

Multimedia

Gregor Rozinaj, Renata Rybárová, Ivan Minárik

Författarna: Gregor Rozinaj, Renata Rybárová, Ivan Minárik
Titel: Multimedia
Översatt av: Jaroslav Chlumsky
Publicerad av: České vysoké učení technické v Praze
Sammanställt av: Fakulta elektrotechnická
Kontakt adress: Technická 2, Praha 6, Czech Republic
Phone Number: +420 2 2435 2084
Print: (only electronic form)
Antal sidor: 123
Utgåva, utgåvan: 1:a utgåva

ISBN 978-80-01-05313-3

Granskad av: Dalibor Svoboda

Innovative Methodology for Promising VET Areas
<http://improvet.cvut.cz>



**Programmet
för livslångt
lärande**

Projektet genomförs med ekonomiskt stöd från Europeiska kommissionen.

För uppgifterna i denna publikation (som är ett meddelande) ansvarar endast upphovsmannen. Europeiska kommissionen tar inget ansvar för hur dessa uppgifter kan komma att användas.

FÖRKLARINGAR



Definition



Intressant



Not, notering



Exempel



Sammanfattning



Fördel



Nackdel

ANTECKNING

Denna kurs innehåller grundläggande information om Multimedia. I den första delen finns beskrivna grundläggande kunskaper om analoga och digitala signaler, digitalisering av signaler samt timm- och frekvensmässig representation av signaler. Nästa del innehåller information om bearbetning av signaler, analoga och digitala teknologier, filterning och beskrivning av telekommunikationskanaler. Slutligen är kursen tillägnad audio och video signalkomprimerings- tekniken och aplikationsteorin om signalernas bearbetning till Multimedia: språksyntesen och igenkännande av språk och bild.

MÅL

Huvudmålet med denna kurs är att bekanta studenter med grunderna för signal bearbetning, i synnerhet med den teknik som används vid multimedial bearbetning. För studenter presenteras grundläggande principer såsom Fouriers transformation, digitala filter, lineära systems, komprimerings teknik och hur den moderna vetenskapskunskapen kan användas praktiskt.

LITTERATUR

- [1] DÚHA, J., GALAJDA, P., KOTULIAK, I., LEVICKÝ, D., MARCHEVSKÝ, S., MIKÓCZY, E., PODHRADSKÝ, P. a kol. *Multimedia ICT technologies, network platforms and multimedia services*, Vydavateľstvo STU Bratislava, 2005, ISBN 80-227-2310-X.
- [2] GAMEC, J. *Spracovanie multimedialnych signálov*, publikované v rámci projektu ESF NGN, kód projektu v ITMS: 13120110126, Bratislava, 2007.
- [3] TALAFOVÁ, R. – ROZINAJ, G. – CEPKO, J. – VRABEC, J. *Multimedia SMS Reading in Mobile Phone* In: INTERNATIONAL JOURNAL of MATHEMATICS AND COMPUTERS IN SIMULATION, Issue 1, Volume 1, ISSN 1998-0159, 2007.
- [4] KOTULIAKOVÁ, J. – ROZINAJ, G. *Číslíkové spracovanie signálov I*, FABER Bratislava, 1996.
- [5] KOTULIAKOVÁ, J. – ROZINAJ, G. – POLEC, J. – PODHRADSKÝ, P. a kolektív. *Číslíkové spracovanie signálov II*, FABER Bratislava, 1997
- [6] HARDESTY, L. *Explained: The Discrete Fourier Transform*, 2009, <http://web.mit.edu/newsoffice/2009/explained-fourier.html> [online]
- [7] MINÁRIK, I. Coding of audio signals at low speed, Diploma Thesis, 2011

- [8] MARMOL, F.G., et al. *ANALYSIS: State of The Art on Identity, Security and Trust*, Deliverable D3.1, HBB-Next FP7-ICT-2011-7, 2012
- [9] DEVENTER, O. et al. *ANALYSIS: Multi-User, Multimodal & Context Aware Value Added Services*, Deliverable D5.1, HBB-Next FP7-ICT-2011-7, 2012
- [10] LEVICKÝ, D., RIDZOŇ, R. *Multimédia a multimediálne technológie*, publikované v rámci projektu ESF NGN, kód projektu v ITMS: 13120110126, Vydalo Vydavateľstvo STU v Bratislave v spolupráci s AGROGENOFOND Nitra, 2007, ISBN 978-80-227-2604-7
- [11] PODHRADSKÝ, P. *Fourierov rad a Fourier transformácia*, publikované ako študijný materiál pre predmet inžinierskeho štúdia "Analogové a digitálne signály a systavy I", študijný program Telekomunikácie, FEI STU Bratislava, 2003

Innehållsförteckning

1	Signaler.....	7
1.1	Vad är en signal?	7
1.2	Viktiga signaler	11
1.3	Digitalisering av analoga signal	19
1.4	Modulering	24
2	Tid- och frekvensrepresentation	30
2.1	Fouriers transformation	30
2.2	Diskret Fouriers transformation	34
2.3	Spektrum	36
2.4	Ortogonal transformation	43
3	Analog- och digitalteknik	48
3.1	Multiplexing	48
3.2	Linjära tidsdiskreta och invariantsystem	52
3.3	Filter	59
4	Kommunikationskanal.....	69
4.1	Kommunikationskanal.....	69
5	Komprimeringsmetoder	72
5.1	Komprimering av audiosignaler	75
5.2	Komprimering av tal	81
5.3	Komprimering av stillbilder	87
5.4	Komprimering av video.....	94
6	Multimedia bearbetning Multimedia processing Multimediala bearbetningar.....	103
6.1	Talsyntes.....	103
6.2	Bildigenkänning	106
6.3	Ansiktsanimering.....	109
6.4	Taligenkänning	112
6.5	Multimodal gränssnitt.....	115
7	Exempel i Matlab	118
7.1	Exempel i Matlab	118

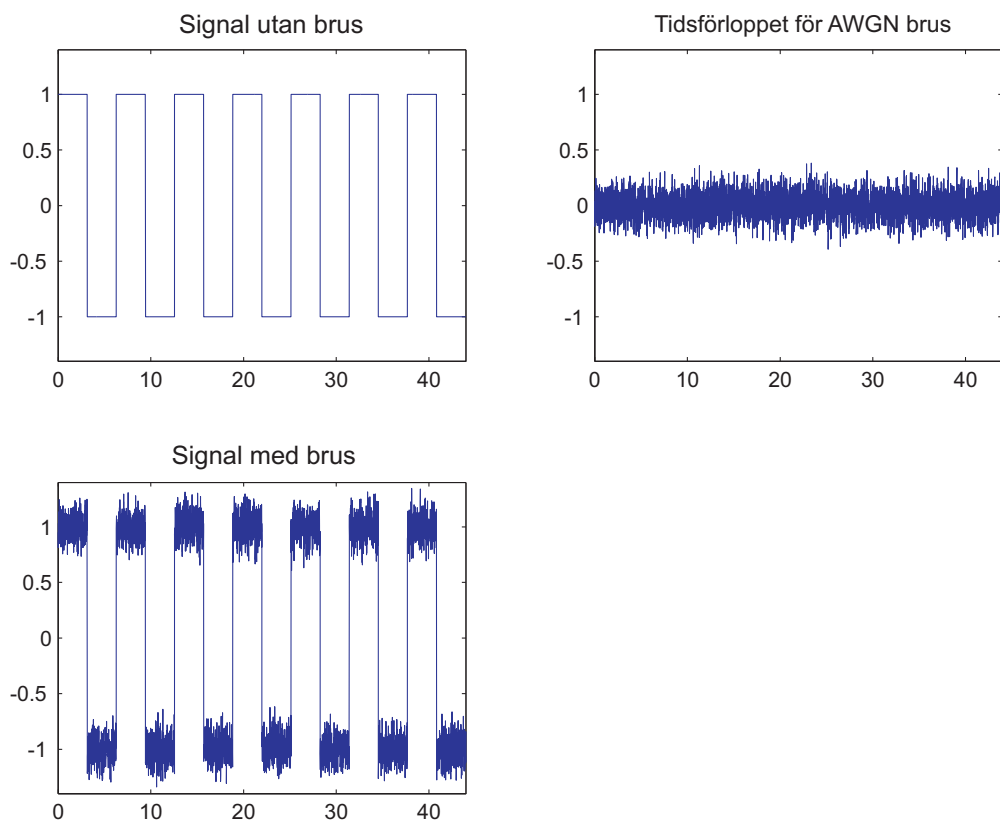
1 Signaler

1.1 Vad är en signal?



Innebörden av ordet signal in denna kurs beskriver en funktion som övergripande och uttömmande beskriver beteende av ett specifikt fenomen.

Signal i den fysiska världen är i principen vilken storhet som helst, storhet som visar ändring i tiden (t.ex. såsom röst) eller i utrymmet (t.ex. såsom bild), vilken kan ge information om den fysiska systemets statusen, eller rapportöverföring mellan observatörer. De verkliga signalerna är alltid ihopblandade med brus



Verklig signal – signal+brus (bakgrundsstörning)

På det elektrotekniska området samt signalbearbetningen skiljer vi mellan två typer av signaler – analoga och digitala signaler.

Analog signal



Analog signal är en associerad signal, som teoretiskt kan lägga till sig vilket värde som helst.



Analog signal kan lägga till sig godtyckliga tidsvärden och amplituder. Dessa representerar beteende med typiska egenskaper eller fenomen i tiden.

Den typiska analoga signalen är en elektrisk signal eller en dagstemperatur ändrad med tiden. Analog signal kan vara deterministisk (vi kan definiera den exakt t.ex. med en matematisk formel), eller stochastisk (händelsen är slumpmässig och vi kan inte förutsäga den).

Digital signal



Digital signal är representerad med ett vanligen fördefinierade diskret sekvensvärde.



Digital signal kan bara ha ultimata mängder med stickprover under ett bestämt tidsavsnitt. Ett enkelt exempel på hur man kan få diskret signal, är matrisassocierad med (analog) signal. Exempel på ett digital signal är lufttemperaturen mätt varje 5:e minut, eller också nollor och ettor vilka används i PC:n.

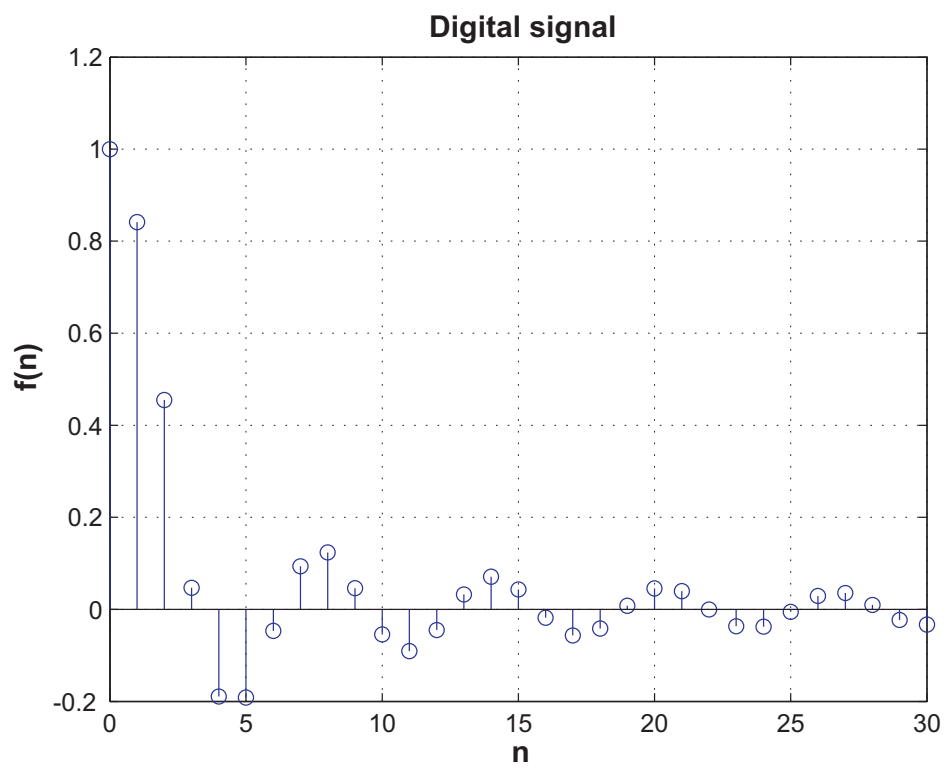
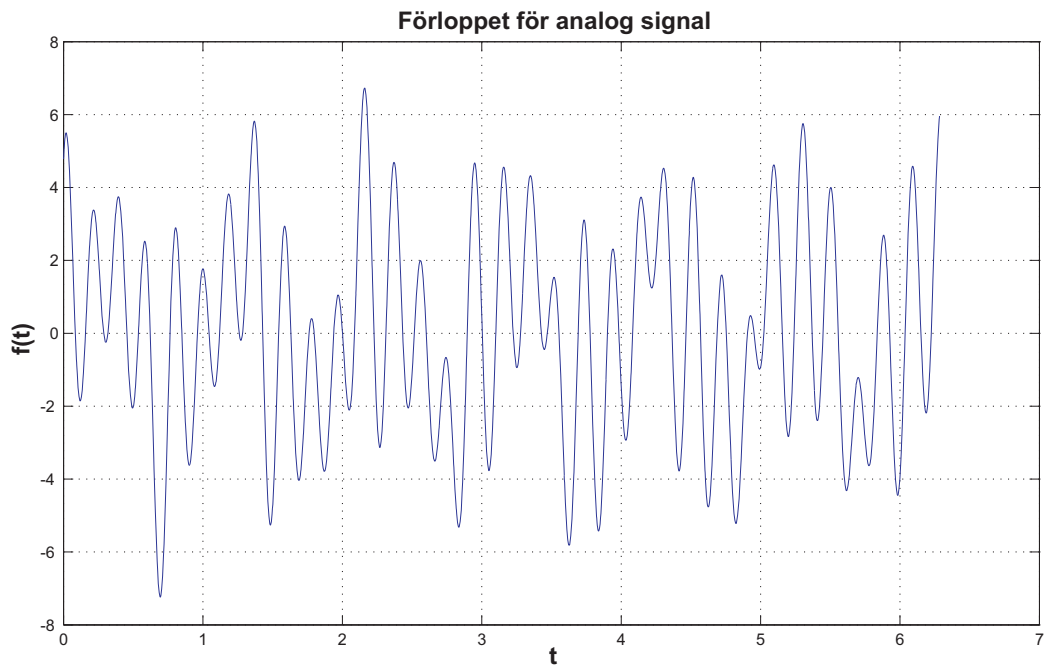
Alla processer i naturen är analoga (föreställ dig t.ex. grafik som visar temperatur ändrad med tiden, eller bilhastighet som ändrar sig under en tidsintervall).

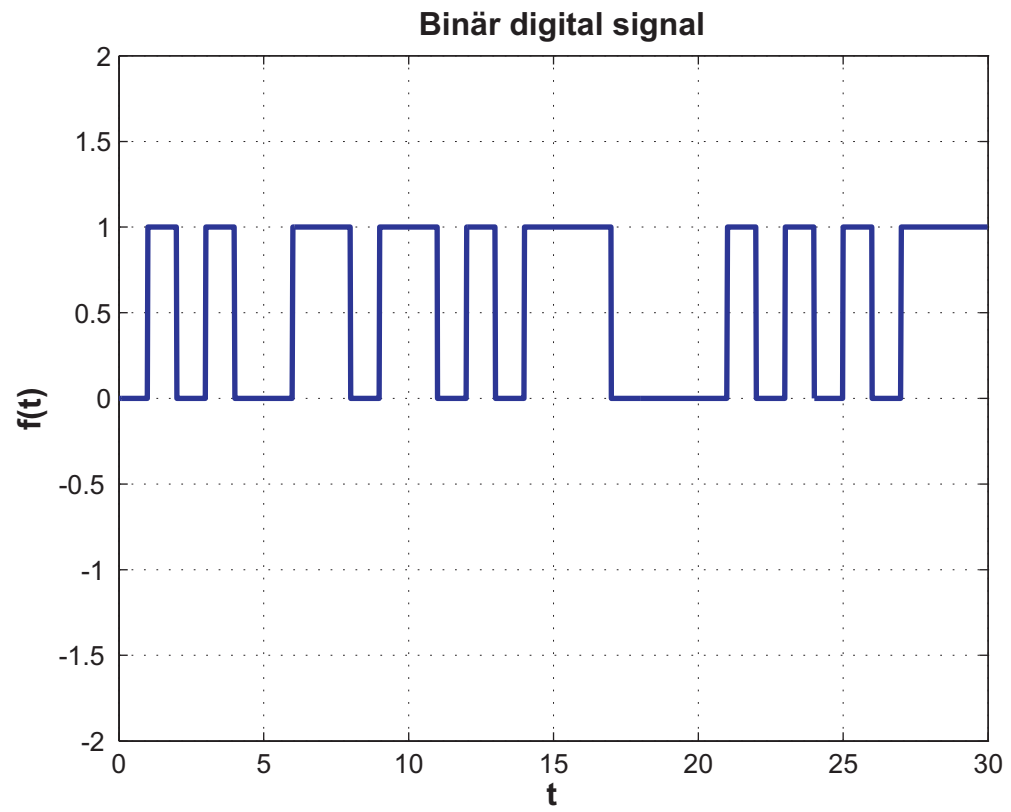
+

Den viktigaste fördel med analoga signalen är det faktum, att den om den utsätts för bearbetning inte förlorar någon information. Dock är digitala signaler mycket enklare att bearbeta (t. ex. CD:n med musik är digital – och kan bli mycket enkelt konverterat till MP3). Digital signal är också mindre sårbar för brus.

–

Analog signal är mer komplex och mer svårarbetad (tänk på en vinylplatta, kvaliteten är bättre än på CD:n, men prova lägga låten från en platta över till en telefon). Fördel med analog signal är, att den inte förlorar information som digital signal med matris processen och kvantisering (föreställ dig en tabell för temperatur med ett värde bara för en dag – det blir bara 24 numeriska värden).





Exempel på en analog och en digital signal

1.2 Viktiga signaler

I föregående delen förklarades betydelsen av ordet signal. Kapitel nämde inte fallet med signaler med mer än en dimension. Denna kapitel beskriver de mest viktiga signalerna på område med digitalbearbetning av signaler och multimedia.

Endimensonelasignaler



Signal, vilken har funktion av en oberoende variabel, kallas för endimensionalsignal. Vanligen är variabel t tiden (t. ex. $f(t)=5t$). I fallet med diskreta signaler ersätter tiden en oberoende diskret variabel n (t. ex. $f(n) = n+1$).

I följande definition representerar formel x mängden av verkliga nummer $\{R\}$ och n representerar de naturliga numren $\{N\}$.

Diracs deltafunktion δ är en allmän funktion av verkliga nummer, där varje värde med undantag av noll är lika med noll. Delta funktion visas ibland som en oändlig hög och oändlig tunn linje baserad på noll. Om vi skulle beräkna area, vilken denna impuls omfattar, får vi numer 1. I område med signalbearbetning är funktion ofta känd under symbolen enhetsimpuls.

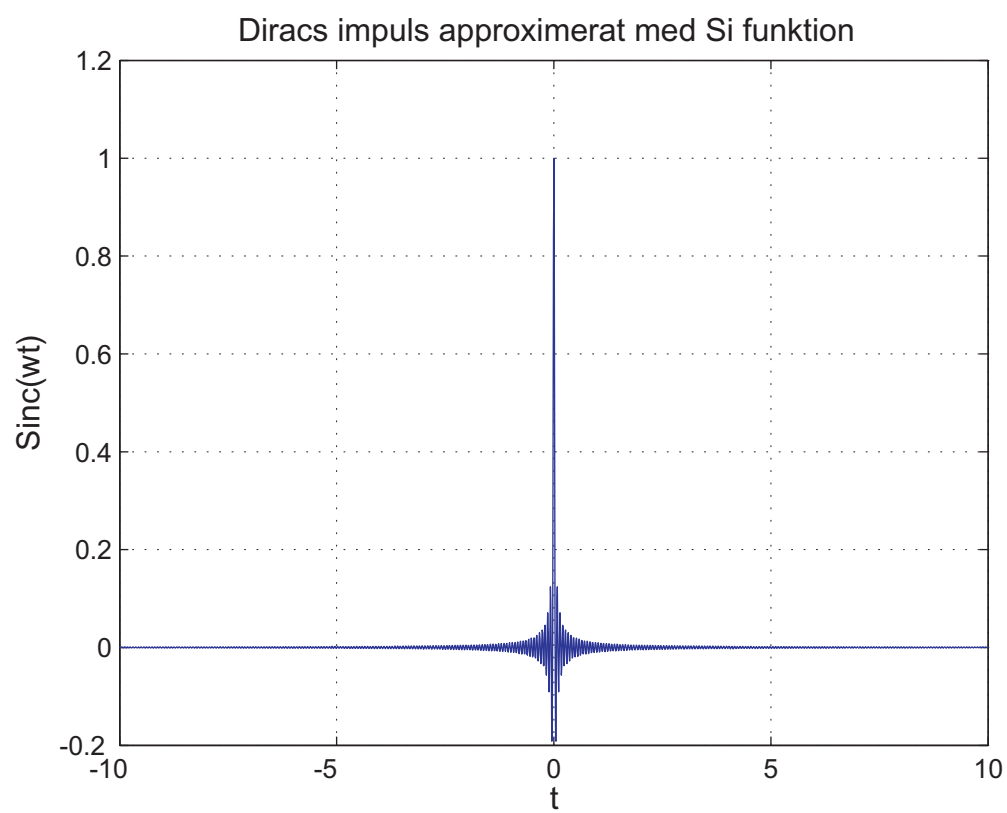
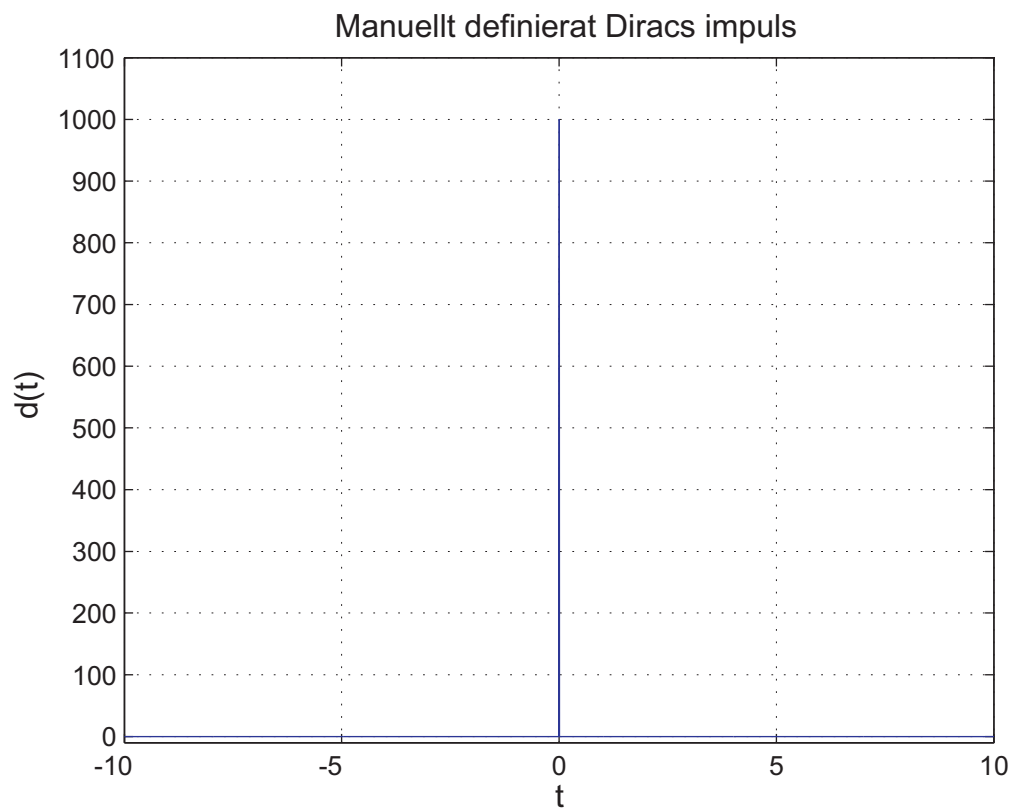
Matematisk definition:

$$\delta(x) = \begin{cases} +\infty & x = 0 \\ 0 & x \neq 0 \end{cases}$$

ibland räcker det med reducerad definition

$$\int_{-\infty}^{\infty} \delta(x) dx = 1.$$

I bilden nedan, visas en idealisk och approximativ Diracs deltafunktion. Approximativ Diracs funktion visas i nästa bild i syfte att bättre förklara hur man kan i den verkliga världen få Diracs deltafunktion.

