallas för ***amplitudspektrum*** respektive ***fasvinkelspektrum*** (Sollervall & Styf [1999:89–94]).

Eftersom reellvärda funktioner uppfyller $\hat{f}(-\omega) = \overline{\hat{f}(\omega)}$, och därför $|\hat{f}(-\omega)| = |\hat{f}(\omega)|$, så räcker det att rita upp $|\hat{f}(\omega)|$ för icke-negativa frekvenser ω för sådana funktioner.

När det gäller ljud så är $f(t)$ lufttrycket vid trumhinnan eller vid en mikrofon vid tidpunkten t (rättare sagt: avvikelsen från ett medellufttryck). Vi kan inte mäta lufttrycket vid alla tidpunkter, utan gör det under ett visst intervall $a \leqslant t \leqslant b$. Det innebär att vi anser att funktionen är noll utanför detta intervall och Fouriertransformen blir då

$$\hat{f}(\omega) = \int_a^b f(t)e^{-i\omega t}dt, \qquad \omega \in \mathbf{R}.$$

Om då $\hat{f}(\omega) \neq 0$ så säger vi att ***frekvensen*** ω ***förekommer*** i signalen (ljudet) f. I praktiken menar man med detta uttryck att $|\hat{f}(\omega)|$ är ganska stort: man negligerar små värden, som uppkommer genom störningar eller genom att man huggit av signalen f.

2.2. Spektrum för en ren ton

Vi vet från Sollervall & Styf hur spektrum för en avskuren cosinussvängning ser ut. Vi studerar svängningen över ett intervall $[c-a, c+a]$ med centrum i punkten c och längd $2a$. Om alltså nu

$$f(t) = \begin{cases} K\cos\alpha t, & c-a < t < c+a, \\ 0 & \text{annars,} \end{cases}$$

så är (se Sollervall & Styf sidan 87)

$$\hat{f}(\omega) = Ke^{-ic(\omega-\alpha)}\frac{\sin a(\omega-\alpha)}{\omega-\alpha} + Ke^{-ic(\omega+\alpha)}\frac{\sin a(\omega+\alpha)}{\omega+\alpha}, \qquad \omega \in \mathbf{R},$$

definierad genom kontinuitet för $\omega = \pm\alpha$. Den har två toppar, vid frekvenserna α och $-\alpha$. Eftersom vi bara behöver titta på positiva frekvenser för att studera amplitudspektrum, så ser vi en enda topp, vid $\omega = \alpha$ om vi väljer $\alpha > 0$, svarande mot den första termen, nämligen

$$\hat{f}(\omega) \approx \varphi(\omega) = Ke^{-ic(\omega-\alpha)}\frac{\sin a(\omega-\alpha)}{\omega-\alpha}, \qquad \omega \geqslant 0.$$

Felet som man gör när man approximerar $\hat{f}$ med φ är inte större än K/α då $\omega \geqslant 0$; vid toppen $\omega = \alpha$ är felet inte större än $K/2\alpha$. Eftersom $|\varphi(\alpha)| = Ka$, så är det relativa felet (alltså felet jämfört med toppen hos $|\varphi|$, som är Ka) högst $1/a\alpha$

för $\omega \geqslant 0$, vid toppen $\omega = \alpha$ bara $1/2a\alpha$. Vi noterar att antalet svängningar som f gör under mätetiden är $a\alpha/\pi$. Det vi behöver är alltså att antalet perioder under mätetiden är tillräckligt stort: $a\alpha \gg \pi$. Om så icke är fallet, kommer de två topparna hos $\hat{f}$ alltför nära varandra; det är inte berättigat att betrakta dem var för sig som vi gör när vi studerar φ. Minnesregeln är alltså att toppen hos $|\varphi|$ är signalens intensitet K gånger signalens halva längd a; toppen hos $\hat{f}$ blir ungefär lika hög förutsatt att störningarna från den negativa frekvensen $-\alpha$ och från andra signaler är små.