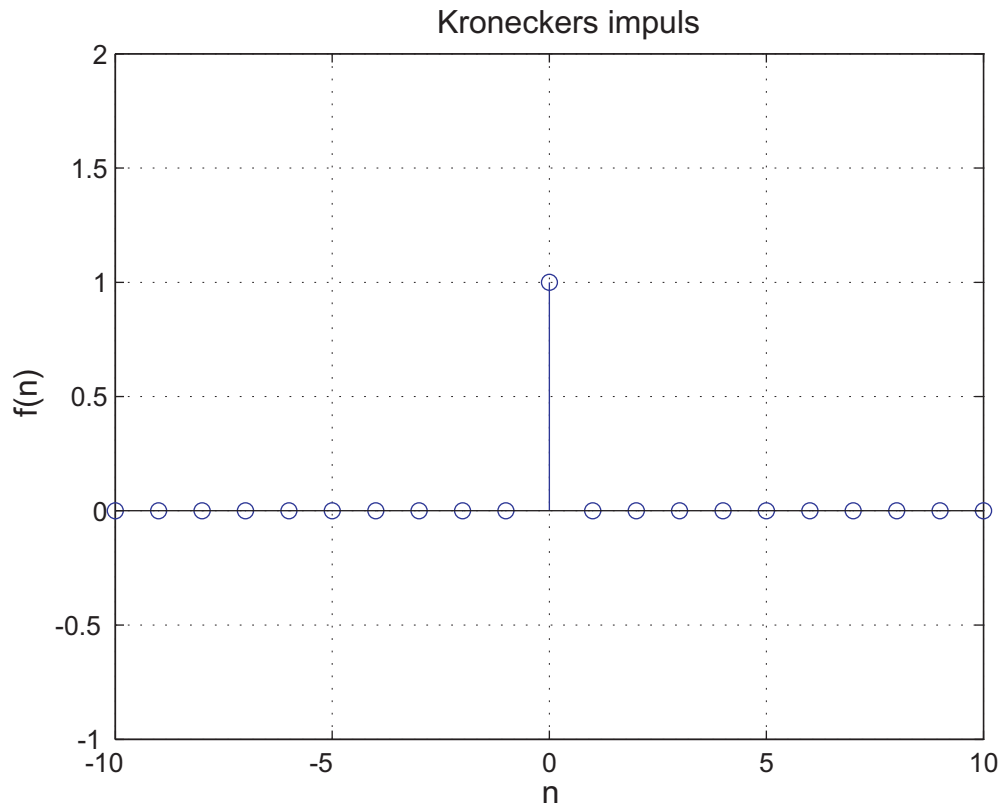


Dirac deltafunktion – idealisk och närmande funktion $\text{sinc}()$.

I det avskilda (diskreta) området motsvarar den så kallade Kroneckers deltafunktion Diracs deltafunktion. I arean för bearbetning av digitala signaler, är funktionen känd som impuls, resp enhetsstegimpuls. Och när denna stimulerar signalprocessingselement kallas resultatet för elementets impulsrespons.

Matematisk definition:

$$\delta(n) = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$$



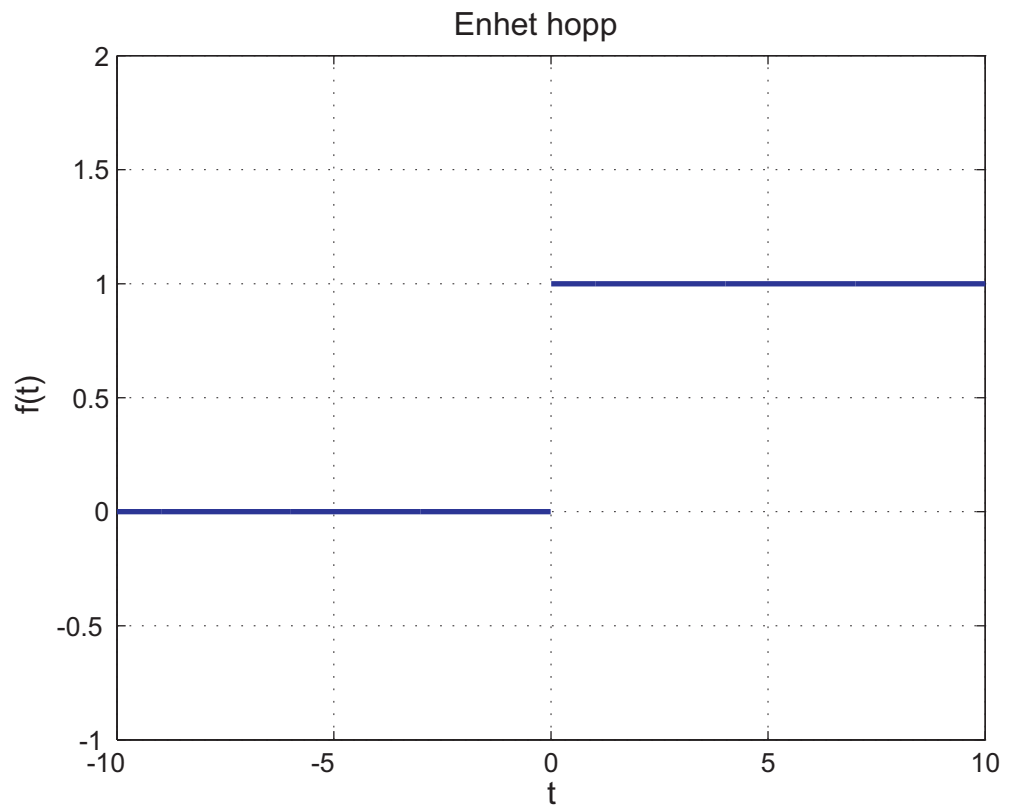
Kroneckers deltafunktion

Enhetsstegfunktion betecknad som u , är en osammanghängande funktion, vilken har ett nollvärde för alla negativa argument (negativa nummer) resp för alla positiva argument. Funktion används för att representera signal vilken kopplas över en specifik tid och förblir så till oändligheten. Enhetsstegfunktion är integralen av Dirac deltafunktion.

$$u(x) = \int_{-\infty}^x \delta(s) ds$$

Matematisk definition:

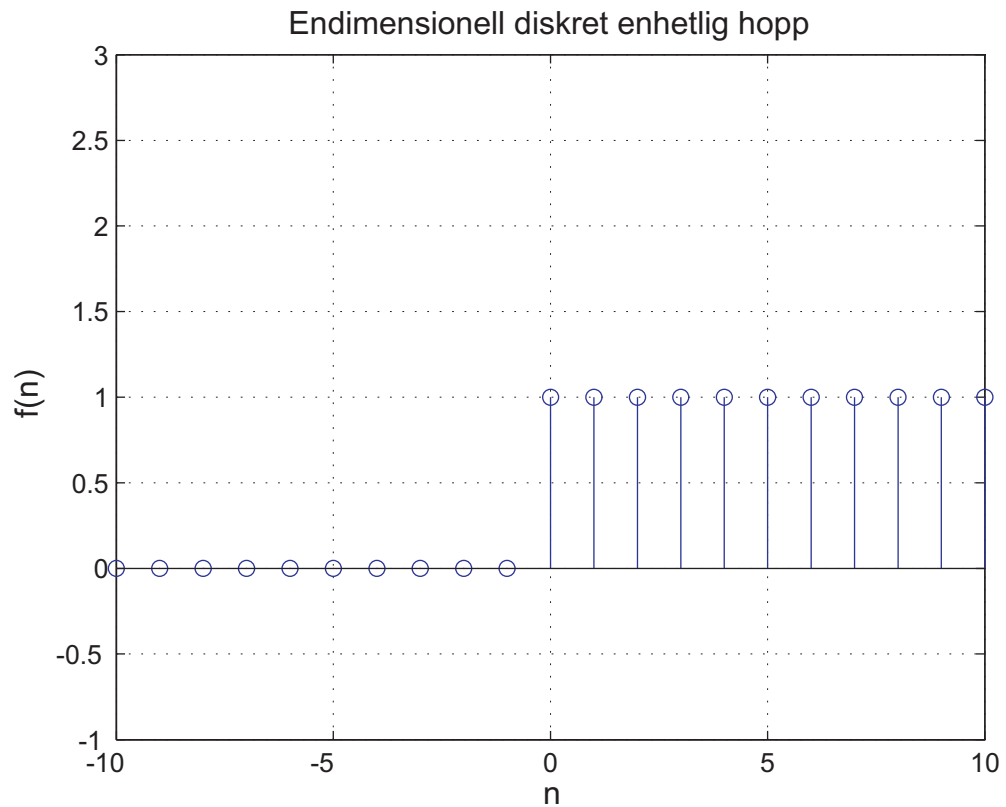
$$u(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$



Enhetsstegfunktion (kontinuerlig)

Diskret enhetsstegfunktion:

$$u(n) = \begin{cases} 0 & n < 0 \\ 1 & n \geq 0 \end{cases}$$



Periodiska signaler är nästa speciella signalgrupp. Periodisk funktion (vilken beskriver upprepade signaler) är en funktion vilken repeterar sina värden i regelbundna intervaller. Till detta domän tillhör alla de trigonometriska funktioner (sinus, kosinus, tangens, kotangens – alla har perioden 2π). Om perioden betecknas med P , är sedan den matematisk definition för periodisk funktion:

$$f(x) = f(x + P)$$

Tvådimensionela signaler

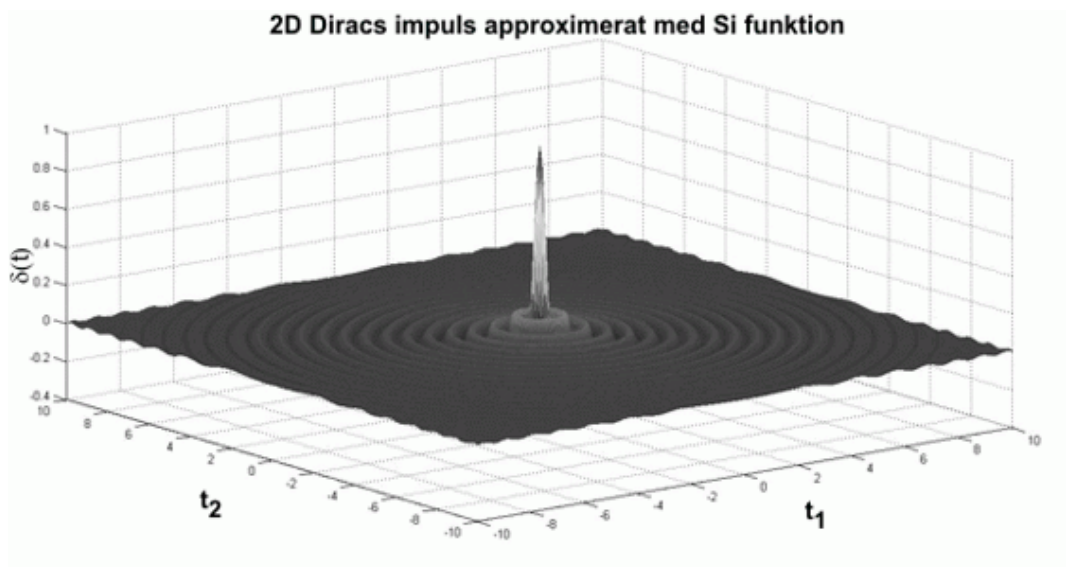
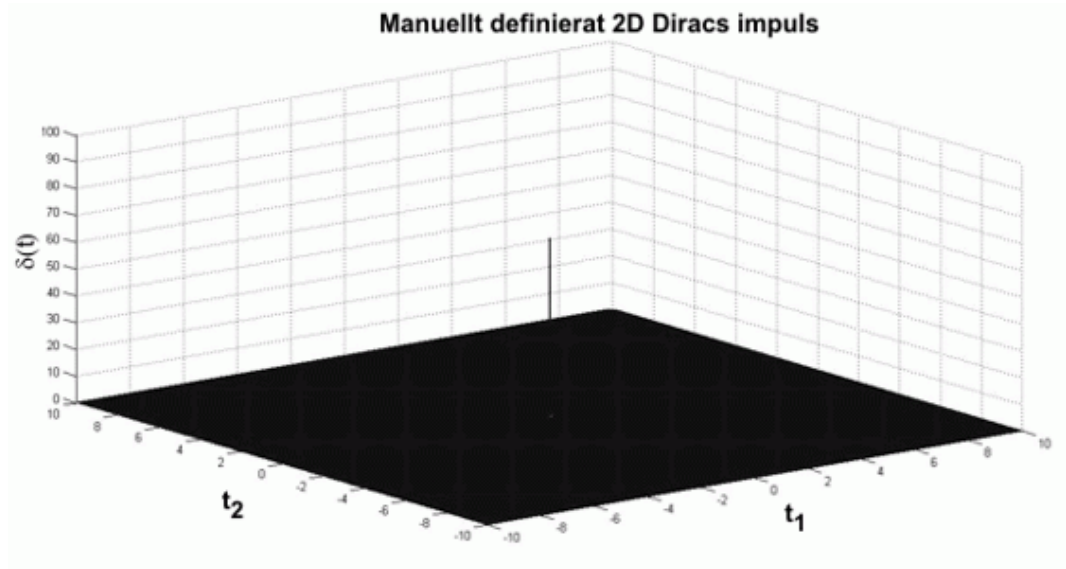


En signal som är funktionen av två oberoende variabler kallas för tvådimensionell signal. Typisk exempel på tvådimensionell signal är en bild. En bild är komponerad av ljus och luminiscenserande signal. 2D bild kan altså blir kontinuerlig rumslig t. ex. klasisk foto eller bild eller bilden kan blir diskretised in i ett utrymme, såsom digital foto t.ex..

Alla viktiga signaler definierade för endimensionela domäner är samtidig definierade också för tvådimensionela domän. Här tar vi upp bara matematiska definitioner.

Diracs deltafunktion

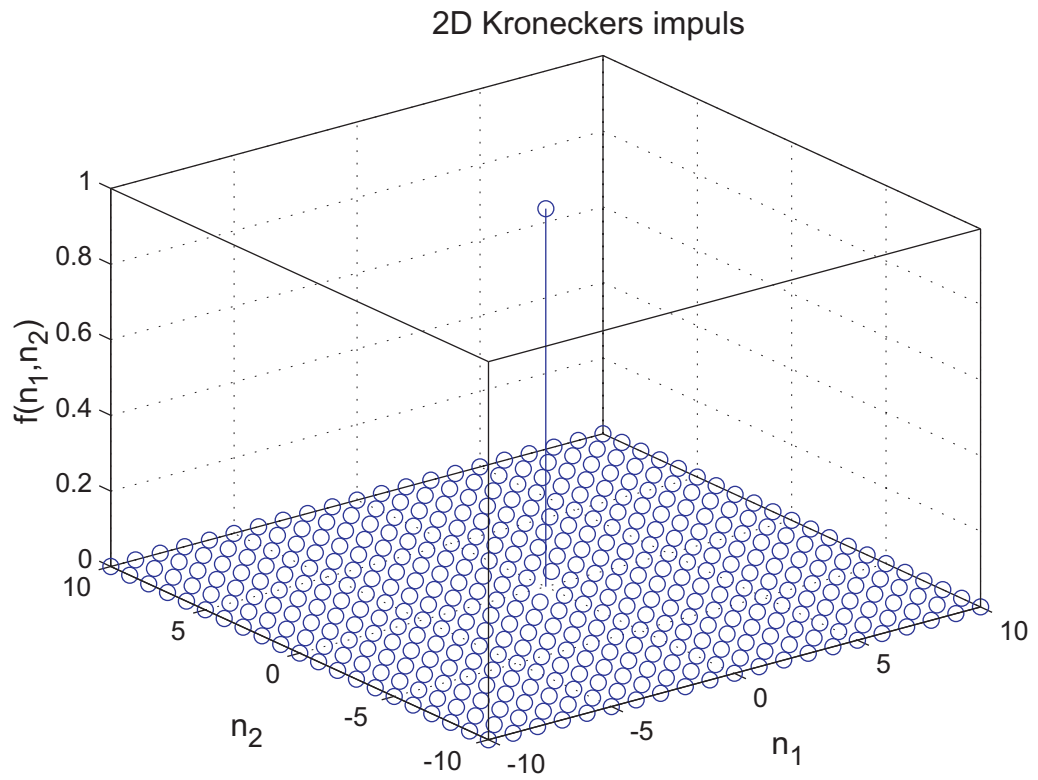
$$\delta(x_1, x_2) = \begin{cases} +\infty & x_1, x_2 = 0 \\ 0 & x_1, x_2 \neq 0 \end{cases}$$



2D Diracs deltafunktion – idealisk och approximated funktion *sinc()*.

Kroneckers deltafunktion

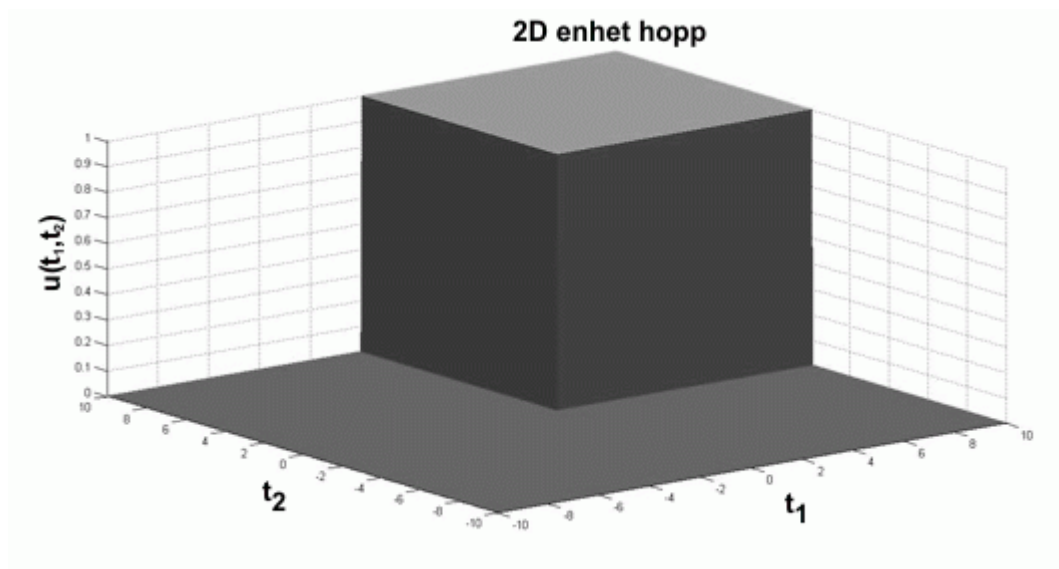
$$\delta(n_1, n_2) = \begin{cases} 1 & n_1, n_2 = 0 \\ 0 & n_1, n_2 \neq 0 \end{cases}$$



2D Kroneckers delta

2D Enhetsstegfunktion (kontinuerlig)

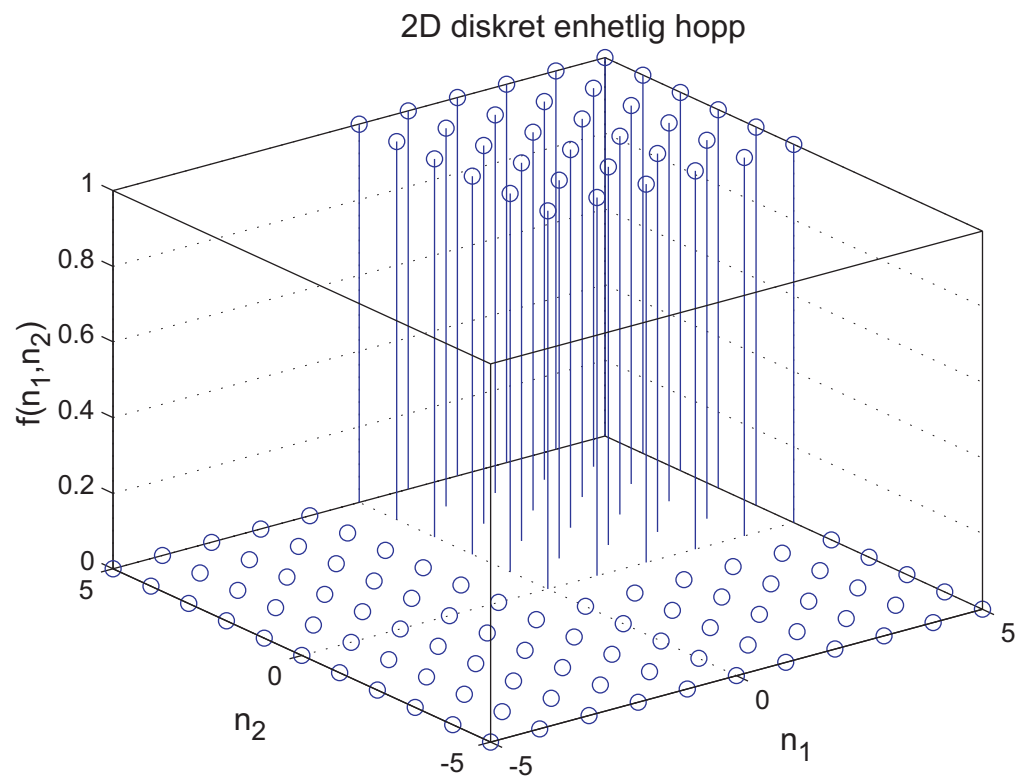
$$u(x_1, x_2) = \begin{cases} 0 & x_1, x_2 < 0 \\ 1 & x_1, x_2 \geq 0 \end{cases}$$



(kontinuerlig)

Enhetsstegfunktion (diskret)

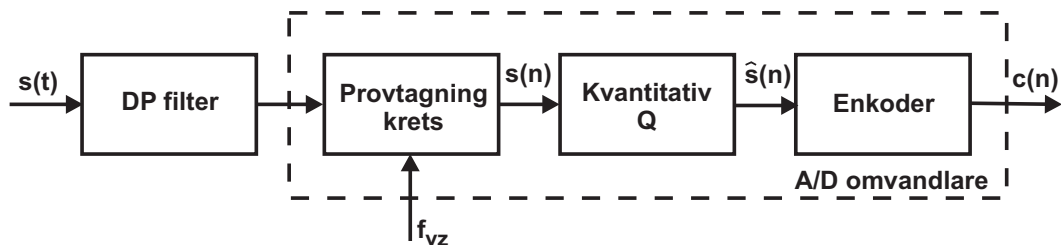
$$u(n_1, n_2) = \begin{cases} 0 & n_1, n_2 < 0 \\ 1 & n_1, n_2 \geq 0 \end{cases}$$



2D Enhetsstegfunktion (diskret)

1.3 Digitalisering av analog signal

Digitalisering av analog signalen är process, i vilken analog signal (som motsvarar t. ex. ljud, video, bild,...) är transformerad till digital form. Signalen är klassificerad, kvantiserad och kodad. Resultat är sekvens (följdsekvens) av binära nummer, vilka kan vidare bearbetas.



Digitalisation av en analog signal (PCM metoden)

Huvudmetoder för kodning av multimediala signaler i en tidsarea som används i multimedial telekommunikation, är följande:

- pulskodad modulation (PCM),
- differential PCM (DPCM),
- adaptiv DPCM (ADPCM).

Signalen är efter en avgränsning i DP filter (antialiasing filter) sorterad in i en provkrets och med detta får vi följdsekvenser av stickprover.



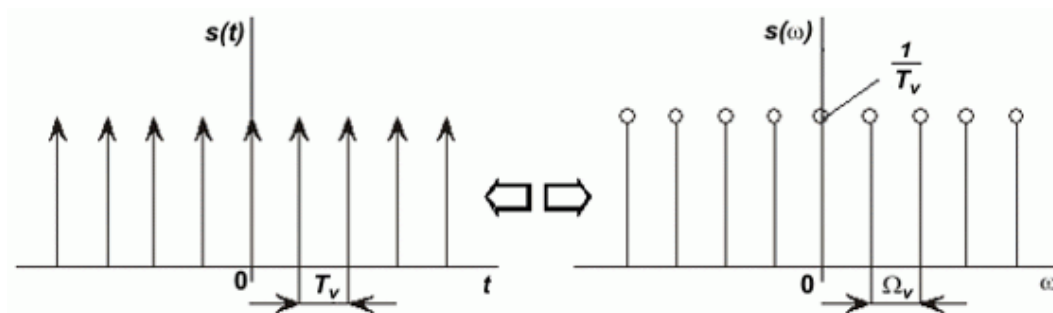
I provkretsen överförs kontinuerliga signaler till diskreta signaler. Prover tas upp i avgränsade tidsmoment och således motsvarar proven de värden resp värdena i den tid eller den area de formades i. Prövningshastighet är bestämd av prövningstermen (känd som Shannon-Kotelnikov term), dvs. prövningsfrekvensen måste bli minimalt två gånger större än signalvärde för den högsta frekvenssignalen.

Värdet för prövningsfrekvensen F_{vz} är bestämd av bandbredden för provsignal, t. ex. får signal för telefonsamtal med bandfrekvensen 300 - 3400 Hz (till 4 kHz) värde $F_{vz} = 8$ kHz.

Prövningsformel kan matematisk uttryckas som: $T_{vz} \leq \frac{1}{2F_m}$

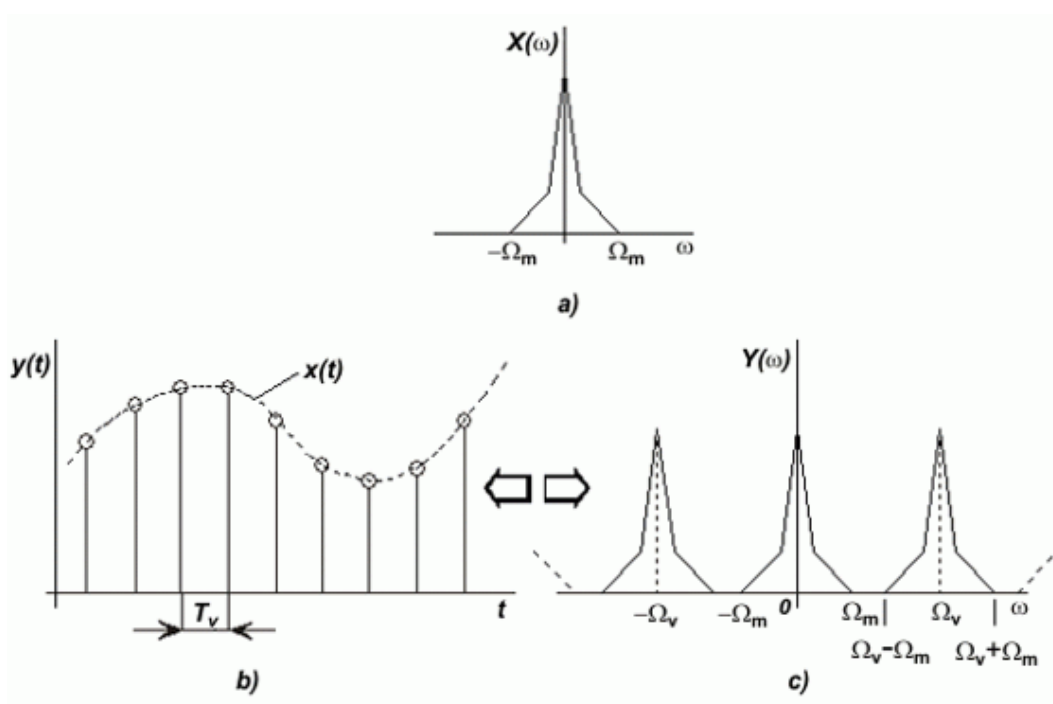
Minimal prövningsfrekvens $F_{vz\min} = 2F_m$ kallas Nyquist frekvens.

Provsignal, märk som $y(t)$, kan uttryckas som produkt från originalsignal $x(t)$ och prövningfunktion $s(t)$, vilken representerar en oändligt serie av Diracs enhetimpulser. Spektrum för prövningsimpulser har avstånd från varandra med tiden $t = T_{vz}$. I en frekvensarea är signalspektrum diskret med avstånd från frekvenskomponenter Ω_{vz} .



Prövningsfunktion i tid och frekvensarea

Spektrum $Y(\omega)$ för provsignal $y(t)$ är resultat av konvolution för spektrum $X(\omega)$ och $S(\omega)$.



a) Spektrum för originalsignal $X(\omega)$ b) Provningsignal in tidsarea $y(t)$ c) Spektrum för provnings signal $Y(\omega)$

Om repetition av spektror gäller $\Omega_v < 2\Omega_m$ för "perioden" inträffar en överlappning av spektrummappar. Det här fenomenet kallas för aliasing. Överlappning av spektror händer också när signalen är obegränsad $x(t)$. Detta resulterar i att den rekonstruerad signalen från provningar är inte detsamma som den ursprungliga originasignalen.

Nästa steg i digitaliseringen är kvantisering. Provsignaler är ersätta med behöriga kvantiserade nivåer (fasta, bestämda värde t.ex. naturliga tal), och med dessa får vi följa kvantiserade provsignal. Kvantiserade nivåer får vi genom att dela upp amplituder till kortare intervaller. Intervalllängden kallas för kvantsteg. Där alla kvantsteg har samma längd talar vi om linjär kvantisering i omvänd fall är det icke-linjär kvantisering.

- Den stora nackdelen med procesen är en s.k. kvantfel eller kvantbrus. Detta är skillnaden mellan analog input till *analog-digital konverter* (**ADC** *analog-digital convertor*) och en digital output. Detta fel är ej linjärt och är beroende av konkret signal. Felet vid kvantisering är också orsaken till problem som uppstår vid omvänd (från digital- till analogsignal) konvertering. Därför är signalen aldrig konverterad tillbaka till sin original form, den är bara approximerad från kvantnivå.

Nästa steg för digitalisering är kodningprocessen.



I i denna process tilldelas varje kvantprov en binär kodord och på detta sätt erhålls en följesekvens av de kodade orden.

PCM metoden är international standard för kodning och överföring av multimediasignaler. Grundprincipen för denna metoden är illustrerad i bilden nedanför.

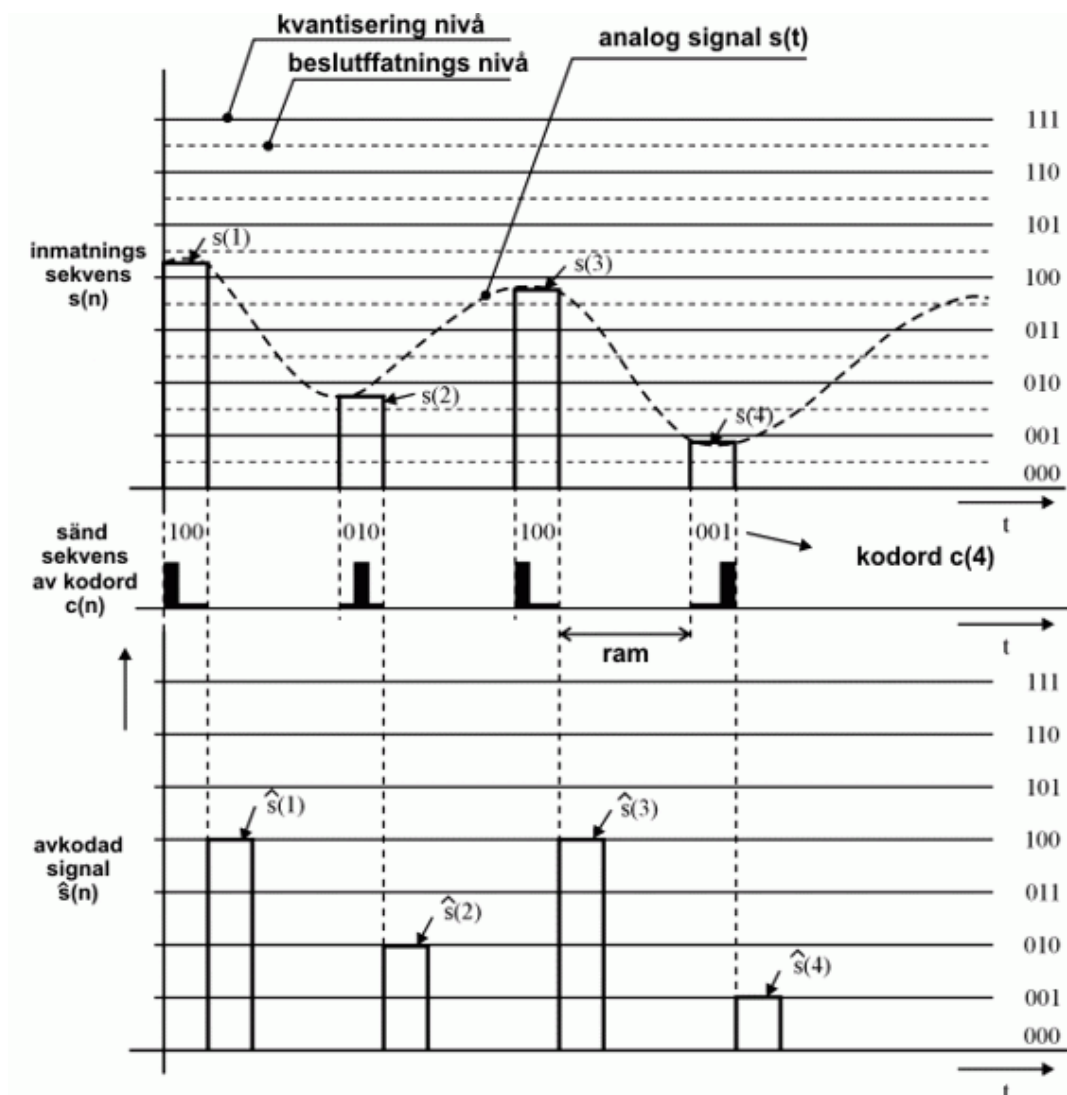


Första systemet med PCM metoder använde 7 bitar lång kodad ordet N där antal kvantiserade värden var 128. Om man antar att provningsfrekvensen är $F_{vz} = 8 \text{ kHz}$ och $N = 8$, behövs sedan transferhastighet för samtaltransfer i telefonband som är $8 \cdot 10^3 \cdot 8 = 64 \text{ kbit/s}$.

- + PCM-metoden har i jämförelse med analogmetoden för signalöverföring fördelen vad gäller motståndskraften för störningar i den överförda digitalsignalen.

- Tyvärr nackdelen med denna metod är krav på bredare frekvensbandbredd för signalöverföring.

Tidsvågformer med kodad PCM-signal visas på bilden.



Tidsvågformer med PCM kodningsmetod av multimediala signaler

Linjär PCM använder samma kvantiseringssteg under hela kvantiseringsomfattning. Av detta följer att omfattningen av den dynamiskt hanterbara inputsignal är beroende av antalet och storleken av kvantiseringssteg. Antalet kvantiseringssteg (grader) vid samma styrka av inputsignal påverkar storlek av kvantiseringsfel. Vid större antal kvantiseringssteg är storleken på kvantiseringsfel mindre men detta ökar kravet på överföringshastigheten. Denna åtgärd har begränsningar vilka kan tas bort med icke-linjära spridning av kvantiseringssteg från digitala signal till analoga sådana. Den icke-linjära PCM arbetar på detta sättet.

Icke-linjär PCM använder icke-linjär layout för kvantiseringssteg för att på detta sättet ökar storleken av högre amplitudsinputsignaler. Denna metod baseras på modifikation av dynamiskt komprimering av inputsignaler från sändningssida och dynamisk expansion på mottagarsida. Kompression på sändningssida förstärker de små värdena av stickprover samtidigt som de stora värdena försvagas. Expansionen på mottagarsidan returnerar prover i den ursprungliga omfattningen med det korrekta förhållande till deras värden.

Avkodning görs på samma sätt som kodningen fast i omvänd ordning. Output från dekodare är sekvenser av kvantiserade prover.

1.4 Modulering

Analog signal, som den var definierad i första delen, är skiftande signal associerad med tiden i en amplitud. Om vi skulle t. ex. presentera växelström grafiskt, skulle den se ut som en våg, med skiftande spänning från minus till plus. Det existerar tre faktorer, vilka beskrivs närmare nedanför: frekvensen, amplituden och fasen.



En frekvens är hastigheten med vilken ström oscillerar över och under nollnivån med andra ord växlar sin polaritet.

När ström kommer över nollvärde rör den sig tillbaka under nollvärde och därefter går den till noll igen, vi säger att den gjorde en „cykel“. Frekvensen anges med antal cykler per sekund och enheten benämns Hertz (Hz). Om analogsignal har 500 cykler per sekund, säger vi säger frekvensen är 500 Hertz (500 Hz).



Amplituden är den högsta höjdpunkt och djupaste djupet på den grafiska vågen.

Alteftersom analogsignal sprider sig igenom ett utrymme minskar amplitudvågen. Denna egenskap kallas dämpning. Analogvågor är inte så påverkade av dämpningsproblem men måste ändå ibland förstärkas. Amplituden i analogvågor mäts i watt, amper eller volt. Enheten för mätningar är oftast decibel. Decibel (dB) tillåter enkla jämförelse mellan två olika styrkor av amplitudssignaler.



Fasen beskriver skillnaden mellan den status i början av en cykel i jämförelse med statusen i påföljande cykel dvs. beskriver förskjutningen mellan två cykler. En signal är alltid referenssignal, andra signal är fasförskjutning.

Fasens förskjutning av signal skapar små förseningar av signaler och denna leder till att toppar och dalar är icke-synkrona i jämförelse med referenssignal. Nivån på icke-synkronisering mäts i grader. När fasförskjutningen av en signal är 180 grader innebär detta att signalförskjutning börjar genom att referenssignal efter nedstigningen från toppen nudar noll.

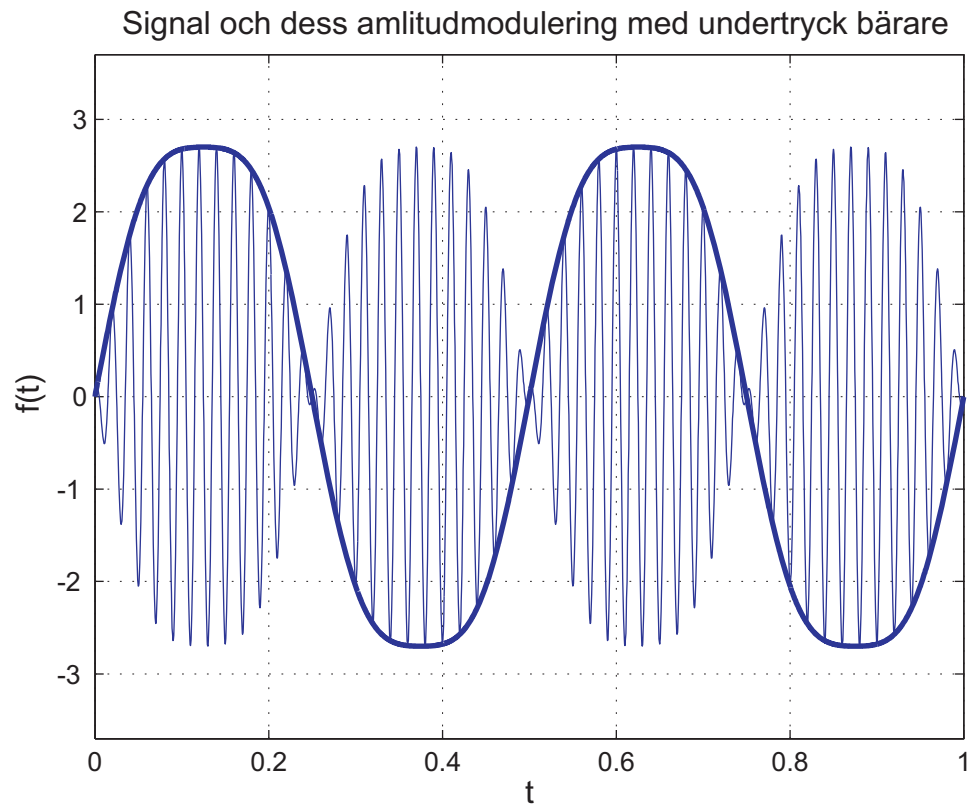
Det betydelsefulla då man arbetar med amplituden, frekvensen och fasen ligger i att dessa tre komponenter kan varieras så att analogsignal kan överföra information genom ändringar av dess egenskaper i tiden.

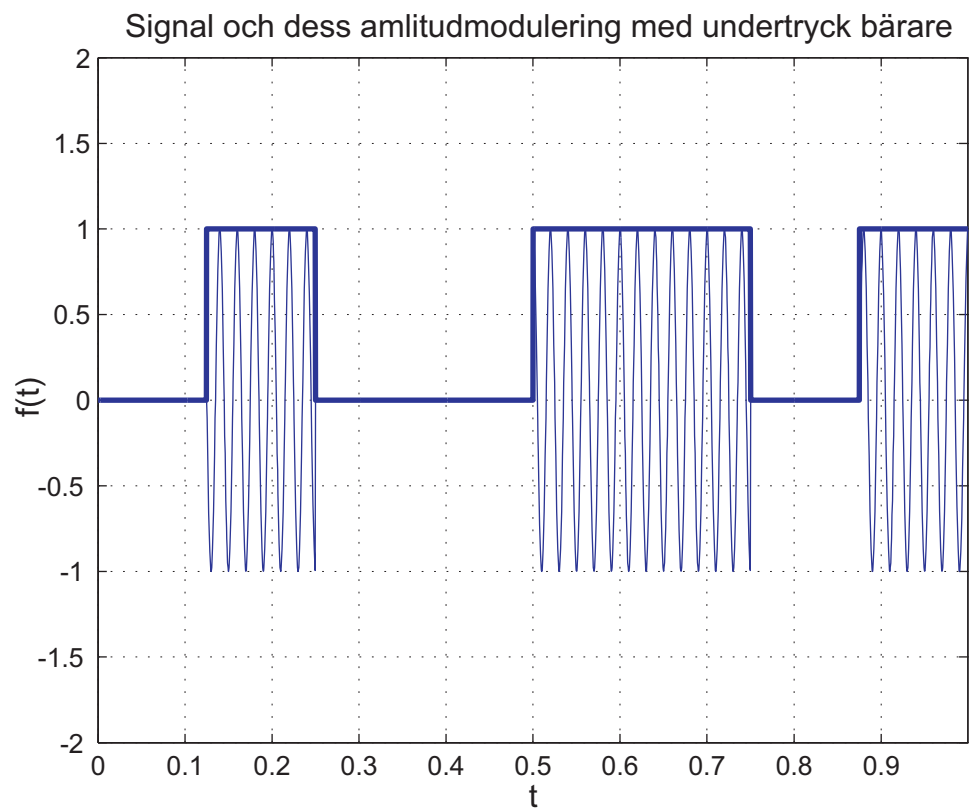


Modulering är insättningprocess av information genom att ström av digitalbites eller analog audiosignal inuti en annan signal kan bli fysisk överförd vidare.

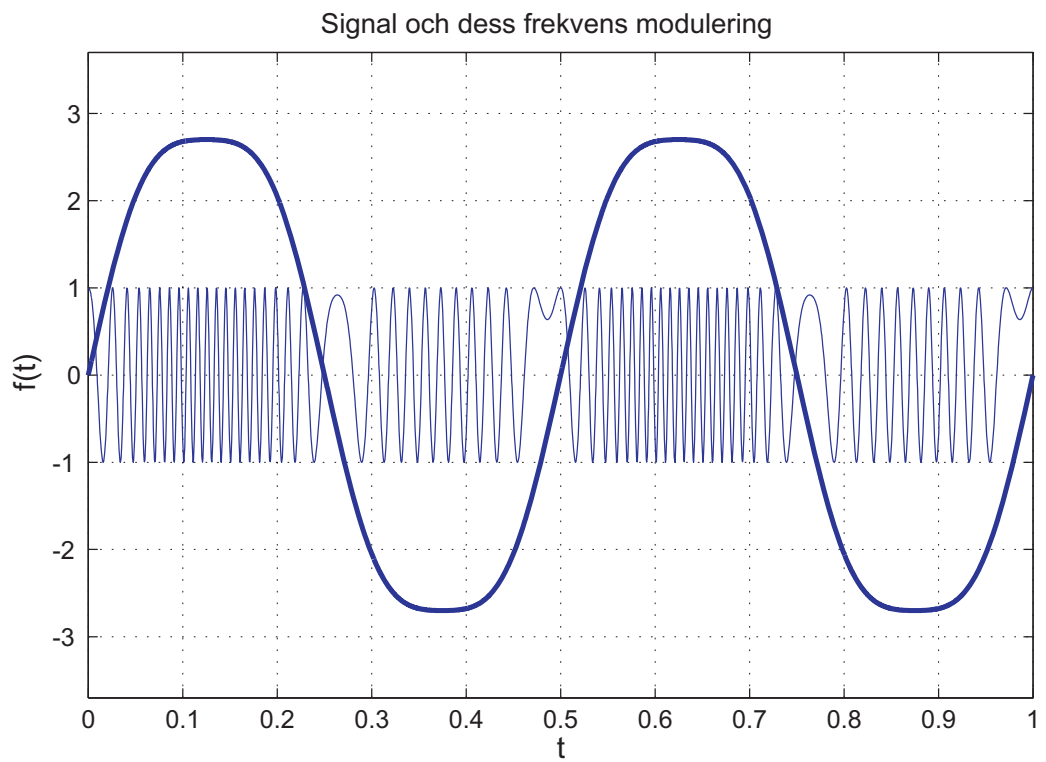
En anledning till modulering är möjligheten till överföring av flera signaler i en kanal med begränsad bandbredd. Varje signal är modulerad med individuell frekvenssvängning. Signalmoduleringen påverkas av möjlig tillgång till en viss sub-bandbredd. Så den största fördelen är att ett överföringssätt t.ex. genom en optisk kabel kan tas i anspråk av många olika signaler.

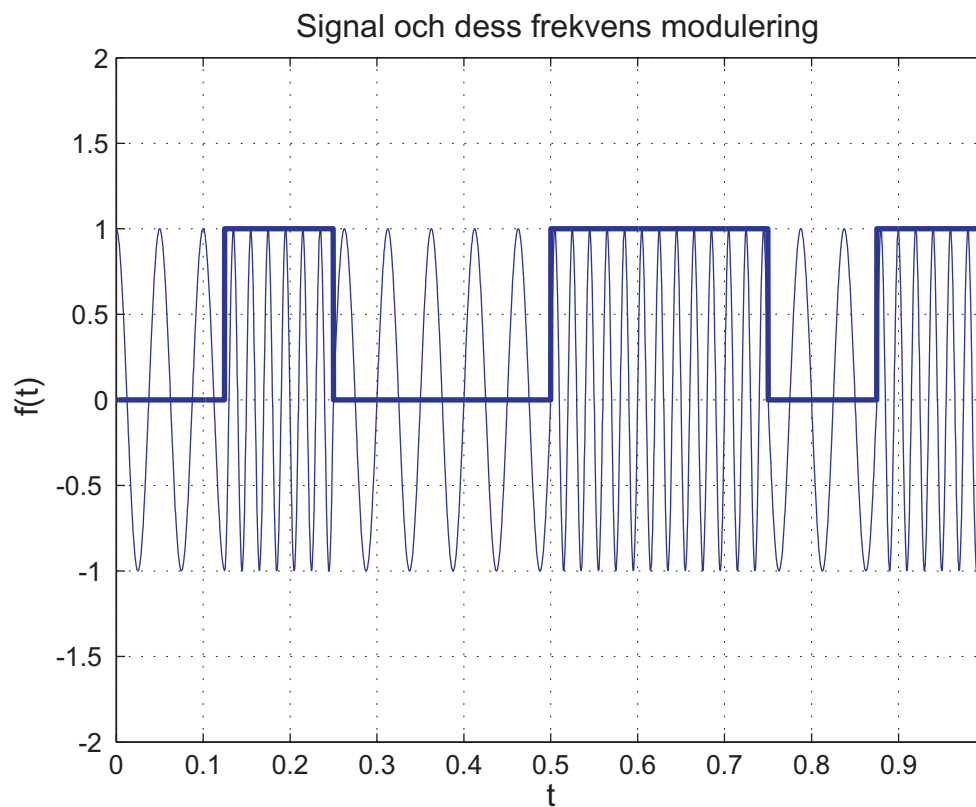
Vid analog modulering tillämpas moduleringen kontinuerlig efter de svar analoginformationssignal lämnar ifrån sig. Det finns många olika typer av analoga moduleringar, de enklaste är *amplitudmodulation (AM)*, *fasmodulation (PM)* och *frekvensmodulation (FM)*.



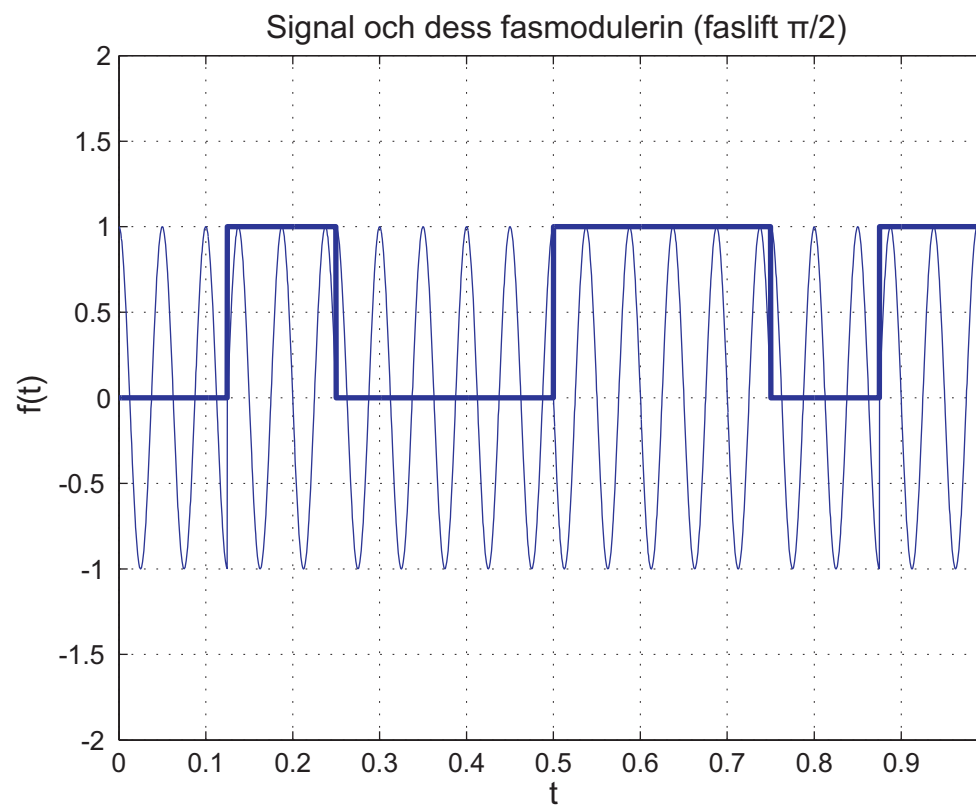


Amplitudmodulation av analog och digital signal





Frekvensmodulation av analog och digital signal



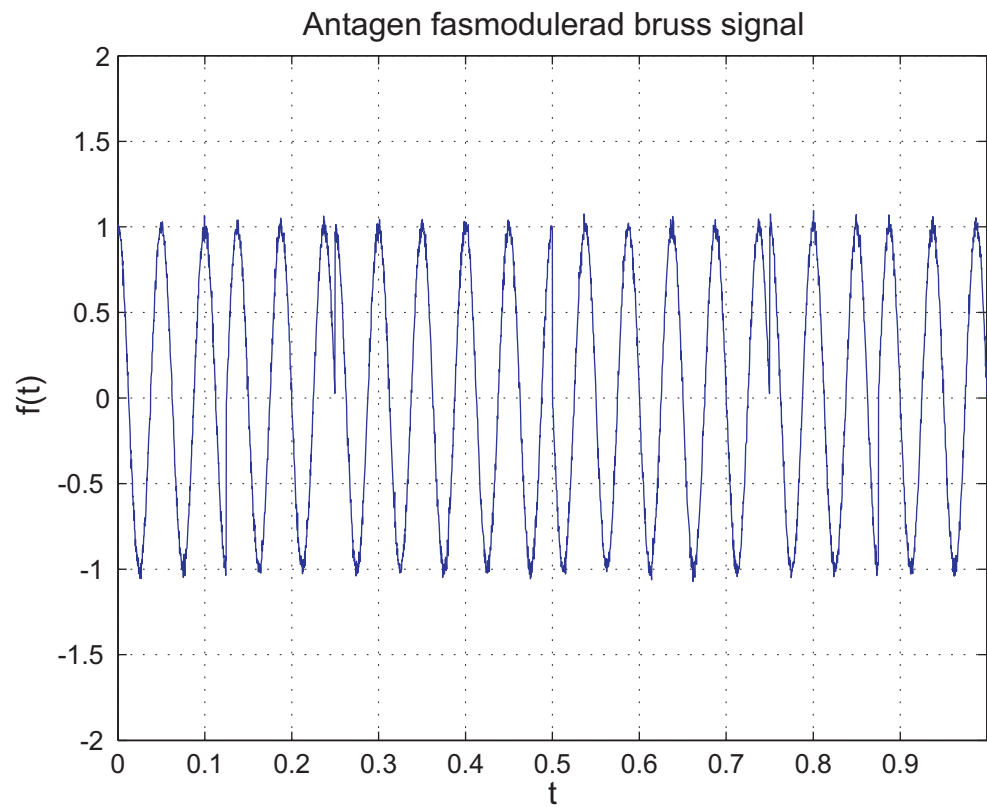
Fasmodulation av digital signal

AM fungerar på grundval av amplitudändringar av överförd signal i en sänd information – amplituden är direkt proportionell till amplituden i informationssignalen. Bärvågen är modulerad (modifierad) inputsignal för att överföra information. Bärvågen har vanligtvis mycket högre frekvens än ingångssignalinformation. Bärvågens amplitud är modulerad med inputsignal (dvs. överförd information) före själva överföringen. Ingående våg modifierar amplitud av bärvåg och med det definierar “det förpackningskuvert” under hela händelse.