Om vi markerar punkter där värdet hos $|\varphi|$ är större än $2Ka/3\pi \approx 0,2122Ka$, och säger att dessa utgör toppen, så får denna en bredd på mindre än $2\pi/a$. De är dessa frekvenser som förekommer, alltså de som ligger på avståndet π/a från α. För stora a blir alltså toppen smal; det innebär att vi måste mäta signalen under tillräckligt lång tid. Det avgörande är om denna tid, som är $2a$, är tillräckligt stor i förhållande till α. Detta innebär som vi sett att signalen skall hinna svänga tillräckligt många gånger under det intervall man mäter. Och intuitivt är det ju klart att en cosinussvängning som bara hinner svänga, säg, en halv period under mätetiden inte kan bestämmas särskilt väl.

För att kvantifiera detta kan vi tänka oss en signal med en frekvens av 440 Hz. Denna ton kallas ettstrukna a, och vi tänker oss att vi mäter den under en sekund. Då är $a = 0,5\,\mathrm{s}$ och vinkelfrekvensen $\alpha = 2\pi \cdot 440\,\mathrm{s}^{-1}$, dvs. 880π radianer per sekund. Antalet svängingar under mätetiden är $a\alpha/\pi = 440$. Toppens bredd, som vi angivit till $2\pi/a$, blir alltså 4π radianer per sekund eller 2 Hz. Toppen kommer alltså att ligga mellan 439 och 441 Hz, en god precision. Om vi mäter under kortare tid blir precisionen mindre.

2.3. Spektrum för en vokal

Vi har sett i kursen att en vokal har ett spektrum som har två eller tre toppar. Exempelvis har enligt Ladefoged [1993: 193] vokalen [i] tre toppar, vid 280, 2250 och 2850 Hz. Man säger att ljudet är uppbyggt av tre ***formanter***. Det är förmodligen författarens egen röst som han analyserat; en kvinnas röst ligger högre, men relationerna är desamma. Det intressanta är nämligen förhållandena mellan formanternas frekvenser. För vokalen [i] ligger den andra formanten mycket högt över den första: kvoten är $2250/280 \approx 8$, vilket är tre oktaver. Som kontrast kan vi ta vokalen [u] (som i det svenska ordet *oro*). Den har tre formanter på 310, 870 och 2250 Hz. Här är förhållandet mellan de två första blott $870/310 \approx 2,8$, vilket är en och en halv oktav ($2^{1,5} = 2\sqrt{2} \approx 2,828$). För [i] är alltså avståndet mellan de två lägsta formanterna dubbelt så stort som för [u] mätt på en logaritmisk skala.

En annan skillnad mellan [i] och [u] är styrkeförhållandena mellan formanterna. Enligt Ladefoged [1973:96–97] är amplituden hos den andra formanten hos [i] ungefär $0,67$ av den förstas, den tredjes är $0,96$ av den förstas. Hos [u] är den andra formanten svagare, ungefär $0,3$ av den förstas.

Det sagda innebär att det verkar som om en rimlig syntes av vokalen [i] ges av

$$f(t) = \cos(\alpha_1(t - t_1)) + 0,67\cos(\alpha_2(t - t_2)) + 0,96\cos(\alpha_3(t - t_3)),$$

där $\alpha_1 = 2\pi \cdot 280$ och $\alpha_2 = 2\pi \cdot 2250$ och $\alpha_3 = 2\pi \cdot 2850$. Tidpunkterna t_j anger när termen har sitt största värde; de är ointressanta, ty dem kan man (troligen) inte höra. Vid filtrering av olika slag brukar man endast bearbeta amplitudspektrum, eftersom man inte vet hur förändringar i fasvinkelspektrum påverkar hörselintrycket.

Vidare verkar en rimlig syntes av vokalen [u] vara

$$f(t) = \cos(\alpha_1(t - t_1)) + 0,3\cos(\alpha_2(t - t_2)),$$

där $\alpha_1 = 2\pi \cdot 310$ och $\alpha_2 = 2\pi \cdot 870$. Om man ritar upp dessa kurvor, så ser man att kurvorna för [i] och [u] har mycket olika karaktär: den för [i] visar tydligt sina höga frekvenser; den för [u] är nära en ren svänging eftersom den andra formanten är så pass svag.

En intressant iakttagelse är att i serien [i e a o u] av kardinalvokaler avtar den andra frekvensen monotont. Den första frekvensen växer först och avtar sedan. Om vi därför låter vokalerna gå genom ett högpassfilter, så att den första (lägsta) formanten filtreras bort, så får vi en monotont avtagande följd av frekvenser. Och vi kan faktiskt utan elektronisk utrustning utföra en högpassfiltrering: när man viskar stänger man av grundfrekvensen. Om man viskar vokalerna [i e a o u], så hör man tydligt att tonhöjden avtar.

2.4. Spektrum för en koncentrerad signal är brett

En signal som

$$f(t) = \frac{1}{\sqrt{2\pi}s} e^{-(t-c)^2/2s^2}, \qquad t \in \mathbf{R},$$

är mer eller mindre koncentrerad till tiden c (då den antar sitt maximum). Ett mått på hur utspridd signalen är utgörs av tiden s, som vi kallar för ***spridningen***. En uppfattning om dess betydelse får vi av observationen att funktionens värde vid tiden $t = c \pm s$ är lika med $1/\sqrt{e} \approx 0,6065$ gånger maximivärdet. Då $t = c \pm 2s$ är värdet blott $1/e^2 \approx 0,1353$ av maximimivärdet. Det mesta av signalstyrkan ligger alltså i intervallet $[c - 2s, c + 2s]$.

Vi har sett på föreläsningarna att

$$\hat{f}(\omega) = e^{-ic\omega} e^{-s^2\omega^2/2};$$

det är alltså en funktion av samma typ som f, koncentrerad kring origo, och med spridningen lika med $1/s$ radianer per sekund. Ju mer koncentrerad signalen är, desto mer utbrett är dess spektrum – och omvänt. Detta är ett konkret exempel på Heisenbergs[2] osäkerhetsrealation.

Ett annat exempel får vi om vi sätter

$$f(t) = \begin{cases} \dfrac{1}{2s}, & -s < t < s, \\ 0 & \text{annars.} \end{cases}$$

Detta är ju funktionen som vi studerat i avsnitt 2.2 med $\alpha = 0$, $c = 0$, $a = s$ och $K = 1/2s$, så dess Fouriertransform blir

$$\hat{f}(\omega) = \frac{\sin s\omega}{s\omega}, \qquad \omega \in \mathbf{R}.$$

[2]Werner Heisenberg, 1901–1976.

Vi ser att frekvenser i intervallet $[-2\pi/s, 2\pi/s]$ förekommer – om vi drar en lämplig gräns för hur små amplituder vi kan negligera. Och intervallet blir stort när s är litet. Om vi till exempel tar $s = 1\,\mathrm{ms}$ så blir det ett kraftigt bidrag från vinkelfrekvenser upp till $2\pi/s$, dvs. frekvenser upp till 1 kHz.

Ett tredje exempel är funktionen

$$f(t) = \frac{1}{2s} e^{-|t|/s}, \qquad t \in \mathbf{R}.$$

Här är $\hat{f}(\omega) = 1/(1 + s^2\omega^2)$. Vi ser att $\hat{f}(0) = 1$, $\hat{f}(1/s) = 0,5$ och $\hat{f}(2/s) = 0,2$. Om till exempel $s = 0,5\,\mathrm{ms}$, så förekommer alla vinkelfrekvenser upp till $2/s = 4\,000$ radianer per sekund, motsvarande frekvenser upp till 637 Hz, med en styrka motsvarande åtminstone en femtedel av amplituden i origo.