FM överför information över en bärvåg med hjälp av modifiering av en omedelbart frekvens. Vid analoga moduleringar är ändringar mellan de omedelbara- och standardfrekvensens bärvåg direkt proportionell till värdet av ingångssignals- amplituden.

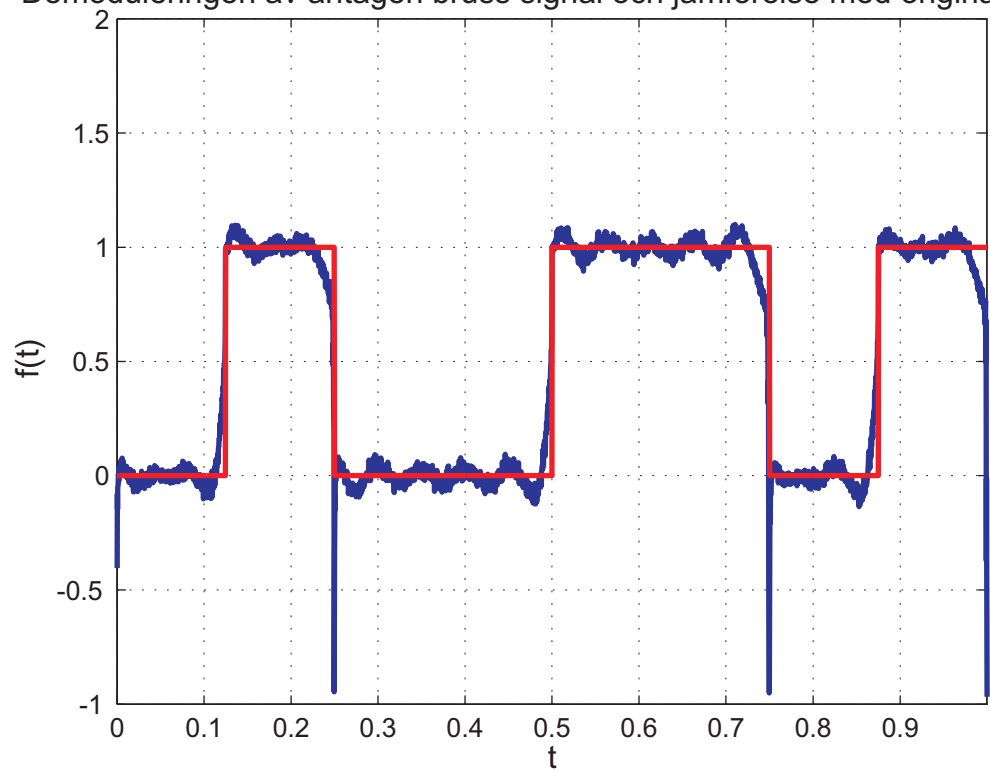
PM är en form av modulering där information representeras som en omedelbart ändrig i fasens bärvåg. Att uttrycka det enkelt, modifikation av fasen enligt datasignal resulterar i fasmodulering. PM är inte den vida använd teknik vid radiosändningar på grund behovet av en komplex mottagande enhet. Med PM behöver man åtgärda tvetydigheter som uppstår t.ex. vid fasändringar $+180^\circ$ eller -180° . PM används framgångsrik i digitala musiksyntar.

Demoduleringen är en process som syftar till att utvinna information från modulerad bärsignal. Demoduleringen är självklart beroende av, vilken signalparametr var modifierad, och vilken moduleringen har använts (amplitud-, frekvens- eller fasmodulering). Så för amplitudmodulerad signal är det möjligt att använda synkrondetektor. Å andra sidan, t.ex. för FM och PM modulering måste man använda en lämplig FM- respektive PM-demodulator.



Mottagen modulerad signal

Demoduleringen av antagen bruss signal och jämförelse med original



Demodulerad signal jämförd med originalsignal

2 Tid- och frekvensrepresentation

2.1 Fouriers transformation

Fouriers transformation efter Joseph Fourier, tillämpas i matematiken, fysiken och diverse andra tekniska vetenskap.



Enkelt kan vi säga att Fourier transformation representerar en frekvensberoende matematisk tidsfunktion. Denna funktion är känd som frekvensspektrum.



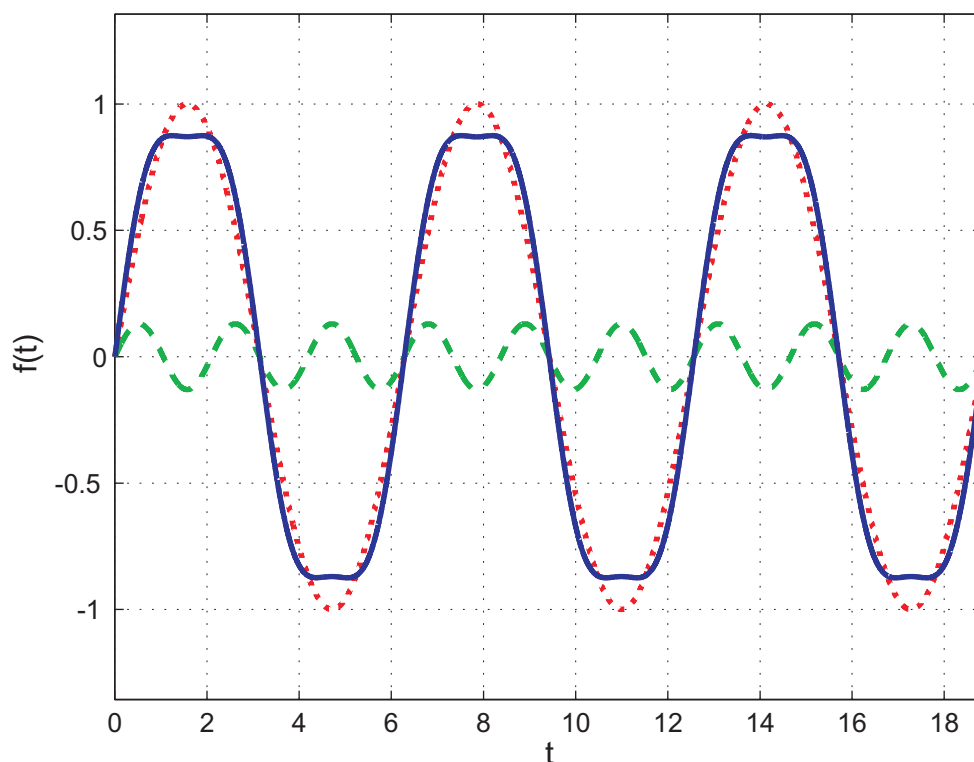
Viktig att notera: Fouriers transformation används endast för icke-periodiska analoga signaler. För periodiska analogasignaler används Fourierserien.

Låt oss anta att det finns en funktion $f(t)$, vilken kartlägger vissa tidsvärde t till något värde $f(t)$.

Vi provar approximera f , som summa av enkla harmoniska svängningar, dvs. sinusoida vågor för viss cirkulär frekvens ω . Självklart, finns det frekvenser vilka är lämpliga till approximation f samt andra som approximerar mindre bra. För detta behöver vi någon lämplig funktion $f(\omega)$, som förmedlar i vilken utsträckning den angivna oscilationen med frekvensen ω är representerad i approximationen f .

Ta t. ex. funktion (visas med svart linje) från:

Demoduleringen av antagen bruss signal och jämförelse med original



Två harmoniska frekvenser skapar signal

vilken är definierad som $f(t) = \sin(t) + 0,13\sin(3t)$. Svängningar (prickade linje) med $\omega=1$ har störst påverkan på resultatet, då kan man säga att $F(1) = 1$. Druhá vlna ($\omega = 3$, streckade linje) har mindre påverkan och då amplitud är mycket mindre. Man kan säga att $F(3) = 0,13$. Nästa frekvens behöver inte överhuvudtaget bli angiven i approximation man skriver ned $F(\omega) = 0$ för denna.

När man känner att $F(\omega)$ är inte bara lämplig för några frekvenser men lämplig för alla möjliga frekvenser ω , kan vi approximera perfekt vår funktion $f(t)$. Det är detta som Fouriers transformation kontinuerligt möjliggör.

Fouriers transformation tar en tidsfunktion $f(t)$ och returnerar behörig funktion $F(\omega) = \mathbf{FT}(f)$, dvs. dess *Fourier transformation*. FT beskriver till vilken nivå är någon från dessa angivna frekvenser representerad i funktion f . Det är bara annan representation av $f(t)$ med samma information, men med registrering i annan domän (t.ex. frekventiell). Oftast kan problem blir löst enkelt i en annan framställning (genom att kunna hitta lämplig koordinatsystem).

Angiven Fouriers transformation kan integreras över alla frekvenser, sätta ihop balanserade sinusformade vågor och får tillbaka vår f . Denna process kallas för *omvänd Fourier transformation* $\mathbf{IFT}$.

Viktigast är att Fouriers transformation har många lämpliga matematiska egenskaper (t.ex. konvolution är i frekvens- domänen bara multiplikation). Oftast

är det mycket enklare att arbeta med Fourier transformation än med bara självaste funktionen. Alltså man utför en enkel matematisk operation med filtrering, transformation och manipulation med sinusvågor och sedan transformerar allt tillbaka igen.



Anta att man vill göra brusreducering på en digital bild. I stället för att manipulera med bildfunktion: Pixel→Ljus, transformerar man hellre det hela och arbetar med $F(\text{bild})$: Frekvensen→Amplituden. De delar av höga frekvenser, vilka är orsaken till brus kan tas enkelt bort – $F(\text{image})(\omega) = 0, \omega > \dots \text{Hz}$.



Fouriers transformation (vanligen känd som framåt transformation) är definierad som:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt$$

och bakåt (omvänd) transformation är definierad:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{j\omega t} d\omega$$

Fouriers funktion $F(\omega)$ är en frekvensiell representation av signal $f(t)$, också kallad spektral funktion. Spektral funktion är avhändig av variabel ω , och kan blir definierad som:

$$F(\omega) = A(\omega) + jB(\omega) = \text{Re}\{F(\omega)\} + j\text{Im}\{F(\omega)\} = |F(\omega)|e^{j\varphi(\omega)}$$

där $|F(\omega)|$ har absolut värde $\varphi(\omega)$ och är fasens spektrum

I varje ekvation är j definierad som $j = \sqrt{-1}$. En exponentiell komplex och är transformeringens hjärta. Komplexa exponentiella funktioner är i grunden kompletta tal, där båda mappar, verkliga och imaginära är sinusoider. Den exakta relation kallas för Euler ekvation $e^{j\varphi} = \cos\varphi + j\sin\varphi$, vilken leder till intressant (och vacker) jämförelse $e^{j\pi} + 1 = 0$. Det är enklare att arbeta med ett exponentiellt komplex än med trigonometriska funktioner då detta ger mycket enklare sätt att skriva in sinusoid (det är enklare att skriva $e^{j\varphi}$ än skriva $\cos\varphi + j\sin\varphi$).



Komplexa exponentiala funktioner (såsom sinus och cosinus) är periodiska funktioner och uppsättning av komplexa exponentiala funktioner är komplett och ortogonal. Således kan Fourier transformation representera vilken partiell och associerad funktion som helst och fel mätts med minsta kvadrat metoden (least-square error) mellan funktionen och Fouriers representation, är minimalt.

Det existerar alltså en annan komplett och ortogonal uppsättning med periodiska funktioner; t.ex. Walsh funktion (kvadratiske vågor), vilken är fördelaktig för digitala elektroniska enheter (mera om detta i kapitlet Ortogonal transformationer).



Varför möter man alltid komplexa exponentialfunktioner vid lösningar av fysikaliska problem? Varför har monokromatiska vågor sinusformad karaktär och inte periodiska regelbundna periodiska sekvenser med kvadratiske eller triangulära vågor? Anledningen är att derivat av komplexa exponentiella funktioner är enkelt omräknade komplexa exponentialfunktioner. Med andra ord är komplexa exponentiella funktioner differentiella operatörer. Merparten av fysikaliska systemer lyder under differentiella ekvationer. När elektroniska analogfilter skall konvertera sinusformad våg till en annan sinusformad våg med samma frekvensen (men inte nödvändigtvis med samma amplitud och fas) är detta filtrerad kvadratisk våg inte längre kvadratisk våg. Denna egenskapen av komplexa exponentiella funktioner gör Fouriers transformation användbar alltifrån radiospridning och ändå upp till kvantmekanik.



Fouriers transformation gäller också för tvådimensionella signaler och är matematisk definierad som:

$$F(\omega_1, \omega_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) e^{-j\omega_1 x_1} e^{-j\omega_2 x_2} dx_1 dx_2$$

och omvänd transformation är definierad:

$$f(x_1, x_2) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\omega_1, \omega_2) e^{j\omega_1 x_1} e^{j\omega_2 x_2} d\omega_1 d\omega_2$$

2.2 Diskret Fouriers transformation

Diskret Fourier transformation **DFT** används för att få spektrum av diskreta signaler. Endast ett begränsat antal sinusoider behövs för att beskriva signalen i en frekvensdomän.

För att visa vad DFT gör använder vi följande exempel. MP3 spelare sänder till högtalare audioinformation i form av instabil spänning i elektriska signalen. Resultaten är att luftpartiklar förflyttar sig samt producerar ljud. Instabil audiosignal i tid kan blir grafisk illustrerad med hjälp av axel x som visar tid och axel y som visar spänningen i den elektriska signalen. Detta ser ut som oregelbunden och oberäkneligt vågliknande krumelur vilken i själva verket är summan av ett antal mer regelbundna krumelurer vilka representerar olika ljudfrekvenser. Med frekvensen menas den hastigheten som luftmolekyler går förflyttar sig fram och tillbaka när spänningen varierar.



DFT skapar matematiskt vad mänsklig örat gör i verkligheten: bryter ner en signal i enskilda frekvenskomponenter. Till skillnad från analog signal från en skivspelare är digital signal från MP3 spelare bara en sekvens av nummer vilka representerar mycket korta stickprover av den verkliga ljuden: t.ex. CD-kvalitet i digital ljudspelare samlar 44 100 prover per sekund. Om vi väljer några efter varandra gående värden från en digital signal: 8, 128 eller 1 000 – DFT representerar den ”viktigare” summan av samma antal frekvenser, („viktigare“ betyder bara att några frekvenser har större andel i helheten än andra).



Diskret Fouriers Transformation (DFT) uttrycks matematiskt som:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-n2\pi jk/N}$$

och Inversen Diskret Fouriers Transformation (IDFT) är definierad som:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{n2\pi jk/N}$$

där N är antal prover av diskret signal och $n = 0, 1, 2, \dots, N-1$, $e^{2\pi jk/N}$ är mycket ofta ersätt som Ω ($\Omega = e^{2\pi jk/N}$).

DFT sekvens av N -punkter från tidsserien, är N -punkter frekvensspektrum med Fouriers frekvens k som spänner alltifrån $-(N/2 - 1)$ genom noll frekvensen som kallas enkelriktad sekvens, och upp till den högsta Fouriers frekvens $N/2$. Varje spektral sekvens representerar heltal av sinusformad period i tidsserien. Amplituder och faser är representerade med amplituder A_k och faser φ_k av dessa är sinusformad. Sammanfattningsvis kan varje spektral sekvens bli beskriven som:

$$X(k) = A_k e^{j\varphi_k}$$

Diskret Fouriers transformation kan också definieras för tvådimensionella signaler och då representeras som en serie utvidgade bildfunktioner (över 2D utrymme).



Definition av framåt och omvänd **2D FT** är följande:

$$X(\omega_1, \omega_2) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} x(n_1, n_2) e^{-j\omega_1 n_1} e^{-j\omega_2 n_2}$$

$$x(n_1, n_2) = 1/(2\pi)^2 \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} X(\omega_1, \omega_2) e^{j\omega_1 n_1} e^{j\omega_2 n_2} d\omega_1 d\omega_2$$

Spektrum $X(\omega_1, \omega_2)$ är komplex när följande $x(n_1, n_2)$ är verklig. Spektrum kan definieras som en verklig summa och imaginärdel eller som produkten av magnituden och fasen.

$$X(\omega_1, \omega_2) = |X(\omega_1, \omega_2)| e^{j\theta_x(\omega_1, \omega_2)} = X_R(\omega_1, \omega_2) + jX_I(\omega_1, \omega_2)$$

2.3 Spektrum



Frekvensspektrum för en tids-domän-signal är representation av en signal i en frekvensiell domän. Frekvensspektrum kan genereras via Fouriers transformation (eller diskret Fouriers transformation) som signal och värden presenteras vanligen som amplitud och fas, båda ritade proportionellt till frekvensen.

Signalen kan blir representerad som amplitud som skiftar med tiden, och har ett motsvarande frekvensspektrum. Detta inkluderar kända begrepp som synlig ljus (färg), musiknoter, radio och TV-programkanaler och även regelbunden rotation av jorden. När dessa fysikaliska fenomen representeras i form av frekvensspektrum blir vissa bestämda fysikaliska beskrivningar av deras interna processer mycket enklare. Frekvensspektrum pekar ut harmoniska sekvenser, oftast som synliga spetsar eller linjer vilka ger tydlig insyn i de mekanismer, som genererar hela signalen.

Spektrumanalys är teknisk nedbrytningprocess av hela signalen i dess enklaste delar. Som beskrivs ovan många fysikaliska processer är bäst beskrivna som summa av många enskilda frekvenskomponenter. Processen som kvantifierar de olika begreppen (dvs. amplituder, styrka, intensitet eller fasen) som andelar i frekvensen kallas spektrumanalys.

Spektrumanalys kan utföras på hela signalen (vanligtvis periodisk signal). Alternativt kan en signal (främst icke-periodiska eller kvasi-periodiska) delas i korta segment så kallade rutor och spektrumanalys kan tillämpas/på dessa individuella delar.

Fouriers transformationsfunktion skapar frekvensspektrum, som innehåller all information om den ursprungliga signalen dock i en annan form. Detta innebär att den ursprungliga funktionen kan helt rekonstrueras med omvänd Fouriers transformation.

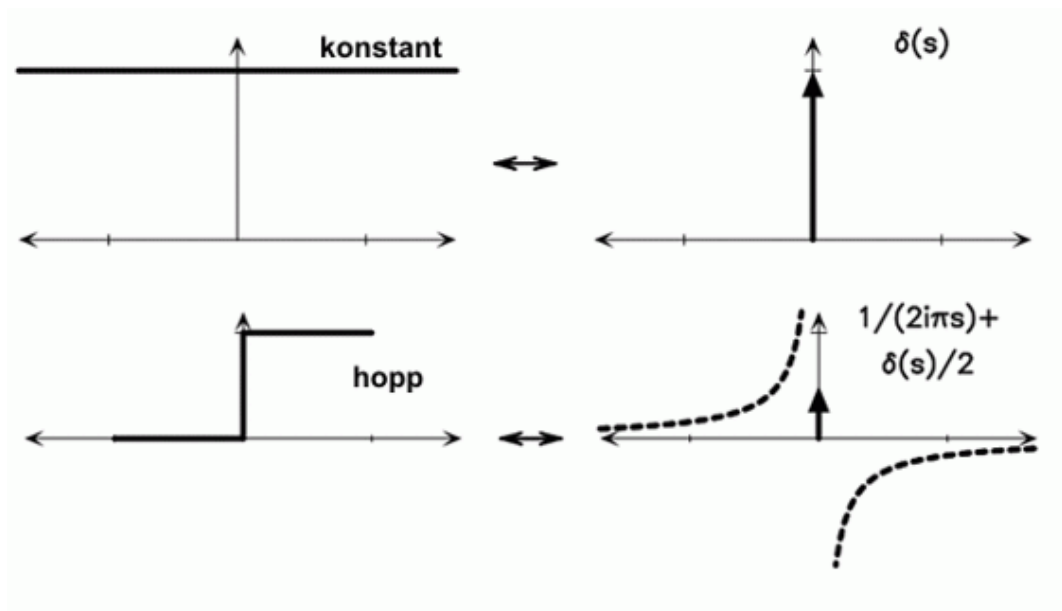


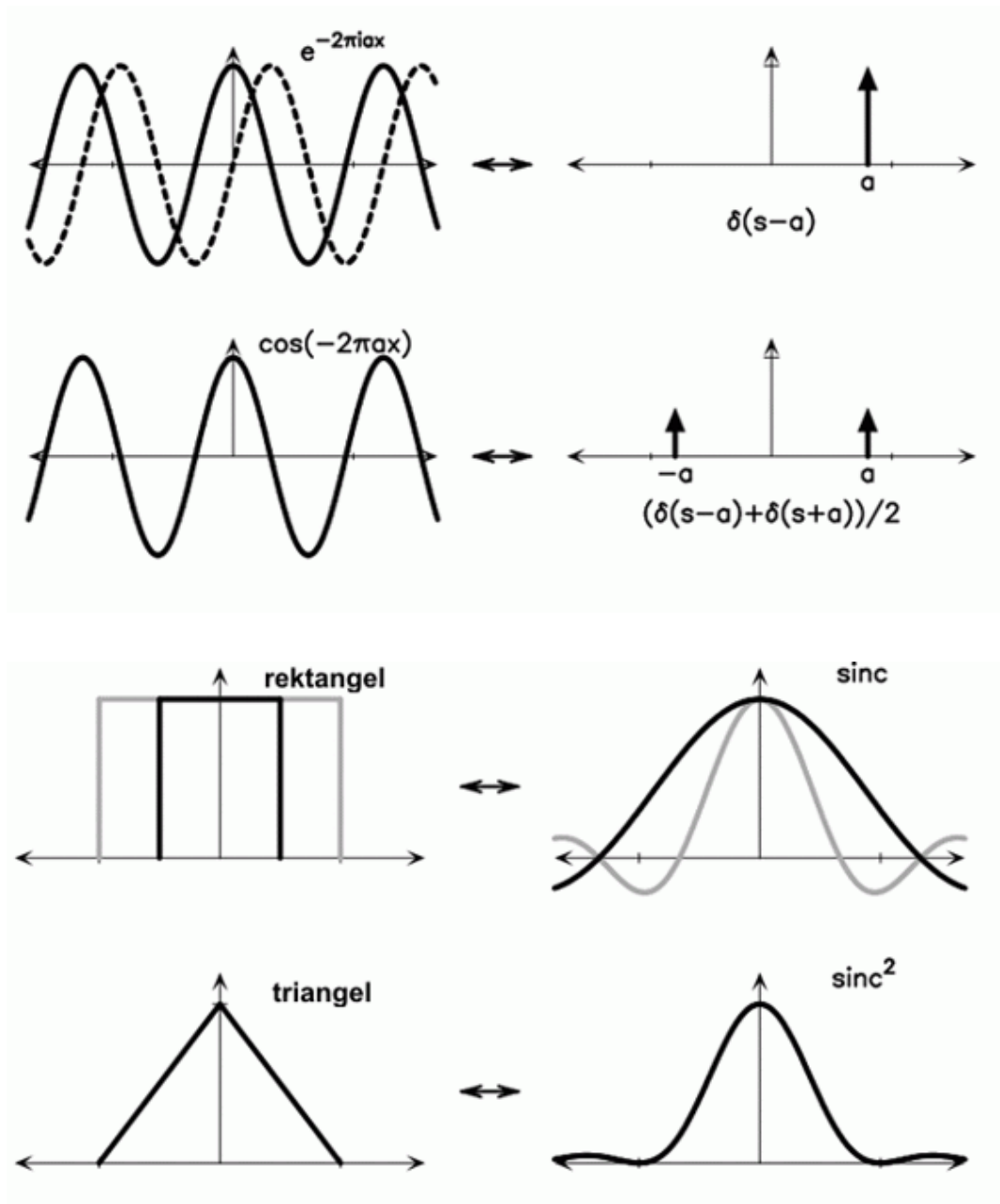
För perfekt rekonstruktion måste spektrumanalysator bibehålla både amplitud och fasen för varje frekvenskomponent. Dessa två informationdelar kan blir representerade som tvådimensionell vektor som komplex tal eller som amplitud och fasen i polära koordinater. Vanlig teknik på området för signalbearbetning måste ta hänsyn till amplitudkvadraten eller styrka. I detta fall kallas det resulterande diagrammet för ett kraftspektrum.