De tre nämnda exemplen kan lätt generaliseras. Vi fixerar en funktion g som har integral 1 och som är noll utanför ett begränsat intervall. Vi fixerar vidare ett positivt tal s och sätter

$$f(t) = \frac{1}{s} g(t/s), \qquad t \in \mathbf{R}.$$

Då får även f integralen 1. Och dess Fouriertransform blir

$$\hat{f}(\omega) = \hat{g}(s\omega), \qquad \omega \in \mathbf{R}.$$

Vi ser att $\hat{f}(0) = 1$ och att funktionen blir mycket utbredd om s är litet. Funktionen $\hat{g}$ är kontinuerlig i origo, så $\hat{f}$ antar värden nära 1 i ett stort intervall. Fenomenet beror alltså inte på den speciella form som funktionerna i de första exemplen hade.

Gränsfallet i alla dessa exempel är ett Diracdelta,[3] som har oändligt kort varaktighet (spridning $s = 0$) och Fouriertransformen 1, alltså $\hat{\delta} = 1$.

2.5. Spektrum för brus

Ett brus kännetecknas av att det innehåller många frekvenser. En rimlig modell är alltså en signal

$$f(t) = \sum_{j=1}^{N} A_j \cos(\alpha_j (t - t_j)), \qquad t \in [a, b],$$

där N är ett mycket stort tal, och där vinkelfrekvenserna α_j ligger mer eller mindre jämt fördelade i ett helt intervall $[\omega_0, \omega_1]$. Det visar sig då att amplitudspektrum fyller ut hela detta intervall $[\omega_0, \omega_1]$.

Ett exempel är *s*-ljudet, [s], och *sje-ljudet*, [ʃ]. Det första ljudet använder alla frekvenser mellan 4 kHz och 10 kHz; det andra ligger lägre och använder alla frekvenser mellan 3 kHz och 6 kHz enligt Ladefoged [1973:53].

2.6. Spektrum för temperaturen och tidvattnet

Temperaturen varierar under dagen och under året. En modell för dess variation kan vara

$$T(t) = \frac{1}{2} A_0 + \sum_{j=1}^{N} A_j \cos(\alpha_j (t - t_j)), \qquad t \in \mathbf{R}.$$

[3] P. A. M. Dirac, 1902–1984.

Om vi mäter tiden i dagar, så blir α_1, som beskriver den dagliga variationen, lika med 2π, medan α_2, som beskriver den årliga variationen, blir $2\pi/365$. Medeltemperaturen $A_0/2$ kan vi sätta till 5°C, medan den dagliga amplituden kanske är $A_1 = 6°\text{C}$, den årliga amplituden kanske $A_2 = 12°\text{C}$, allt tänkt för Uppsala. Vidare anger tidpunkterna t_1 och t_2 när det är varmast. Under dagen är det kanske varmast klockan 13:30, vilket ger $t_1 = 13,5/24$; under året är det kanske varmast den 15 juli, vilket ger $t_2 = 365 \times 6,5/12$. Om man förfogar över temperaturmätningar var tredje timme under några års tid, så kan man beräkna de värden på A_0, A_1, A_2 och t_1, t_2 som ger den bästa anpassningen. Frekvenserna är alltså givna, α_1 motsvarar $1/86400\,\text{Hz} \approx 11,57\,\mu\text{Hz}$, medan α_2 motsvarar $31,7\,\text{nHz}$.

Så kan vi fundera på om det finns en komponent i temperaturen med samma period som solfläckscykeln, som är 11 år, motsvarande en frekvens på $2,88\,\text{nHz}$. Vad blir i så fall amplituden A_3? Om man har temperaturdata under ett eller två sekel – och det har vi ju – så kan man se om det finns en komponent med perioden 11 år.

Och var det inte varmare under bronsåldern? I så fall skulle man kanske leta efter en komponent i T med en period på, säg, 4 000 år, motsvarande en frekvens på $7,925\,\text{pHz}$. Och med vilken period kommer istiderna? Om man har data under tillräckligt lång tid, så går det att Fourieranalysera och se vilka frekvenser som förekommer...