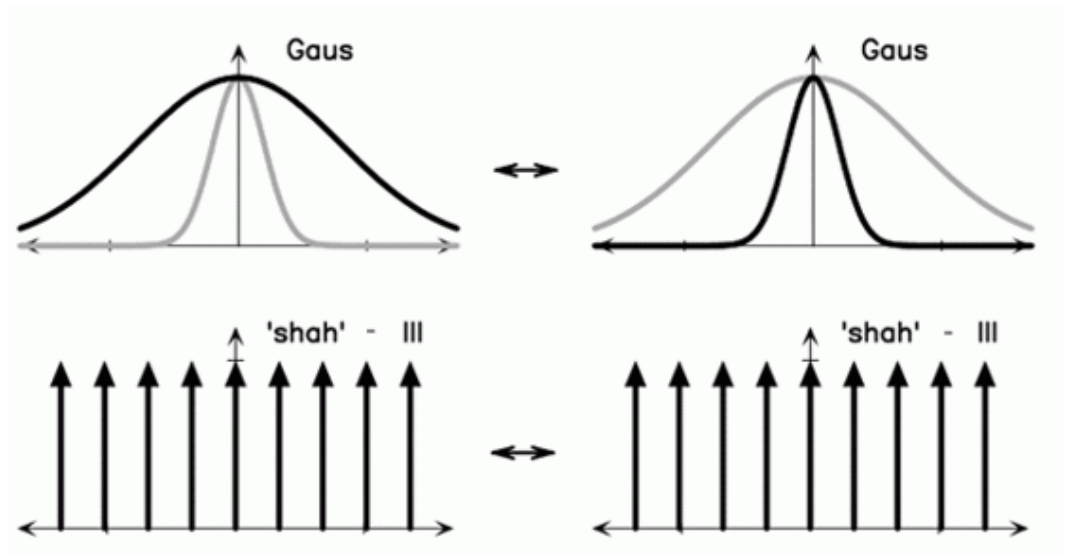
I följande tabell sammanfattas typer av signaler och deras spektrum.

Signal	Spektrum
kontinuerlig periodisk	diskret icke-periodisk
kontinuerlig icke-periodisk	kontinuerlig icke-periodisk
diskret periodisk	diskret periodisk
diskret icke-periodisk	kontinuerlig periodisk

Följande bild visar olika grundläggande par för Fouriers transformation (på bild visas bara amplituder). Dessa kan bli kombinerade med användning av den angivna teorin om Fourier transformation för att generera Fourier transformation till olika funktioner.

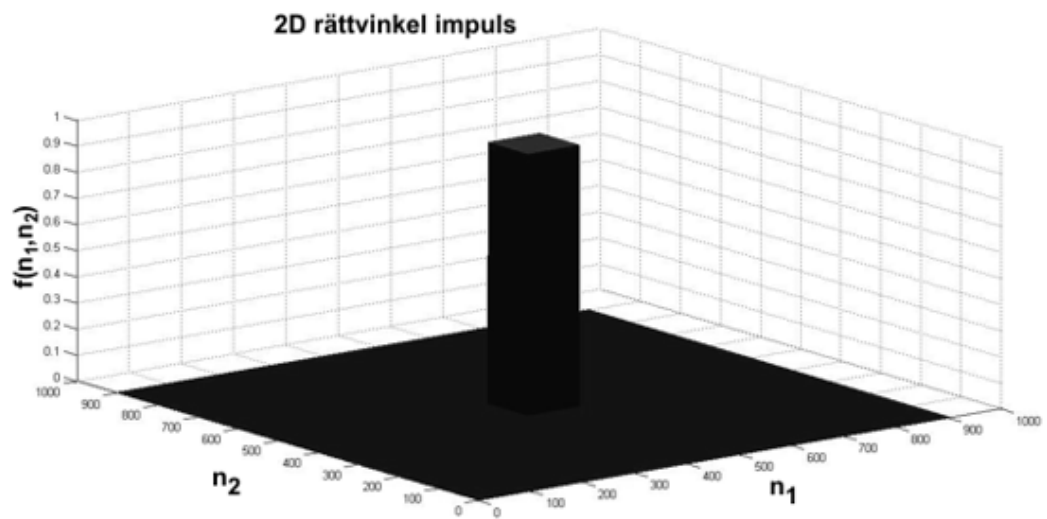




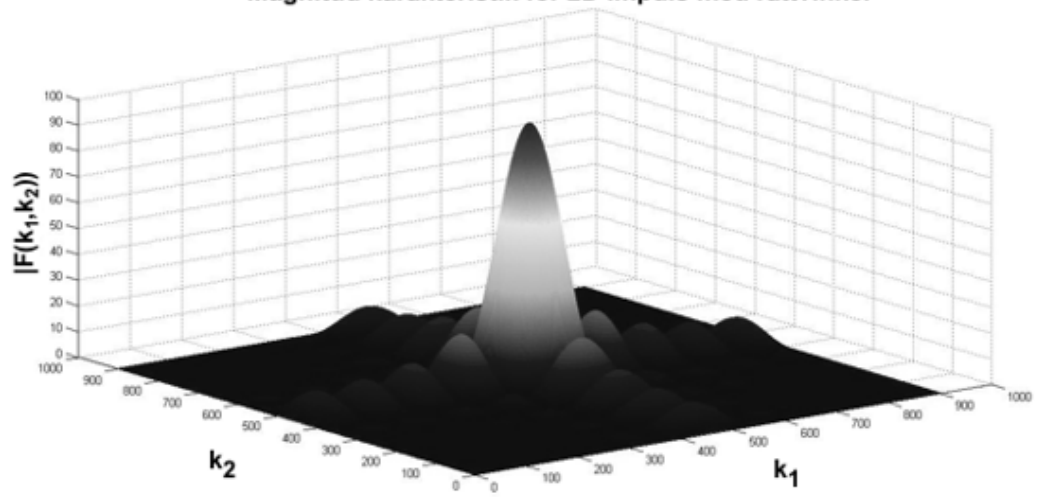


Grundläggande par för Fourier transformation

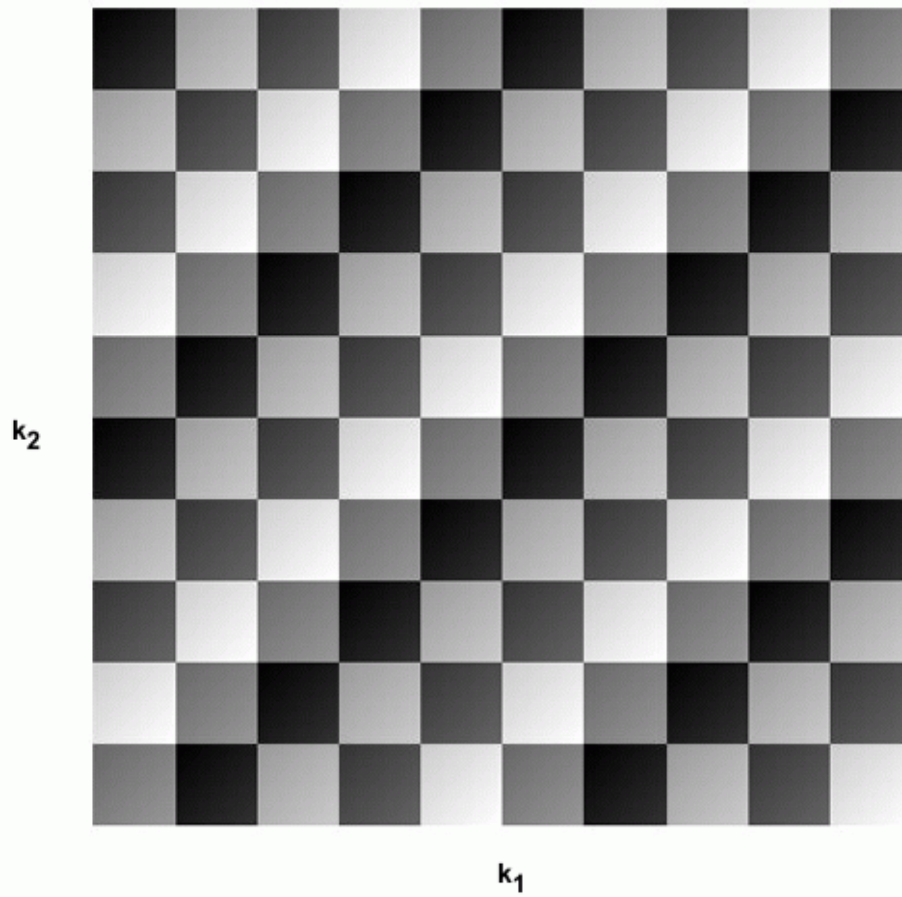
Bild nedan visar tvådimensionella signaler och deras spektrum (med karaktäristisk magnitud och fasfrekvens).



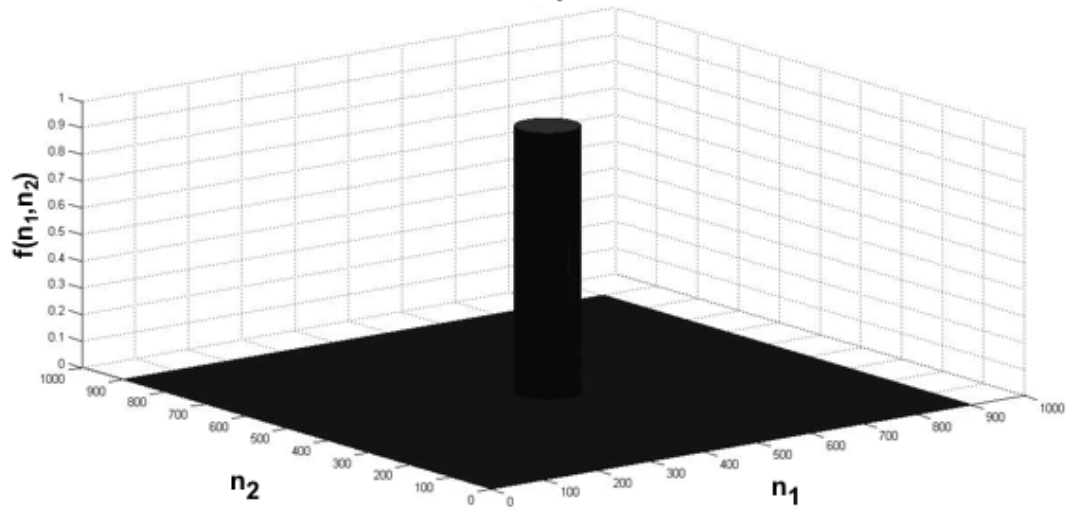
Magnitud karakteristik för 2D impuls med rättvinkel



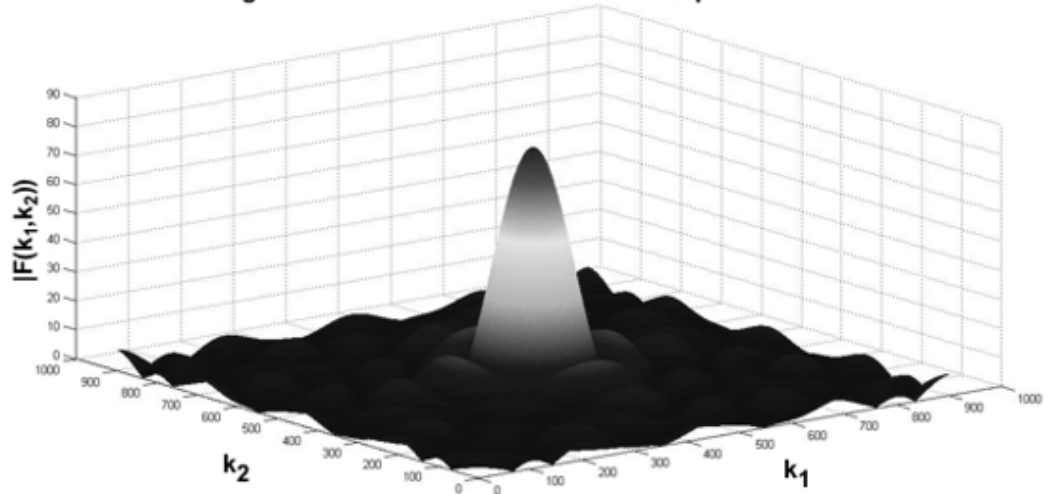
Fasens karakteristik för 2D impuls med rättvinkel



2D cirkulär impuls med rättvinkel

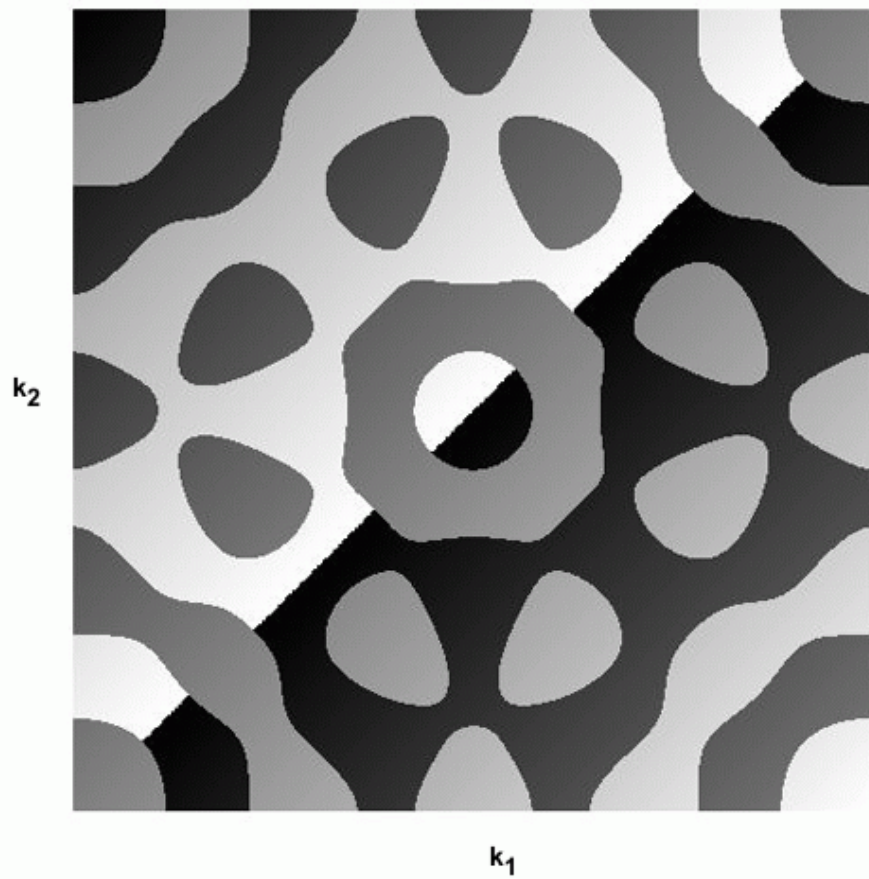


Magnitud karakteristik för 2D cirkulär impuls med rättvinkel



2D kvadratisk funktion, med kararkteristisk magnitud och fasfrekvens karakteristik

Fasens karakteristik för 2D cirkulär impuls med rättvinkel



2D cirkulär funktion, med karakteristisk magnitud och fasfrekvens

2.4 Ortogonal transformation



Ortogonal transformationer tillåter representera tidsförloppet för signalen i form av ett allmänt spektrum. Transformation används för att efteråt utföra i spektral domän behövande matematiska operationer vilka möjliggör t.ex. snabbare utvärdering av signaler eller minskning redundans.

Diskret ortogonal transformation finner brett användning på många domäner: vid komprimering av data, igenkänning av bilder, vid analys och syntes av röstsignaler osv. Huvudmålen i denna kurs är beskrivning av endimensionella ortogonala funktioner och transformationer.

Det enklaste sättet att matematiskt uttrycka endimensionella signaler är linjär kombination (summan av multiplar) av några grundläggande basfunktioner. Sådant uttryck för endimensionella signaler är mycket fördelaktig i synnerhet i linjära system därför att den tillåter hantering av många uppgifter enligt principen för superposition. Det är önskvärt att grundfunktioner $u(k, t)$ beräknas lätt samt har enkel form och att dessa krav överensstämmer med en ortogonal funktion.



Matematisk definition för ortogonala funktioner $u(0, t), u(1, t), \dots, u(N-1, t)$ i intervall $\langle t_1, t_2 \rangle$ är följande:

$$\int_{t_1}^{t_2} u(k, t) u(m, t) dt = 0, \quad k \neq m$$

$$\int_{t_1}^{t_2} u^2(k, t) dt = U_k, \quad k \neq m$$

I exempel, att $U_k = 1$; kallas funktionen ortonormal (dvs. har storlek en).

T. ex. att matematisk uttrycka endimensionell signal $x(t)$, med hjälp av linjär kombination med några basfunktioner:

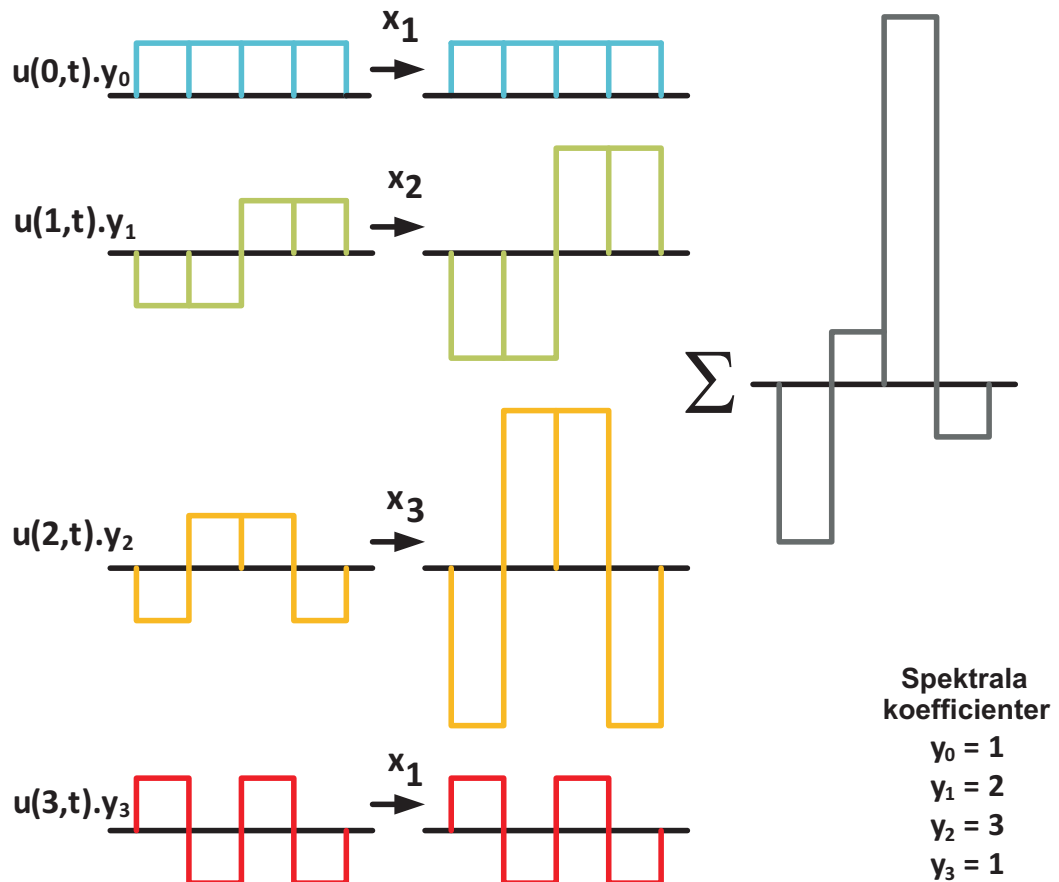
$$x(t) \cong \sum_{k=0}^{N-1} y_k u(k, t)$$

där y_k är spektrala koefficienter definierade som:

$$y_k = \frac{1}{U_k} \int_{t_1}^{t_2} x(t) u(k, t) dt$$

Studera följande exempel:

System för de första fyra Walsh funktioner



Exempel för matematisk uttryck för endimensionell signal, med hjälp av linjär kombination med några basfunktioner

Walshs funktion likaså som Haars- och Rademachersfunktion tillhör de mest använda ortogonala funktionerna på domäner för bearbetning av signaler.



Approximation av den diskreta signalen $x(nT)$, där T är en skapande period med antal prover M med några diskret ortogonallabasegenskaper ges av formeln:

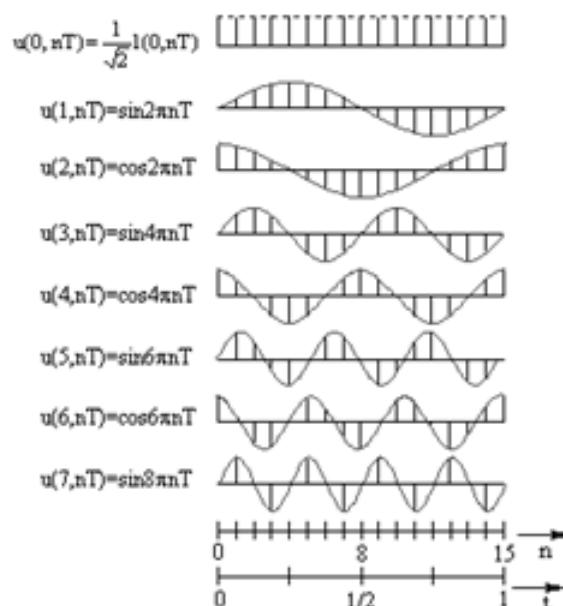
$$x(nT) = \sum_{k=0}^{N-1} y_k u(k, nT) \quad n = 0, 1, \dots, M-1$$

Medan de optimala koefficienterna är:

$$y_k = \sum_{n=0}^{M-1} x(nT) u(k, nT)$$

Vid harmoniska signaler är den viktigaste parametern en frekvens. Vid icke-harmoniska signaler är den viktigaste parametern en sekvens vilken anges med antalet skärningspunkter i nollnivå per sekund. I samband med diskreta funktioner, fastställs sekvensen som ett antal teckenändringar i stickprover per sekund.

Diskreta ortogonala basfunktioner kan skapas med diskreta funktioner associerade med ortogonalt baserade harmoniska funktioner.



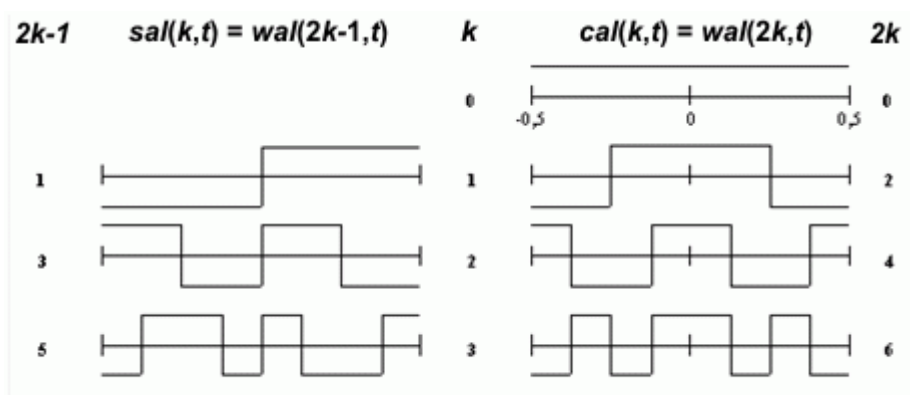
Walshs basfunktioner



Walshs funktioner är en ordnat uppsättning av rektangulära pulser, som har endast två möjliga värden för amplitud (+1 och/eller -1) och utgör en komplett system för ortogonalfunktioner. Dessa är beroende av de två faktorerna tiden (t) och ordningstalet (k). De är markerade $wal(k, t)$ och man kan skilja mellan jämna (kosinus-Walsh) $cal(k, t)$ - och udda (sinus-Walsh) $sal(k, t)$ funktioner, vilka görs:

$$cal(k, t) = wal(2k, t), \quad k = 0, 1, 2, \dots$$

$$sal(k, t) = wal(2k-1, t), \quad k = 0, 1, 2, \dots$$



Exempel på jämna och udda Walshs basfunktioner

I enlighet med Walshs funktioner kan dessa delas i tre grupper:

1. arrangemang av Walsh (sekventiell) $walw(k, t)$
2. arrangemang av Paleyho (diadyc – didactic) $walp(k, t)$
3. arrangemang Hadamar (naturliga) $walh(k, t)$

Alla funktioner har egna arrangemang för genomförande. I grunden innehåller de samma funktioner men är sorterade på olika sätt. Som exempel kan du lista Hadamars (naturliga) arrangemang.

+

Fördelen med detta arrangemang är att på ett enkelt sätt skapa grundfunktioner av högre dimensioner.

I bilden nedan är den första av de åtta kontinuerliga och diskreta Walshs funktioner med en naturlig ordning. Om vi skriver Walshs diskreta matrisfunktioner skapar i detta fall Hadamars matris $U_h(3)$ med en storlek på 8 x 8.



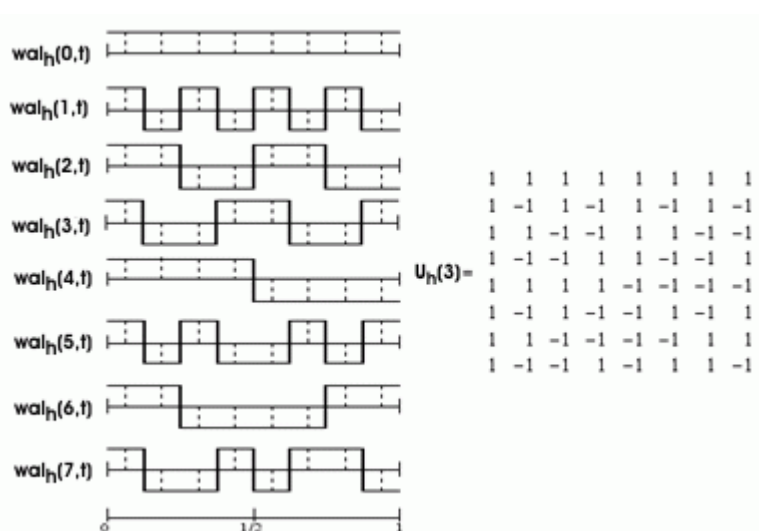
I allmänhet kan man beräkna Hadamars matrix $U_h(r)$ med dimension $M \times M$, där $M = 2^r$ med hjälp av Kroneckers direkta produkten av matriser från $U_h(r-1)$.

$$U_h(r) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \otimes U_h(r-1) = \begin{bmatrix} U_h(r-1) & U_h(r-1) \\ U_h(r-1) & -U_h(r-1) \end{bmatrix}$$

där $U_h(0)=1$



Kroneckers produkt betecknas som $\otimes$ en operation med två matriser av godtyckliga dimensioner. Denna operation resulterar i en block-matris.



En naturlig placering av kontinuerliga Walshs diskreta funktioner och en matris gjord av dessa funktioners värde.

Det finns också andra transformationer med harmoniska grundfunktioner. Förutom Fouriers diskreta transformation finns det *discrete cosines transformation* (**DCT**), *discrete sinus transformation* (**DST**), *discrete Hartley transformation* (**DHYT**).

3 Analog- och digitalteknik

3.1 Multiplexing

Multiplex kombinerar flera analoga eller digitala signaler i en telekommunikationskanal. Multiplex (mux) används främst till att öka datamängder som skickas över ett nätverk med en viss bandbredd och vid en viss tidpunkt.

Och tvärtom, demultiplex (demux) är en enhet som tar emot en enda inputsignal och delar denna till många output-signaler. Multiplexer används oftast med kompletterande demultiplexer på mottagarsidan.

Multiplex frekvenskrets

Multiplex frekvenskrets (*Frequency division multiplexing FDM*) är en analog teknologi. Multiplex frekvenskrets kombinerar flera digitala signaler till gemensamma resurser (datapaket) för att sända dessa signaler i olika frekvens domäner.

+

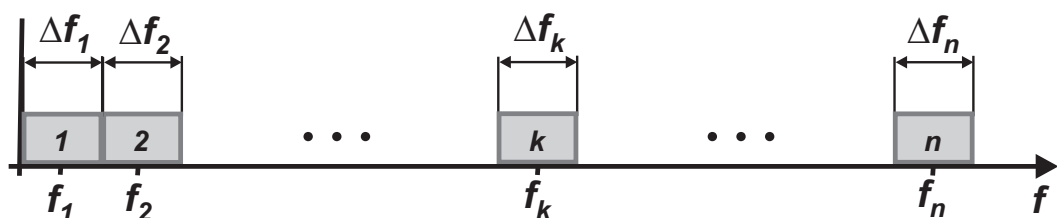
Anledning till användningen av system med frekvensmultiplex är förmågan att sända samtidig flera signaler från olika källor samt att dessa signaler kan sändas från en mängd olika avlägsna områden.

—

Nackdelen är ömsesidiga kanalstörningar som skapas av överlappande signalspektra. Detta sker pga icke-idealiska filter samt uppkomsten av oönskade frekvensmappar som åstadkoms av icke-linjära elektriska kretsar (även kallade överhörning).



Frekvens multiplex åstadkoms genom att till varje individuell signal, på en frekvensaxeln tilldelas ömsesidigt icke-överlappande frekventiella band $\Delta f_1, f_2, \dots, f_n(\Delta)$. Vanligen $\Delta f_1 = \Delta f_2 \dots = \Delta f_k = \Delta f_k$. Signalspektrum av behöriga kanaler måste passa in i det reserverade frekvensbandet Δf_n .



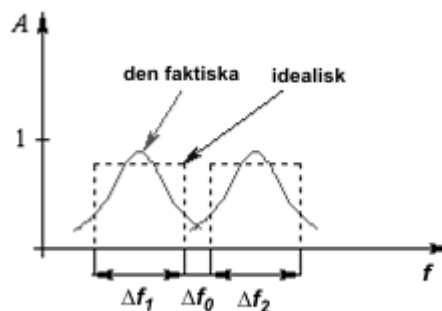
Flerkanals frekvensspektrum

Frekvensdomän för överförings väg F bestämmer gränshänsyn för kanaler eller minimal frekvensintervall Δf_l ($\Delta f_{l_{min}}$) samt maximal frekvens intervall Δf_n ($\Delta f_{n_{max}}$).

$$F = \Delta f_{n_{max}} - \Delta f_{l_{min}}$$

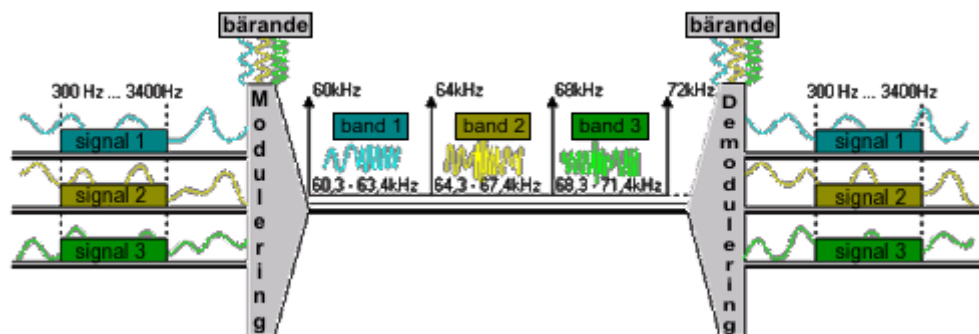
En viktig del av frekvensmultiplexsystem för båda sändare och mottagare är ett bandfilter (PP) som på sändarsidan avgränsar frekvensband för relevanta kanaler. På mottagarsidan delar filtret upp signaler som efter att ha passerat genom demodulator kan accepteras för införsel av införselomvandlare. Vid val av frekvensbandfilter tas i synnerhet hänsyn till den begärda överföringshastigheten för systemet.

I riktiga bandfilter sker dämpningen utanför gränser för frekvensdomän inte med hopp, men fortlöpande, och därför utelämnas mellan frekvensband för enskilda filter frekvensskyddsdomän Δf_0 . Förhållandet mellan $\Delta f_0/\Delta f = k_f$ beror på vilka amplitudfrekvens egenskaper används i bandfiltret (se bilden nedan). Mer information om filter kommer i avsnittet Filter.



Amplitud-frekvens kännetecken (idealisk och faktisk) för bandfiltrar

Genom modulering av olika bärsignaler flyttas enskilda (t.ex. tal) domäner till subdomäner, och på detta sätt överförs de vidare. Efter demodulering på mottagarsidan finns dessa enskilda talband sedan tillgängliga i sin ursprungliga talform. Med denna procedur kan man överföra flera talband via en enkel kanal samtidigt.



Princip för frekvensmultiplex

Frekvensmultiplex användes framförallt i analoga telekommunikationssystem. För närvarande används detta främst av TV-bolag (varje kanal har en något annorlunda frekvens, med kanalbyte växlar man till ett annat frekvensband) och i optisk kommunikation.

Multiplex tidskrets

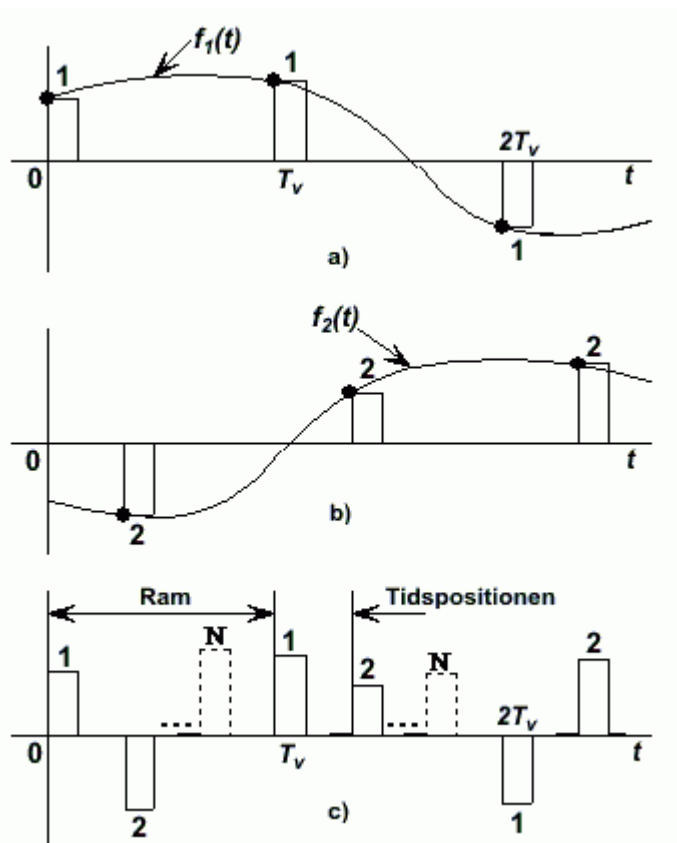


Multiplex tidskrets (*Time division multiplex* **TDM**) används både i nätverk och telefon och gör exakt vad den säger i titeln: data samlas tidsregelbundet från flera signaler med lägre hastighet för att sedan överföras förenade via en överföringskanal med en högre hastighet. Därefter återställs de till de ursprungliga signalerna.

Inmatningsenheten även kallad multiplex insamlar data från olika signaler för att sedan kombinera dessa till ett enda paket som sänds via en kanal. På detta sätt är flera data från olika källor ihopkopplade och överförda av en höghastighetskanal. Detta är möjligt tack vare det faktum att de enskilda källorna sänder data med relativt låg moduleringshastighet (som 300 Baud) och den bärandekanalen överför moduleringshastigheten från alla källor till (1200 Baud). I slutet av överföringen delar en annan multiplex det överförda datapaketet för att vidare befodra det till destinationen med samma låga moduleringshastighet med vilken data förts in i TDM-systemet.

Denna teknik används i främst av telefoniföretag som måste överföra stora mängder av telefonsamtal med hjälp av ett begränsat antal kablar. I händelse av att samtalet bryts delas data upp för att återhämta sig snabbare än det mänskliga örat kan upptäcka och ingen kommer att upptäcka avbrottet. Det är varför telefontrafik använder TDM med höghastighetsbärare där flera samtal överförs på en gång.

- Nackdelen med multiplex som utnyttjar tidsmultiplexing är det faktum att den tilldelar en tidsposition även när källan inte har något att sända. Detta leder till ineffektivitet.



Tidsmultiplex a) information 1, b) information 2, c) multiplex impulssekvens

Synkronisering

Synkronisering av provtagningsprocessen är nödvändig för att upprätthålla rätt position i impulssekvensen. Detta säkrar sekvens för klockimpulser som är referens för alla provtagningkretsar. Mottagaren måste ha liknande sekvens av klockimpulser för att säkerställa korrekt mottagning av enskilda impulser med relevant information. Annars kan mottagares demultiplex inte avgöra till vilken destinationkanal tillhör de enskilda källsignalerna och hur den ska rekonstruera en signal. Sekvens av klockimpulser i sändaren och mottagaren måste ha en definierad nyckeltalfas och synkroniseras med varandra. Klocksynkronisering är därför nödvändig för en korrekt fungerande TDM-teknologi.



Synkronisering av klockimpulser i sig garanterar fortfarande inte korrekt mottagning av information. Det är också nödvändigt att göra en avgränsad bildsynkronisering. För synkronisering av klockimpulser och för avgränsad bildsynkronisering bestäms vanligtvis också en tidsposition i vilken speciella impulser placeras t.ex. större än det beräknade maximala värdet i följande tidsposition. Början av varje bildavgränsning är påvisbar med unika enkla tröskelkretsar.

Tidssynkroniseringen baseras på sändning av referenstidsimpulser från en sändare. Mottagaren måste synkronisera sig själv med hjälp av de mottagna impulserna. Det finns också en bildsynkronisering där varje bildruta har en särställning då en bild med en speciell karaktär utses.

3.2 Linjära tidsdiskreta och invarianta system

Linear time-invariant system (LTI) Linjära tidsinvarianta system används direkt i seismologi, signalbearbetning, kretsar, teorin om reglerteknik och andra tekniska domäner. Analys av kontinuerliga LTI och diskreta LTI liknar varandra en hel del. Diskret domän är mindre tekniskt utmanande och därför fokuserar vi oss på dessa system.



LDTI (Linear discrete time-invariant system) Linjära diskreta, tidsinvarianta system har en signal på inkommande data och en på utdata och för den anges följande egenskaper:

- Systemet är linjär. Detta innebär att både ingång $x_1(n)$ och $x_2(n)$ genererar utdata $y_1(n)$ och $y_2(n)$, och med detta är a_1 och a_2 konstanter medan insignalen $a_1x_1(n) + a_2x_2(n)$ genererar en utsignal $a_1y_1(n) + a_2y_2(n)$.
- Systemet är oföränderlig i tid. Detta innebär att när ingångssignalen $x(n)$ genererar utdatasignal $y(n)$ ges sedan för varje reellt tal s att offset för insignalen i tiden $\hat{x}(t) = x(t-s)$ genererar offset för utsignalen i tiden $\hat{y}(t) = y(t-s)$.

Differentierad ekvation

Första viktiga kunskap om uppförandet av en diskret linjär slutlig tidsinvariantsystem är att systemets svar på input är exakt definierad med systemets svar på en specifik input i tid 0, och detta är Kronecker impuls (se definitionen i avsnittet Viktiga signaler). Systemets svar kallas för impulssvar. Vi definierar nu en relation som uttrycker utdata som genereras av insignal $x(n)$.



$$y(n) = \sum_{k=0}^N a_k x(n-k) - \sum_{k=1}^N b_k y(n-k)$$

Ekvation beskriver en rekursiv LDTI (med feedback). Stickprover för denna utsignal ges som en linjär kombination av kalibrerade prover av insignalen. Viktade koefficienter är betecknade som a_k och b_k . System som beskrivs av följande ekvation kallas system med oändlig impulssvar (*infinite impulse response IIR*).

Ett speciellt fall är den differentiella ekvationen för icke-rekursiv system. I detta fall är utsignalen endast beroende av insignalen, och ej av tidigare prover från utsignalen. Systemet är slutligt impulssvar (*finite impulse response FIR*) och matematiskt definieras som:



$$y(n) = \sum_{k=0}^N a_k x(n-k)$$

Konvolution



Konvolution är ett annat sätt att beskriva systemet LDTI. Om man vet att systemets impulssvar kallas $h(n)$ och insignalen ges som $x(n)$, då kan utsignalen kan uttryckas som:

$$y(n) = \sum_{k=0}^{D_y-1} x(k)h(n-k) = x(n) * h(n)$$

där operatör* är produkt av konvolutionen. Längden på utsignalen definieras med förhållandet $D_y = D_x + D_h - 1$ där D_x är längden på insignalen och D_h är längden på impulssvaret.

Verksamheten bygger på principen om överlagringen av linjära system. Systemets outputsignal ges av summan av viktade och flyttade impulsegenskaper. Granska följande exempel för att förstå:



Låt oss ha ett FIR-system med impulssvar $h(n) = \{1, 2, 3\}$. Insignalen är fastställd som:

$$x(n) = \{x(0), x(1), x(2), x(3)\}.$$

Utsignalen bestäms med hjälp av konvolutionen.

Längden på impulssvar $h(n)$ är $D_h = 3$ och längden på insignalen $x(n)$ är $D_x = 4$, då längden för output är $D_y = 6$. Vi kan helt enkelt beräkna konvolution med hjälp av tabellen:

Konvolution

n	0	1	2	3	4	5	6
x(0)	x(0)	2 x(0)	3 x(0)	0	0	0	0
x(1)		x(1)	2 x(1)	3 x(1)	0	0	0
x(2)			x(2)	2 x(2)	3 x(2)	0	0
x(3)				x(3)	2 x(3)	3x(3)	0
y(n)	y(0)	y(1)	y(2)	y(3)	y(4)	y(5)	0

I varje tabellrad är behöriga stickprover på insignalen viktade med impulsegenskaper. Genom att förskjuta raderna i högerriktning får man fram fördröjningen av insignalen från tillhörande provet. I den sista raden är stickprover från utsignalen vilka vi får från deras överlagring i varje kolumn för $n = 0, 1, 2, \dots$ Till exempel:

$$y(1) = 2 \cdot x(0) + x(1)$$

$$y(2) = 3 \cdot x(0) + 2 \cdot x(1) + x(2), \text{ osv.}$$

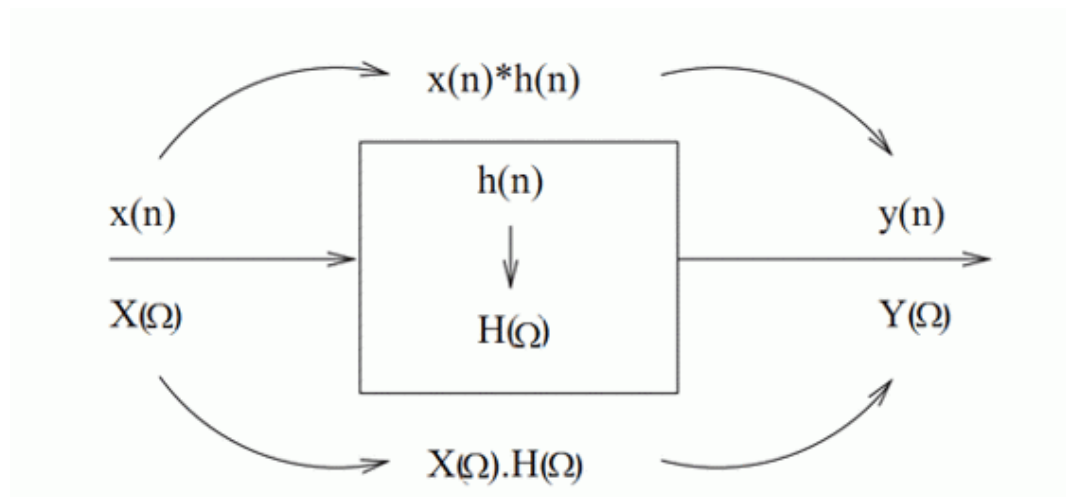
Produkten av konvolution av de två sekvenserna beräknas så, att den ena av sekvenserna kommer att organiseras i omvänd ordning och sedan flyttas den under den andra från höger, och för varje steg bestäms summan av de framställda produkterna.

Överföringsfunktion

Överföringsfunktion $H(\omega)$ representerar förhållandet mellan ingående och utgåendesignal från LDTI-system med status noll vid startvillkor i frekvensdomän. Överföringsfunktion härstammar från differentiella ekvationer eller impulsegenskaper. I båda fallen används DFT (diskret Fouriers transformation) för transformation till frekvensdomän. Enligt definitionen ovan $y(n) = h(n) * x(n)$ får man med hjälp av DFT förhållandet:



$$Y(\omega) = H(\omega) \cdot X(\omega) \Rightarrow H(\omega) = \frac{Y(\omega)}{X(\omega)}$$



Sambandet mellan input- och outputsignalen och deras bilder

Det är uppenbart att $h(n)$ och $H(\omega)$ kännetecknas av samma system i olika domäner. Om vi tillämpar DFT på den differentiella ekvationen beskrivs överföringsfunktion med hjälp av viktade koefficient a_k och b_k , som är identiska med koefficienterna i den differentiella ekvationen:

$$H(\omega) = \frac{a_0 + a_1 e^{-j\omega} + a_2 e^{-j2\omega} + \dots + a_N e^{-jN\omega}}{1 + b_1 e^{-j\omega} + b_2 e^{-j2\omega} + \dots + b_M e^{-jM\omega}}$$

Förhållandet är registrerat som rationell vinkelbruten funktion då genom att dividera täljare med nämnaren får man stickprover direkt från impulssvar. För IIR-system är antal prover oändlig. I FIR-system är överföringsfunktionen inte i form av en bråkdel, eftersom nämnaren är lika med 1.

$$H(\Omega) = a_0 + a_1 e^{-j\Omega} + a_2 e^{-j2\Omega} + \dots + a_N e^{-jN\Omega}$$

Överföringsfunktionen är mycket viktig och med hjälp av $H(\Omega)$ denna bestäms frekvensgenskaper. Dessa används främst i teorin om filter.

Frekvensomfång

Frekvensomfång bestäms av dynamiken i systemet.



Det är förhållande mellan amplitud (eller magnitud) och fasensutdata som man får frekvensfunktion för ingång.

Enkelt uttryckt när man till systemets input infört sinusfunktionen med ett visst frekvensomfång svarar LDTI med en frekvens där amplituden och fasen skall ha värde i förhållande till indata.

Frekvensomfång för överföringsfunktionen $H(\Omega)$ definieras som:



$$H(\Omega) = |H(\Omega)| \cdot e^{j\varphi(\Omega)} = \text{Re}\{H(\Omega)\} + j\text{Im}\{H(\Omega)\}$$

$$\text{där } |H(\Omega)| = \sqrt{\text{Re}\{H(\Omega)\}^2 + \text{Im}\{H(\Omega)\}^2} \text{ a } \varphi(\Omega) = \arctan \frac{\text{Im}\{H(\Omega)\}}{\text{Re}\{H(\Omega)\}}.$$

Det absoluta värdet av överföringsfunktion $|H(\Omega)|$ kallas fasens frekvensomfångsmagnitud $\varphi(\Omega)$. Funktionen $\varphi(\Omega)$ är inte kontinuerlig funktion, men resulterar i en 180 graders hopp. När vi tar bort detta hopp får vi fasens frekvensomfång $\Theta(\Omega)$ som kommer att vara kontinuerligt. Att ta bort detta hopp ger möjlighet att ändra tecken för frekvensomfångsmagnitud som sker alltid vid varje hopp i fasens frekvensomfång. Förhållandet mellan amplitudens och magnitudens frekvensomfång är:

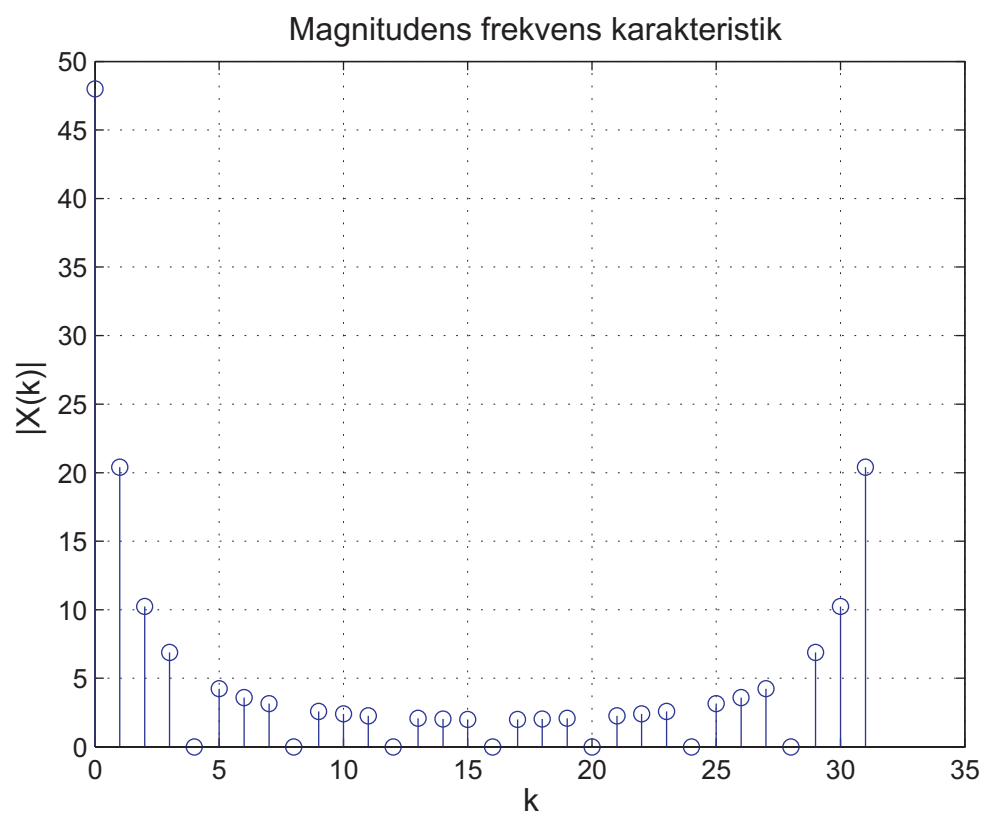
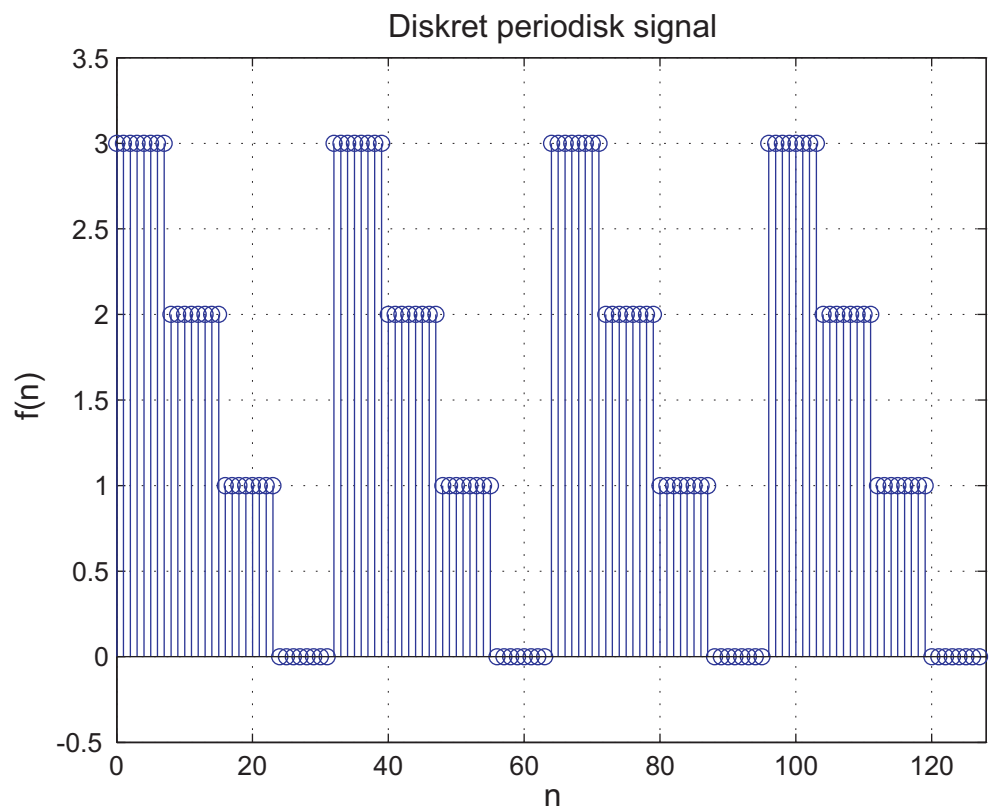
$$A(\Omega) = \pm M(\Omega)$$