Spektrum för T i Uppsala visar alltså en kraftig formant vid frekvensen motsvarande perioden 1 år, och en något svagare formant vid frekvensen motsvarande perioden 1 dygn. I tropikerna är styrkeförhållandet omvänt. Huruvida det finns en formant vid frekvensen motsvarande solfläckscykelns 11 år vet jag inte. Här fordras forskning.

Tidvattnet vid en viss ort kan beskrivas av samma serie som för temperaturen. Här gjordes Fourieranalysen, eller som den också kallas, den harmoniska analysen, av Lord Kelvin[4] på 1870-talet. Den största amplituden hör till perioden ett halvt måndygn, cirka 12,5 timmar; den näst största till perioden ett halvt soldygn, alltså 12 timmar. Totalt behövs för goda prognoser tjugo termer, varav tolv har astronomiskt ursprung och de återstående åtta geografiskt ursprung. Kelvin byggde en analog maskin som användes för att förutsäga tidvattnet ända fram till 1960-talet.

3. Spektrogram

3.1. Konstruktion av spektrogram

Spektrum visar alltså vilka frekvenser som förekommer, men de säger inget om när under mätetiden de förekommer. Kanske har vi en ton på 440 Hz under en halv sekund, och sedan en ton på 880 Hz under en halv sekund. På amplitudspektrum kan man då inte skilja en sådan signal från en där den högre tonen kommer under den första halvan och den lägre därefter.

Ofta vill man kunna säga att en viss frekvens förekommer vid en viss tidpunkt. Den totala tiden från a till b är kanske lång, så lång att en analys av hela signalen blir ointressant. Vi delar därför upp signalen i lämpliga bitar och Fouriertransformerar varje bit. Det kan till exempel gå till så här.

[4] Lord Kelvin (tidigare namn William Thomson), 1824–1907.

Man delar upp signalen i ett antal kortare signaler, till exempel så att man sätter, för varje $j = 1, ..., n$,

$$f_j(t) = \begin{cases} f(t), & t_{j-1} \leqslant t < t_j, \\ 0 & \text{annars.} \end{cases}$$

Här har vi infört en massa delningspunkter $a = t_0 < t_1 < t_2 < \cdots < t_n = b$. Så beräknar man Fouriertransformen av varje f_j:

$$\hat{f}_j(\omega) = \int_{t_{j-1}}^{t_j} f(t)e^{-i\omega t}dt, \qquad \omega \in \mathbf{R}, \quad j = 1, ..., n.$$

På så sätt får vi n funktioner $\hat{f}_j$, $j = 1, ..., n$, vilkas absolutbelopp kan åskådliggöras genom en funktion av två variabler j och ω:

$$F(j, \omega) = |\hat{f}_j(\omega)|, \qquad \omega \geqslant 0, \quad j = 1, ..., n.$$

I praktiken är index j då en tid. Vi kan nämligen bestämma att alla differenser $t_j - t_{j-1}$ skall vara lika och definiera $F_1(t, \omega) = F(j, \omega)$ då $t_{j-1} \leqslant t < t_j$, och så rita upp funktionen F_1 för vissa värden av t och ω, säg för $a \leqslant t \leqslant b$ och $\omega_0 \leqslant \omega \leqslant \omega_1$. Rättare sagt, vi kan rita en tvådimensionell projektion av den tredimensionella grafen av funktionen F_1, dvs. alla tripplar $(t, \omega, F_1(t, \omega))$. Men ofta ritar man inte hela denna graf, utan nöjer sig med att svärta de punkter (t, ω) i planet där $F_1(t, \omega)$ är större än ett visst värde. De punkter som blir svarta är då de punkter där frekvensen ω förekommer hos f vid tidpunkten t. Det sista uttrycket är egentligen felaktigt. Man kan strängt taget aldrig veta vilken frekvens som förekommer vid en viss tidpunkt. Det handlar om de frekvenser som förekommer i ett helt intervall $[t_{j-1}, t_j]$, där j väljes så att en given tidpunkt t ligger i intervallet.

Med ett ***spektrogram*** avser jag här en grafisk framställning av spektra av alla funktionerna f_j, $j = 1, ..., n$, ordnade i tidsföljd. Medan ett spektrum har blott en oberoende variabel (mätt i radianer per sekund eller perider per sekund, hertz), så har alltså spektrogrammet två oberoende variabler, nämligen såväl tiden (mätt i sekunder) som frekvensen (mätt i hertz).

Under kursens gång har jag visat många sådana spektrogram. Därvid har jag alltid avsatt tiden längs den horisontella axeln och frekvensen längs den vertikala. När jag i fortsättningen talar om horisontella streck och vertikala streck så syftar jag just på detta sätt att rita spektrogrammen. Detta stämmer ju också med det vanliga bildspråket för toner: vi talar om *höga* och *låga* toner och om *tonhöjd*, som är just frekvensen mätt i perioder per sekund, alltså Hz.

Spektrogrammet visar alltså inte bara *vilka* frekvenser som förekommer utan också *när* de förekommer. Trots att Heisenbergs osäkerhetsrelation egentligen förbjuder oss att tala om detta, så har förfarandet en praktisk betydelse, ty svängningarna hos f är så snabba att varje intervall $t_{j-1} \leqslant t < t_j$ innehåller många av dem, och då blir det meningsfullt att säga att en av dessa frekvenser förekommer hos f vid tiden t_j. Om vi däremot vill studera en frekvens som är så låg att bara några få perioder ryms i varje intervall $t_{j-1} \leqslant t < t_j$, så blir det inte längre meningsfullt.

Låt oss kvantifiera. Säg att vi använder intervall med längden $t_j - t_{j-1} = 100\,\mathrm{ms}$, och centrerade kring punkter $\frac{1}{2}(t_{j-1} - t_j)$. Det betyder att $a = 50\,\mathrm{ms}$. Om frekvensen är $1\,\mathrm{kHz}$, så är alltså $\alpha = 2\,000\pi$ radianer per sekund. Under mätetiden förekommer $a\alpha/\pi$ svängingar, dvs. 100 svängningar. Toppens bredd blir $2\pi/a = 40\pi$ radianer per sekund, motsvarande 20 Hz. Det betyder att de frekvenser som syns i spektrogrammet ligger i intervallet från 990 Hz upp till 1010 Hz. Precisionen blir sämre än i det förra exemplet, i slutet av avsnitt 2.2, där vi mätte under en hel sekund.

För att registrera tal använder man faktiskt så korta intervall som 7 ms. Då blir toppens bredd fjorton gånger större, hela 280 Hz, vilket gör att de markerade frekvenserna ligger i intervallet från 870 Hz upp till 1140 Hz. En ren ton blir alltså ganska utsuddad, men dålig upplösning i frekvensrummet är det pris man måste betala om man önskar få hög upplösning i tiden.

Kan vi få bättre precision genom till exempel mer förfinad elektronisk utrustning? Nej, det är omöjligt på grund av Heisenbergs osäkerhetsrelation; se också avsnitt 2.3 om relationen mellan spridningen hos f och spridningen hos dess Fouriertransform.

3.2. Spektrogrammet för en ren ton

Vi har sett hur spektrum för en ren ton ser ut i avsnitt 2.2. Olika värden på talet c, som anger centrum för intervallet, syns inte i amplitudspektrum (faktorn $e^{-ic(\omega-\alpha)}$ har ju absolutbeloppet 1). Därför blir amplitudspektra för alla f_j lika. Om vi tar flera intervall, så ser vi att spektrogrammet blir en horisontell linje av viss tjocklek, men ganska smal, och av samma längd som tonen.

Om tonen till exempel sjunker (frekvensen blir lägre) med tiden, så visar spektrogrammet en kurva som går nedåt.

Övertoner visar sig som linjer som ligger ovanför den nämnda linjen. Om vi har en överton som är en oktav högre än grundtonen, så ligger den på den dubbla frekvensen. Därför använder man ofta en logaritmisk vertikalskala, så att avståndet mellan 440 Hz och 880 Hz är lika stort som det mellan 880 Hz och 1760 Hz.