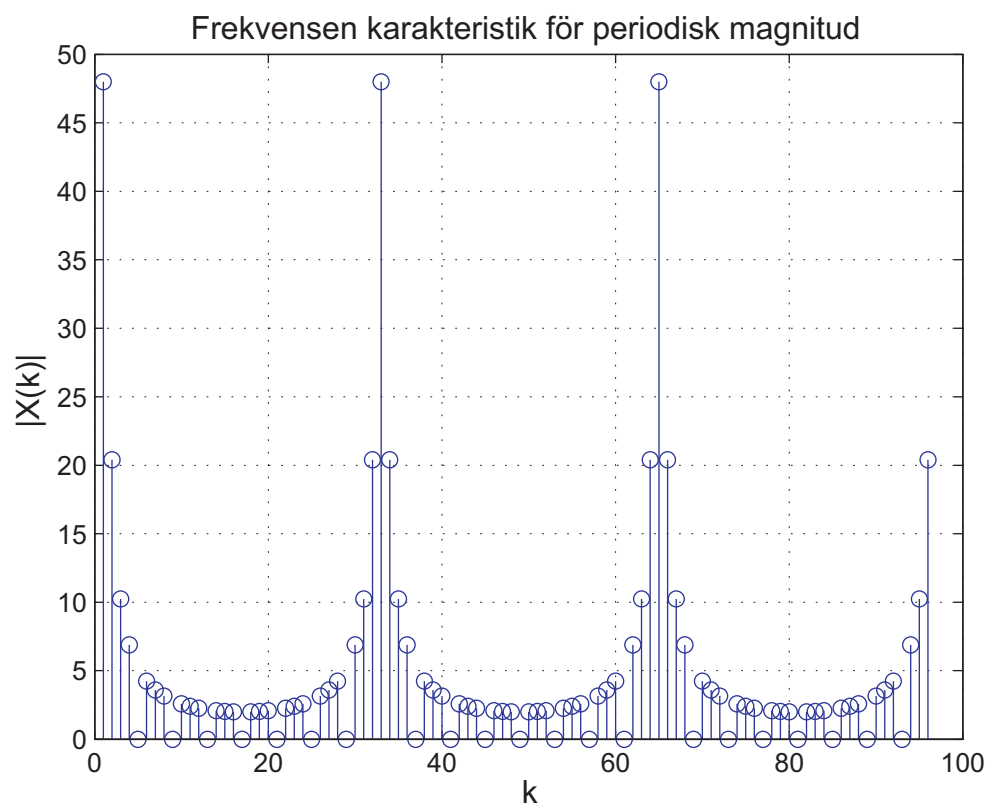
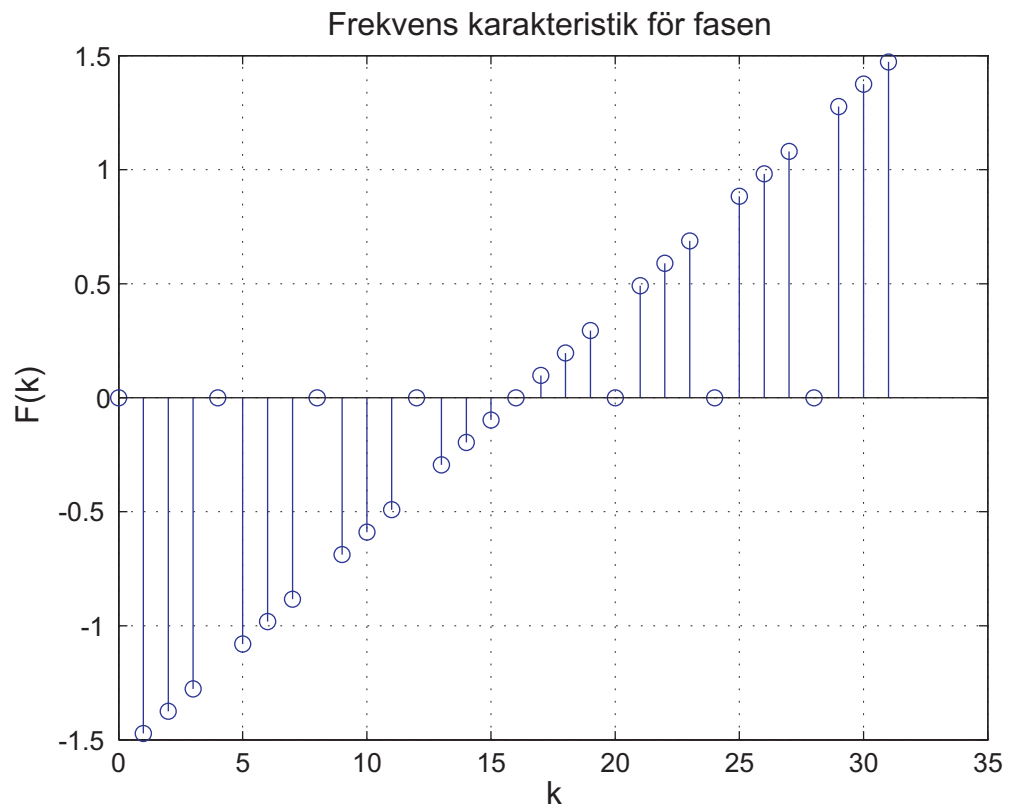
Med beaktande av ovanstående information kan följande definieras för frekvensens egenskaper:

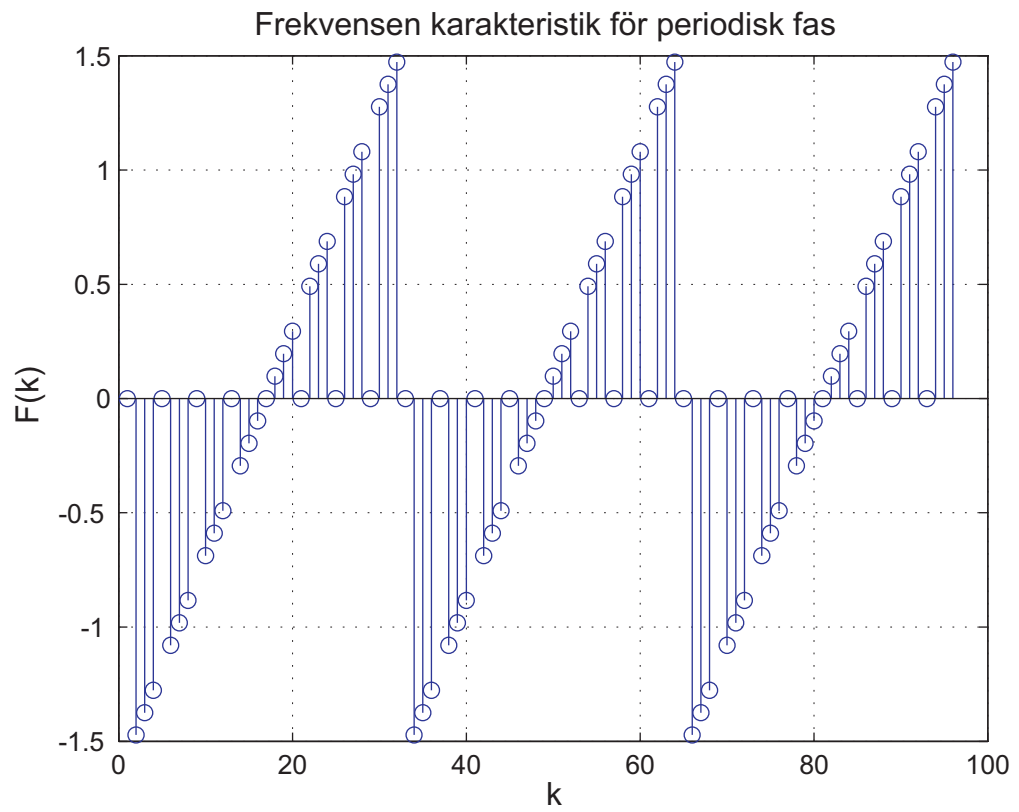


$$H(\Omega) = A(\Omega) \cdot e^{j\varphi(\Omega)}$$

$$H(\Omega) = M(\Omega) \cdot e^{j\varphi(\Omega)}$$







Ett exempel på LDTI-överföringssystem magnitudens och fasens frekvensegenskaper (under en och/eller flera perioder)

3.3 Filter



I signalbehandlingen är filter en enhet eller en process som tar bort en oönskad del eller egenskap från signalen.

Detta innebär att filter oftast tar bort vissa frekvenser för att dämpa signalstörningar samt minska bakgrundsbruset. Dock fungerar filter inte uteslutande på frekvensdomän. Särskilt inom domän för bildbehandling är det många olika variabler som kan filtreras.

Nackdelen med filtrering är förlust av information i samband med redan filtrerad del. Signal kombination i Fouriers utrymme är ett alternativt sätt för borttagning av specifika frekvenser från signalen.



Det finns flera olika klassificeringar av filter, som i många kriterier överlappar varandra. Det finns alltså inte en tydlig uppdelning. Därför kan vi dela filter i:

- analoga eller digitala
- tidsdiskreta eller värdediskreta
- linjära eller icke-linjära
- med oändliga impulssvar (IIR) eller med finita impuls svar (FIR), dessa med sin utformning befinner sig mellan diskreta och digitala filter

Viktiga begrepp som används i klassificeringen av linjära filter:

- Lågpasfilter – låga frekvenser går igenom, höga frekvenser dämpas.
- Högpasfilter – höga frekvenser går igenom, låga frekvenser dämpas.
- Domän för passfilter – bara frekvenser från definierad domän går igenom.
- Domän för stoppfilter – dämpar frekvenser från en definierad frekvens domän.
- „Ta bort“-filter – tar bort endast en definierad frekvens (detta är en speciell typ av domän stoppfilter).

Filter kan vara inbyggda i olika tekniska utföranden. Samma överföringsfunktionen kan genomföras på flera sätt. Detta innebär att de matematiska egenskaper och parametrar är desamma men det fysiska genomförandet är olika. Komponenter i varje teknik är oftast desamma och dessa utför i de använda filterna samma uppgift.

Elektroniska filter gjordes ursprungligen endast från passiva element (motstånd, induktion och kondensatorer). Aktiv teknik förenklade produktion och öppnade upp nya möjligheter i filterspecifikationerna.

- Digitala filter arbetar med signaler definierade i digital form. Kärnan i digitalt filter är att de har direkt i mikrokoden inbyggd matematisk algoritm som motsvarar önskad överföringsfunktion.

Linjäranaloga filter



Tidsbeständig linjärkrets med länkande funktioner anses oftast som filter i domänen för signalbearbetning. För förenklingens skull kallas denna endast "filter". Dessa filter har som uppgift att dämpa vissa frekvenser och låta andra passera.

En sådan filter måste nödvändigtvis vara en linjär filter. Icke-linjära filter kan orsaka att i de utgående signalerna kan det finnas frekvenser som inte fanns i den ingående signalerna.

Den mest kända gruppen av linjära tidsbeständiga filter är:

- Chebyshevs filter som har den bästa approximation med tanke på de idealiska egenskaperna i andra jämförbara filter för angiven ordning och vågrörelse.
- Butterworths filter vilken har maximalt balanserad frekvensegenskap.
- Bessels filter vilken har högst balanserad fasfördröjning.

Skillnaden mellan dessa filtergrupper ligger i det faktum att varje typ använder en annan polynom funktion för approximation av perfekt frekvensegenskap. Resultatet är en mängd överföringsfunktioner. Således väljer man beroende på vilken frekvensegenskap behövs (t.ex rullande- resp balanserade egenskaper) olika typer av filter.



Filter används främst inom telekommunikationsdomän. De har en viktig del i många tekniska framsteg och gav betydande vinster till telekommunikationsföretag. Så det borde inte vara någon överraskning att utvecklingen av de första filter pågick samtidigt med utvecklingen av överföringsvägarna.

Digitala filter



På det elektroniska, datatekniska och matematiska området anses digitala filter som ett system som utför matematiska urvalsoperationer. Målet är att förbättra signalfunktionerna.

Detta är den stora skillnaden jämfört med analoga filter som består av en elektronisk krets som arbetar fortlöpande med signaler. En analog signal kan behandlas av ett digitalt filter först när den har digitaliserats och representeras av en talsekvens. Därefter kan signalen gå via ett digitalt filter tillbaka och rekonstrueras som den filtrerade analoga signalen. Med analoga filter är insignalen direkt ansluten till elkretsen.



Ett digitalt filter kännetecknas av dess överföringsfunktion eller dess differentiella ekvationen. Matematisk analys av överföringsfunktioner kan beskrivas av deras möjlighet att svara på varje ingångssignal.

Finita impulssvarsfilter



Filter med finita impulssvar med förkortningen FIR är ett filter vars impulssvar är av begränsad varaktighet eftersom den under begränsad tid blir satt till noll.

FIR är en typiskt filter utan feedback då utsignalen beror endast på stickprover från insignalen. Om vi har N -prover i insignalen då är $N-1$ graden av filtret.



Differentialekvationer som beskriver systemet i tidsdomän:

$$y(n) = \sum_{k=0}^N a_k x(n-k)$$

Överföringsfunktion definieras som:

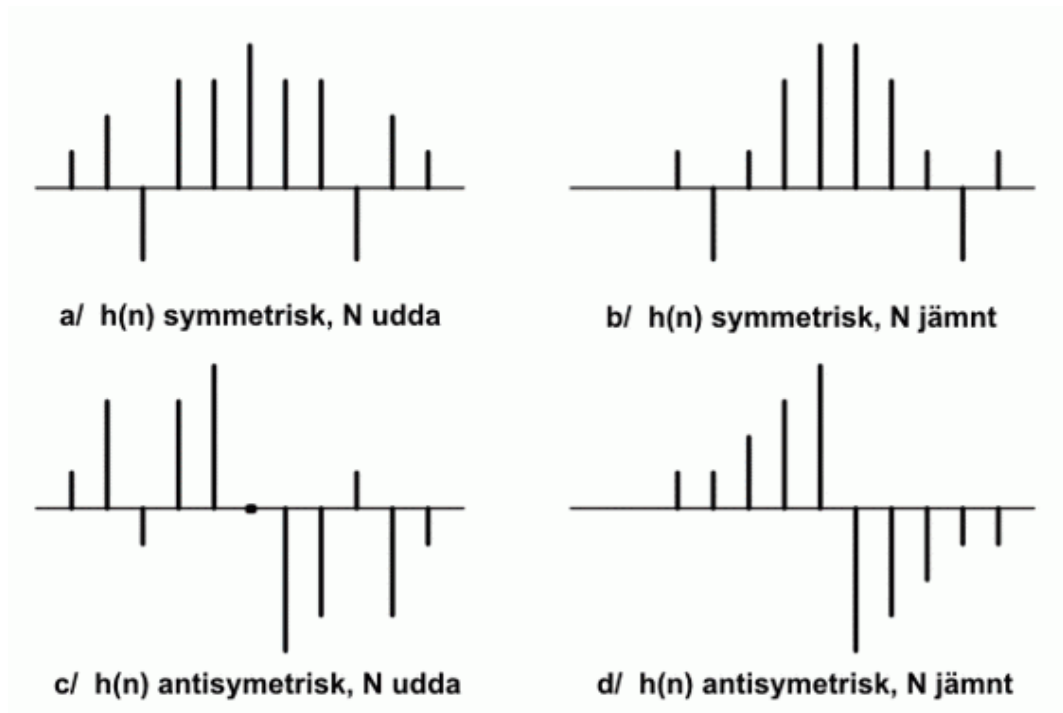
$$H(\Omega) = a_0 + a_1 e^{-j\Omega} + a_2 e^{-j2\Omega} + \dots + a_N e^{-jN\Omega}$$



De främsta fördelarna med FIR är följande:

- Ackumulering av fel undviks då feedback ej finns.
- Enkel implementering.
- Stabilitet. Då det inte finns feedback är alla poler* placerade i början (i princip "finns inte poler"). Ett system är stabilt om och endast om det absoluta värdet av varje pol är mindre än ett.
- Det kan helt enkelt skyddas med linjär fas. Detta görs genom symmetriska eller asymmetriska impulssvar

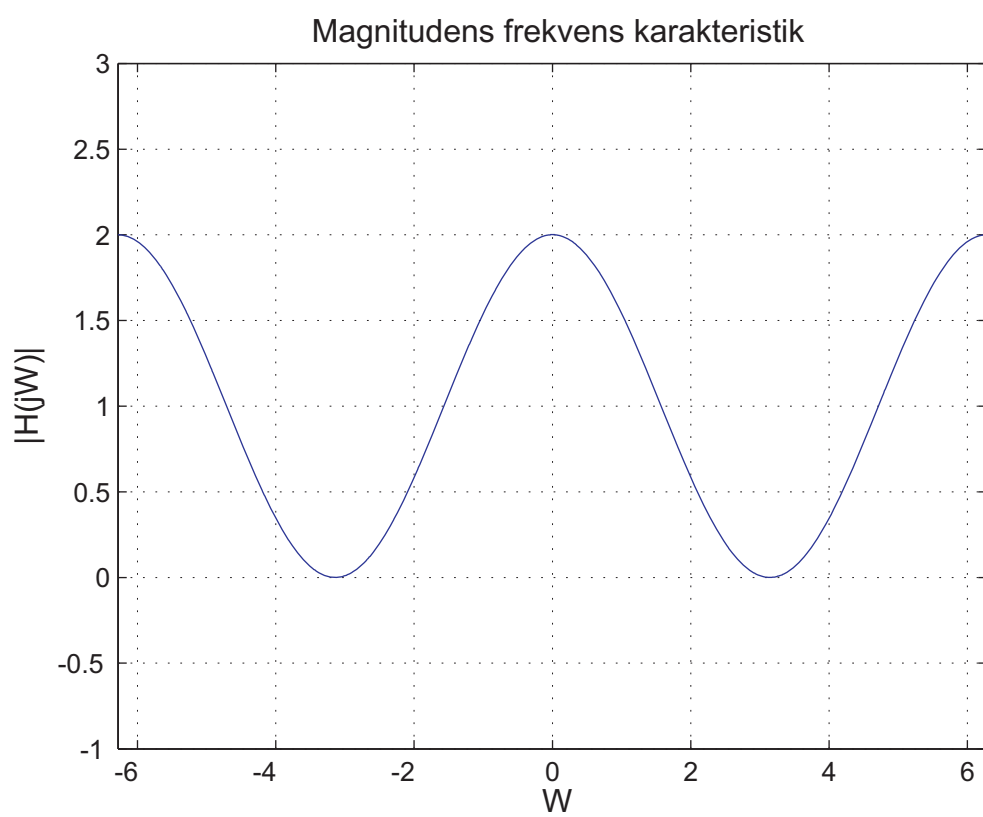
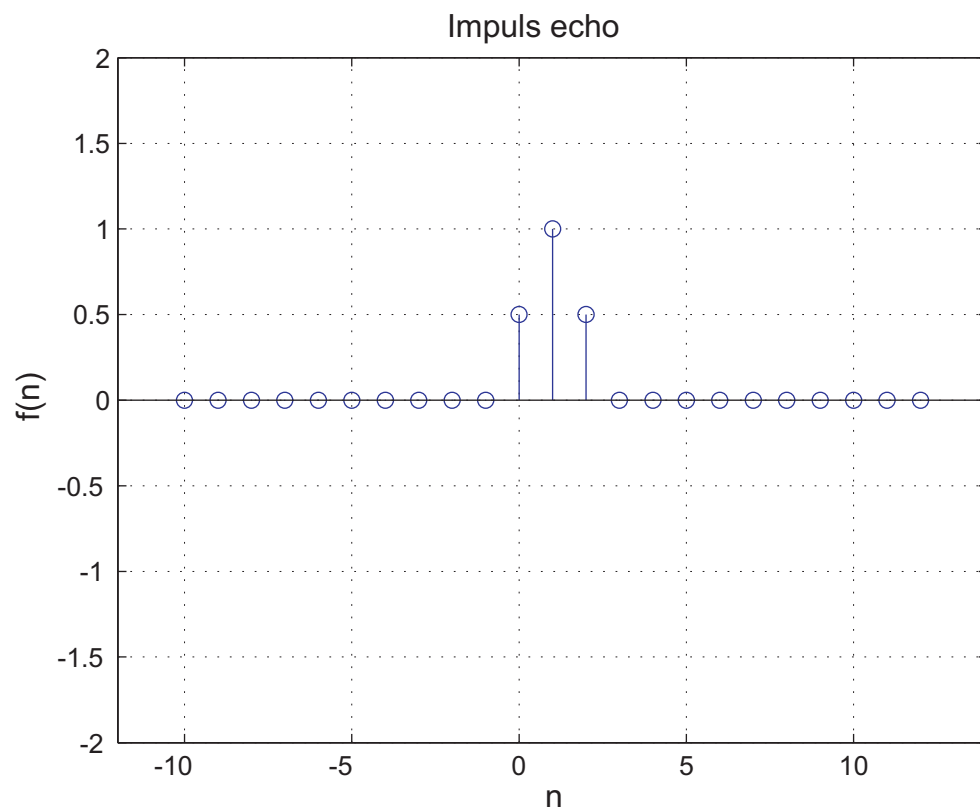
*Obs: poler är rötter i nämnaren av filteröverföringsfunktionen.

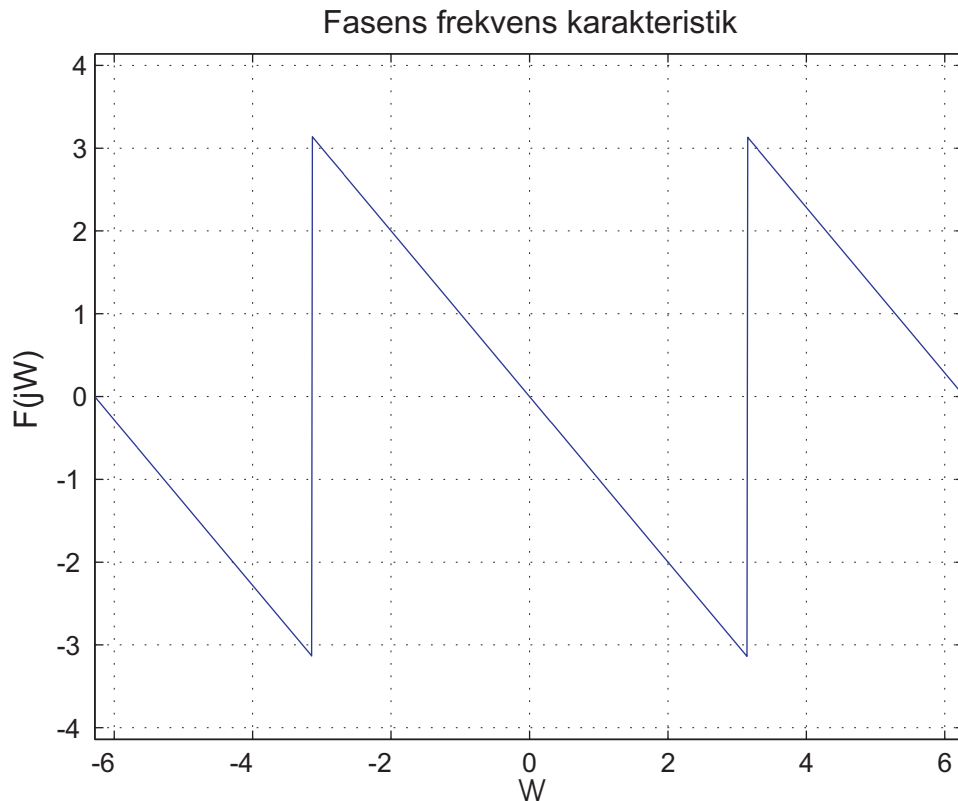


Impulssvar (för säkring av linjär fas frekvens kännetecken)

När man skapar filter måste man välja sådana koefficienter att filtret har de egenskaper som krävs. Dessa brukar anges i filterspecifikationen vilka oftast anger frekvenskännetecken. Det finns flera metoder för att hitta koefficienter. Bland de mest använda är:

- Intuitiv metod
- Frekvens provtagningsmetod
- Metod med fönster funktioner
- Metod för vägning av den minsta kvadraten





Impuls svar, magnitud och frekvensfas egenskaper för FIR-filer

Filter med oändligt impulssvar



Ett filter med oändligt impulssvar med förkortningen IIR är ett filter vars impulssvar förvärvar aldrig nollvärde.

Ett utdataprover av utsignalen ges från summan av N -prover från insignalen, vägda koefficienter a_k och utsignalens vägda koefficienter b_k . Detta är också uppenbart från den differentiella ekvationen som beskriver IIR system.