3.3. Spektrogrammet för en vokal

Vi har sett hur en vokals spektrum ser ut i avsnitt 2.3. I spektrogrammet blir bilden ett antal någorlunda horisontella linjestycken eller band under den tid man uttalar vokalen. Banden kanske inte är helt horisontella utan lutar litet. Det är i så fall ett tecken på att tonhöjden för någon av formanterna ändras under uttalets gång.

3.4. Spektrogrammet för en smäll

Vi har sett att spektrum för en koncentrerad signal blir brett: alla frekvenser i ett stort intervall förekommer. Hur blir då spektrogrammet? Först konstaterar vi att smällen bara förkommer under ett av delintervallen $t_{j-1} \leqslant t < t_j$, så utanför detta finns ingen svärta alls. Och vid tidpunkten $a = \frac{1}{2}(t_{j-1} + t_j)$ registreras en svärta över ett stort frekvensintervall: alla frekvenser som kan mätas upp förekommer. Spektrogrammet blir ett vertikalt linjestycke. Detta gäller för en smäll, en knäpp och vilket som helst mycket kortvarigt ljud. Vi såg ju nämligen att förhållandena blir likartade för varje koncentrerad funktion.

3.5. Spektrogrammet för brus

Vi har sett hur ett brus' amplitudspektrum ser ut i avsnitt 2.5: det fyller ett stort intervall någorlunda jämt. Och när vi delar upp signalen i småbitar sker ingen förändring. I spektrogrammet bildar därför en brussignal, eller ett väsande ljud, en fylld rektangel, där basen är det tidsintervall under vilket väsandet pågår, och höjden är intervallet av alla förekommande frekvenser.

3.6. Spektrogram för fåglarnas sånger

Lars Gårdings bok [1987] innehåller spektrogram av 245 fågelarters sånger och läten. Det visar sig att fåglarnas läten är mycket mer varierade än människans, nämligen om man tittar över alla arterna... människan kanske har rekordet för en enda art. Det förekommer smällar, till exempel hos tjädertuppens[5] parningsläte: *täckupp täckupp täck täck täck pang*; mycket riktigt syns smällarna som vertikala streck i spektrogrammet, där det sista, återgivande den smäll som har liknats vid utdragandet av en champagnekork, är längst.

Sångsvanen[6] kan sjunga mycket bättre än knölsvanen.[7] Spektrogrammet avbildar den förstnämndas sång som ett antal horisontella linjer på litet olika höjd (jämför avsnitt 3.2), medan den senares väsande fyller en rektangel i spektrogrammet, helt i enlighet med överläggningen i avsnitt 3.5.

Referenser

Lars Gårding

1987 *Fåglars sång och läten.* Bokförlaget Signum. 85 sidor. ISBN 91-85330-75-2.

Peter Ladefoged

1973 *Elements of Acoustic Phonetics.* Chicago & London: The University of Chicago Press. vii + 120 sidor. ISBN 0-226-46785-6.

1993 *A Course in Phonetics.* Forth Worth: Harcourt Brace College Publishers. xi + 308 sidor. ISBN 0-15-500173-6.

Håkan Sollervall & Bo Styf

1999 *Transformteori för ingenjörer.* Skebobruk: Bokförlaget Kub. 139 sidor. ISBN 91-89104-03-X.

Författarens adress: Uppsala universitet, Matematiska institutionen,
Box 480, SE-751 06 Uppsala.

Telefon: 018-4713216 (a); 018-300708 (h) Fax: 018-4713201

Datoradress: kiselman@math.uu.se URL: http://www.math.uu.se/~kiselman

[5] Tjäder, *Tetrao urogallus* L., urogalo, capercaillie, grand coq de bruyère.

[6] Sångsvan, *Cygnus cygnus* (L.), kantocigno, whooper swan, cygne chanteur à bec noir.

[7] Knölsvan, *Cygnus olor* (Gmelin), muta cigno, mute swan, cygne muet à bec rouge.