Differentialekvation som beskriver IIR-filtret:

$$y(n) = \sum_{k=0}^N a_k x(n-k) - \sum_{k=1}^N b_k y(n-k)$$

Definitionen för överföringsfunktion:

$$H(\Omega) = \frac{a_0 + a_1 e^{-j\Omega} + a_2 e^{-j2\Omega} + \dots + a_N e^{-jN\Omega}}{1 + b_1 e^{-j\Omega} + b_2 e^{-j2\Omega} + \dots + b_M e^{-jM\Omega}}$$

Överföringsfunktionen ges som andel av två polynomer och därmed kan stabilitet i ett system inte alltid garanteras. Som nämndes tidigare är systemet stabilt när det absoluta värdet för varje pol är mindre än ett. Vid instabilitet finns det några sätt

att stabilisera systemet, t. ex. med hjälp av fasfilter och eller med hjälp av PLSI algoritm.

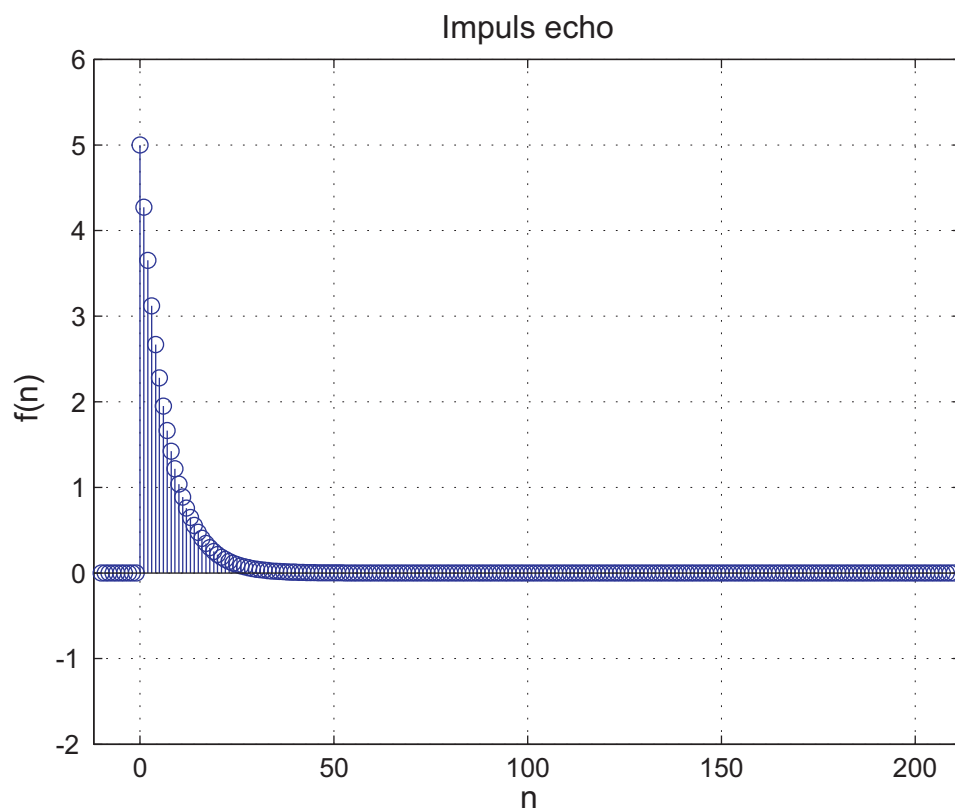
När ett filter skapas väljs koefficienter så att systemet har de föreskrivna frekvensegenskaperna. Detta innebär för IIR-filer att bestämma graden av täljare, nämnare och koefficienter för a_k och b_k . Designmetoder för IIR-filer kan delas i två grupper.

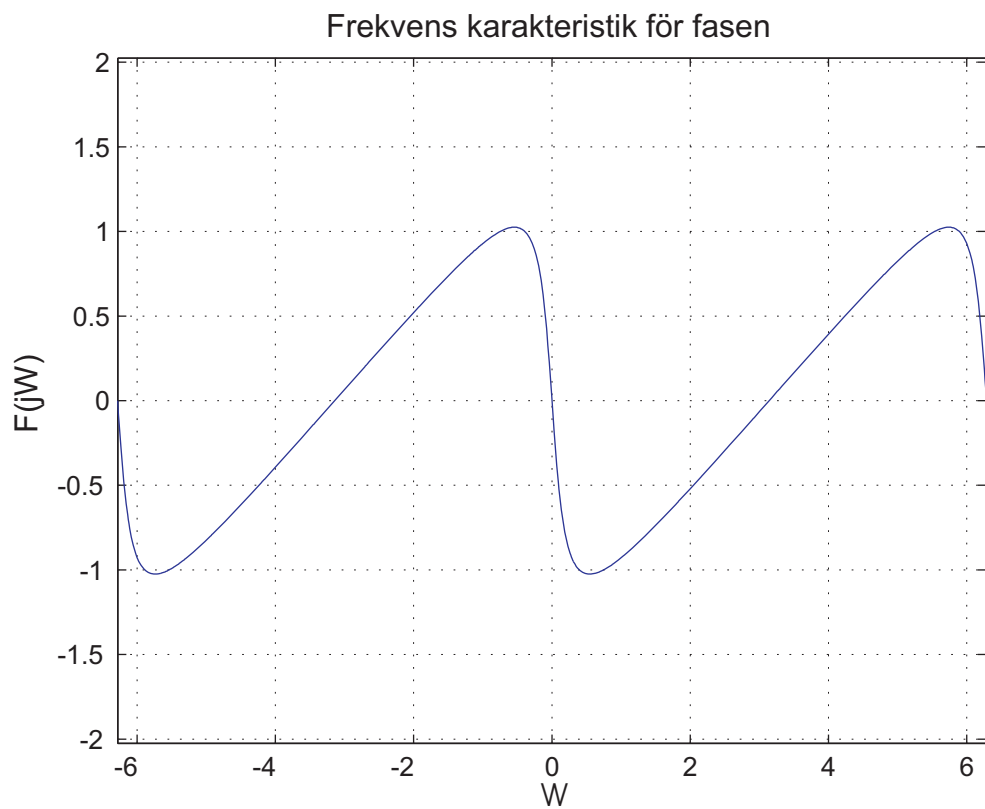
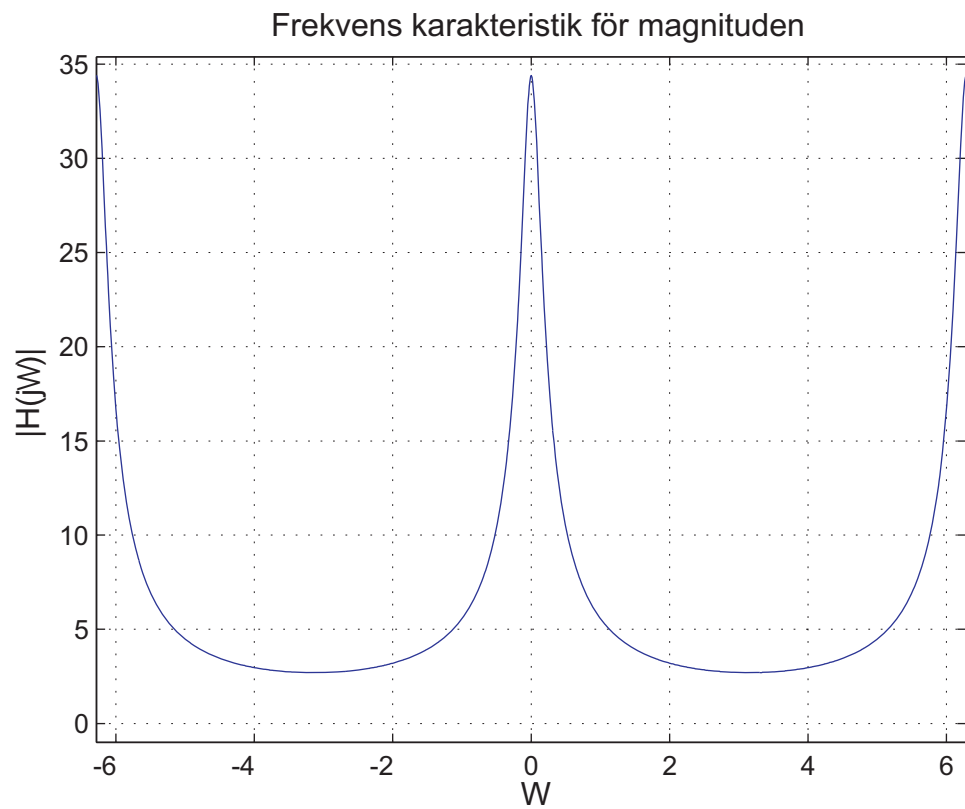
Till den första gruppen tillhör de med direktdesignmetoden:

- Intuitiv metod
- Frekvens provtagningsmetod
- Prony-metod

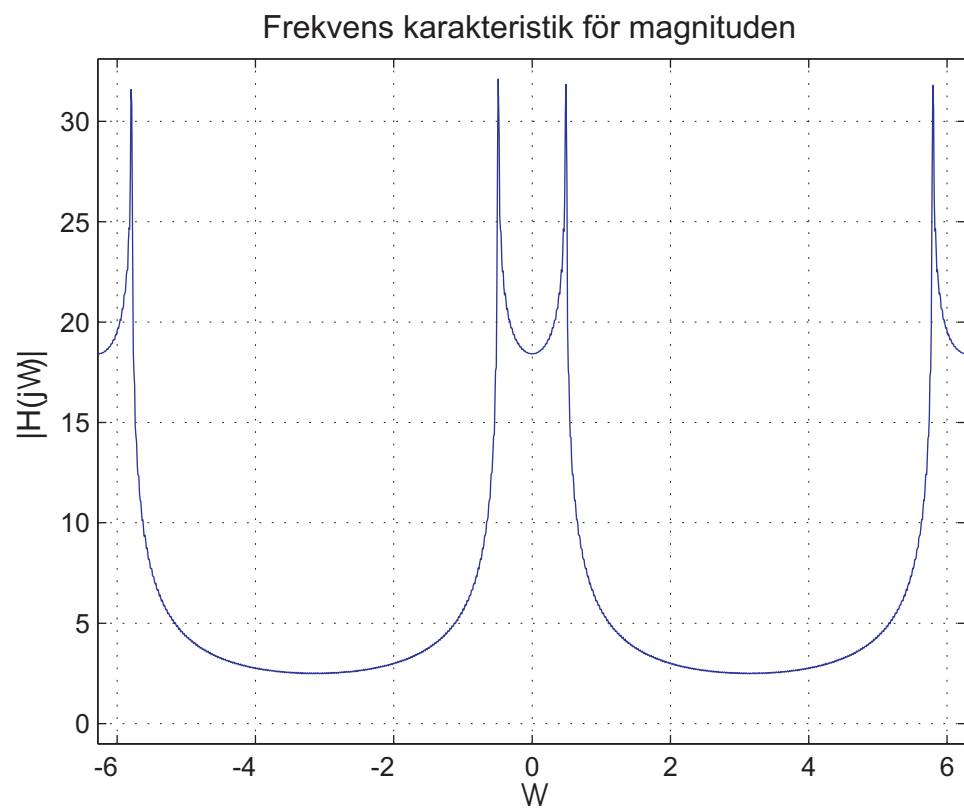
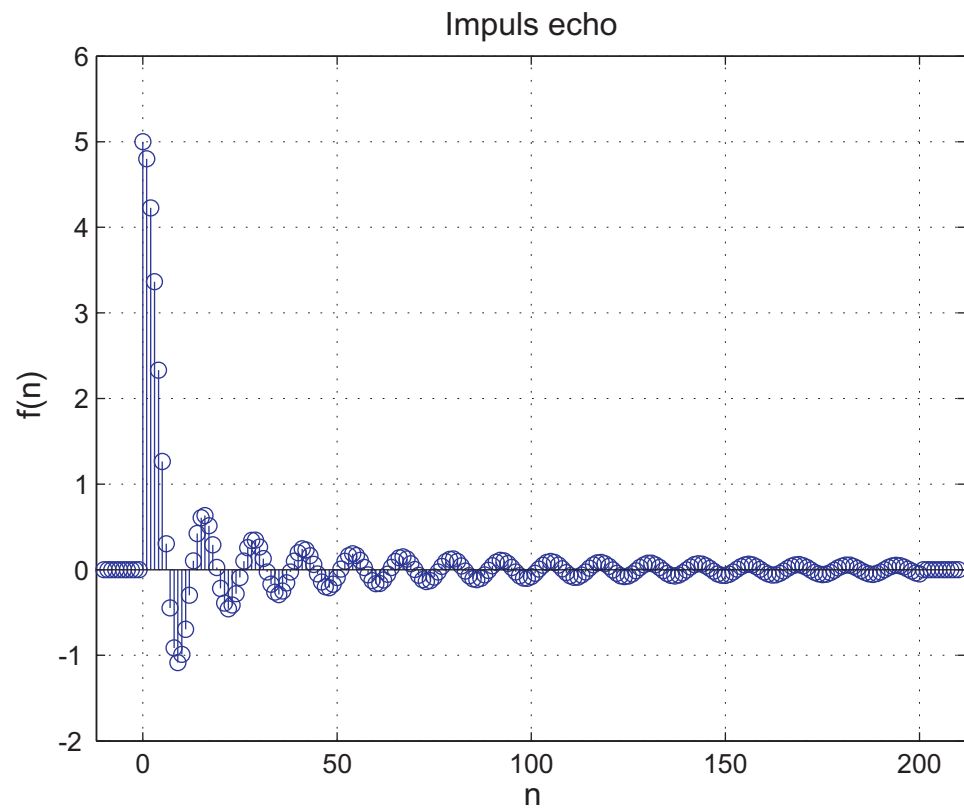
Den andra gruppen av designmetoder kallas den indirektdesignmetoden och är baserad på de reviderade designmetoderna för analoga filter. Vid formgivningen formuleras krav för IIR-filer. Sedan föreslås analog filter (Chebyshevs filter, Butterworths filter eller Eliptiskfilter), som konverteras till ett digital filter genom att tillämpa diskretiserings teknik. Bland de vanligaste finns:

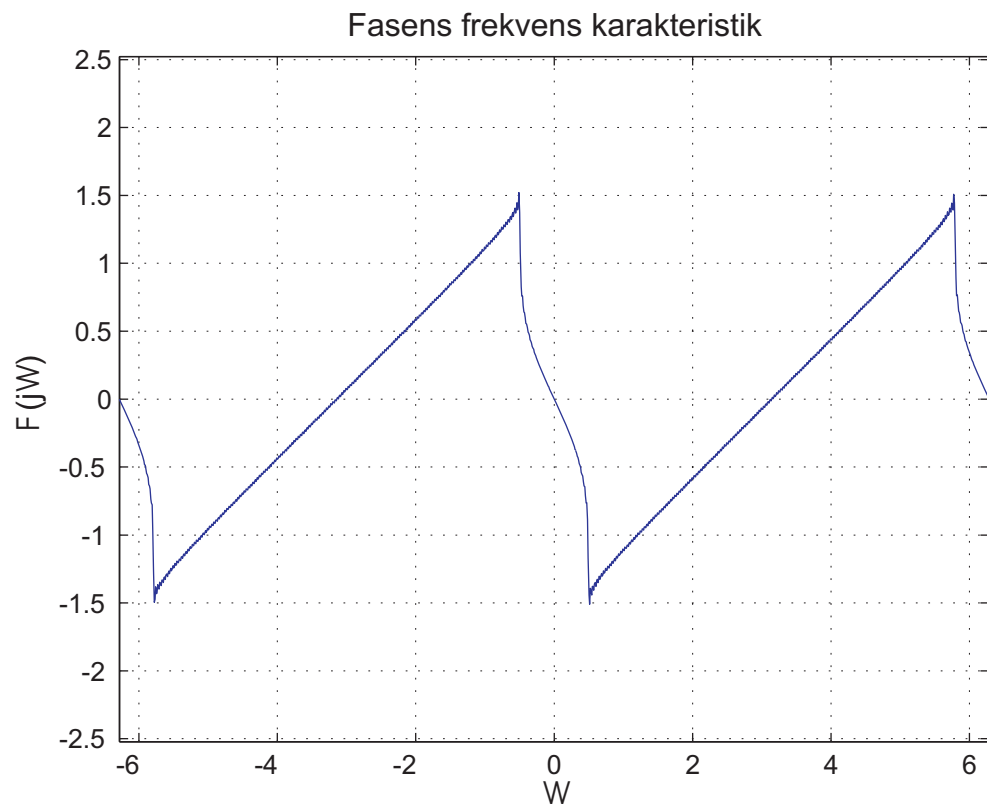
- Bilinjär transformationsmetod
- Impulsinvariant metod





Impulssvar, magnitud och frekvensfas egenskaper för IIR filter (exponential funktion)





Impulssvar, magnitud och frekvensfas egenskaper för IIR filter (sinc funktion)

+

Den största fördelen med IIR-filter är dess användning av färre rekursionsavtappningar och därmed konsumerar mindre dataresurser än motsvarande FIR-filter. De är därför användbara inom digital signalbehandling.

—

Nackdelen med IIR-filter är att de kan vara instabila. Implementering av IIR-filter är mer komplicerat än en implementering av FIR-filter.

4 Komunikationskanal

4.1 Komunikationskanal

En typisk kommunikationssystem består av ett sändningssystem (sändare), ett mottagningssystem (mottagare) och överföringslänk (överföringsmedium) som tillsammans bildar en telekommunikationslänk.

Telekommunikationskanal representerar en uppsättning av tekniska hjälpmedel som bara tillåter enkelriktad signalöverföring mellan två punkter, oavsett arten av de medel som används.

Telekommunikationskrets består av ett par ömsesidigt tillhörande omvända kanaler som även tillåter fram och tillbaka kommunikation. Denna kommunikation kan antingen vara simplex dvs den överför en signal i kretsen växelvis i ena eller andra riktningen eller duplex dvs den överför en signal i kretsen i båda riktningarna samtidigt).

Baserat på överförd signal finns det två typer av kommunikationssystem - analoga och digitala. Följande avsnitt kommer att beskriva de två systemen och deras grundläggande funktioner.

Analog kommunikationskanal



Analog kommunikation är ett sätt att överföra information som använder sig av kontinuerlig signal för överföring av meddelanden såsom konversation, musik, bild, video etc.

Analog kommunikationskanal tillåter föra över meddelanden från en punkt till en annan punkt eller flera punkter. För överföring av data med hjälp av analoga signaler används modulering. Därför är det nödvändigt att sändare och mottagare har kompatibel utrustning dvs lämplig modul på sändarsidan och demodulator på mottagarsidan.



Utrustning, som används för att sända och ta emot datasignaler består av modulator och demodulator och kallas MODEM (MODulator+DEModulator).



Analog kretsar omfattas inte av informationskavntisering som digitala kretsar. Detta innebär stor nackdel då slumpmässiga fel och signalskador som man inte kan reparera. Detta resulterar i störande brus i överföringar av ljud eller video. Regeln är att ju större överföringsavstånd desto större störande brus.

Digital kommunikationskanal

Sändare

Digitalt kommunikationssystem kan hantera analogsignal (exempelvis samtal). Med *analog-digital omvandlare* (*analog-digital convertor* **ADC**) kan en analog signal omvandlas till en digital form. I analog-digital omvandlare digitaliseras signalen. Detta förfarandet beskrivs mer detaljerat i avsnittet Digitalisering av analogsignal.

Ett annat block som arbetar med en digital signal är källkodens enkodare vars allmänna egenskaper gör att den effektivt omvandlar varje diskret symbol till en lämplig digital (vanligtvis binär) representation. Kodning processen görs vanligtvis redan i ADC. Digitalt meddelande i de flesta fall innehåller redundans.



Principiella egenskaper av källkodsencodare är att eliminera redundans (överskotsinformation). En kodare eliminerar effektivt redundans som leder i slutändan till ett mindre antal bitar (binära siffror "0" och "1") behövs för överföring av informationen. Följande enkla exempel visar hur man minimerar redundans: i stället för "och så vidare" kan "etc" användas.

I vissa system som inte har funktionen för kanalkodning, omvandlas utgångsdata direkt av källkodsencodare till en form lämplig för överföring via kanalen. Som nämnts ovan är rollen av enkodare att minska redundans i den digitala signalen. Vid kanalkodning är dock uppgiften att lägga till redundans som används för att upptäcka och korrigera fel. Utan redundansinformation skulle det inte vara möjligt att identifiera och eliminera de fel som orsakas av brus och störningar. Den sista enheten på sändarsidan är modulator.



När det gäller digitalmodulering moduleras analoga signalbärare av diskret signal.

Vilken typ av modulering används beror på typen av kanal och den överförda data. Vissa modulatorer är utformade för kanaler där brus förekommer ofta (t.ex. radio) andra är istället avsedda för kanaler med ett minimum av störningar (t.ex. optiska kablar).

Mottagare



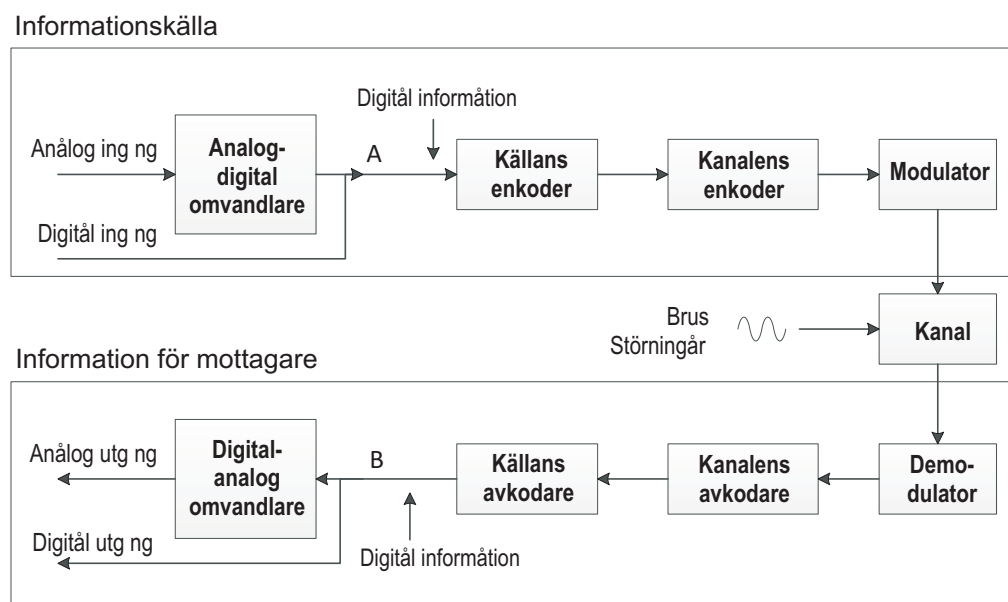
Operationer i mottagaren genomförs omvänt gentemot sändares operationer. En mycket viktig roll i designen och optimeringen av systemet för digital kommunikation är att minimera felen som inträffar under överföringen.

I den mottagna signalen uppstår vanligtvis fel på grund av brus i kanalen och dessa fel skall rättas till på mottagarens sida. Detta är en uppgift för demodulator som ändrar det numeriska värdet till signalvärde (t.ex. nr. 1 till värde 5 där 5 kan vara ampere, volt, etc.). Beroende på hur signalens parametrar (amplitud, fas och frekvens) överförs i en bärande signal väljs demoduleringsmetod. När signalen

moduleras av en linjärmodulering som AM används synkron detektorn. För vinkelmodulering måste användas en FM-demodulator eller en PM-demodulator. Dessa demodulatorer är sammansatta av olika typer av elektriska kretsar.

Kanaldekodarens huvuduppgift är att avkoda utdata från sändarens källkoder så exakt som möjligt. Som regel uppstår det i praktiken oftast fel i en överförd signal. På grund av redundansen orsakad av kanalenkodare kan fel upptäckas och rättas till så att man kan få den korrekta signalen. Detta är omöjligt att göra vid analog överföring.

Källdekodaren utför exakt en omvänd funktion som den som utförs av källkodarsändaren. Den lägger till den tidigare borttagna redundansen. När det gäller de digitala mottagarens utdata är dessa källsekodarens utdata samtidigt mottagarens utdata. När vi behöver de analoga utdata (t.ex. samtal) går data från källdekodern vidare till en *digital-analog omvandlare* (*digital-analog convertor DAC*). Denna omvandlare rekonstruerar i enlighet med teorin om provtagning och kvantisering det analoga originalmedelandet.



Blockdiagram över digital kommunikationssystem

5 Komprimeringsmetoder



Syftet med komprimeringen är att ta bort ovidkommande eller överflödigt information från de ursprungliga data så att man kan överföra och spara data mer effektivt. Detta förfarande spar på överföringskapaciteten samt bidrar till effektivare användning av lagringsutrymme. Med komprimering kodas användbar information med färre antal bitar i jämförelse med den ursprungliga framställningen.

Mängden data som skapas av dagens multimedia enheter är enorm. Om vi ville lagra en bild med 6 megapixel (3000 x 2000 pixlar med 8 bitar/färg) på en USB-sticka i okomprimerad form, skulle vi behöva mer än 17 MB utrymme (exakt 18 miljoner byte). Samma bild komprimerad med JPEG-enkoder skulle ta runt 3 megabyte lagringsutrymme (ca 3 200 000 byte) och vi märker ingen skillnad i bildkvaliten. Detsamma gäller för ljud- och videosignaler både under lagring och överföring i digitala nät.

Komprimeringsalgoritmer kan delas i:

- Förlustfrikodning
- Förlustkodning



Förlustfrikomprimering minskar antalet bitar med sökning och eliminering av statistisk redundans det vill säga information som upprepas eller kan erhållas från annan information i signalen. Förlustfri komprimering (även känd under benämning "compaction") tillåter att rekonstruera den ursprungliga signalen utan några ändringar. Det är därför lämpliga för arkiveringsändamål.

Förlustkomprimering försöker å andra sidan minska datamängden genom att ta bort den information som människor på grund av brister i sina sinnen inte kan registrera. När man använder denna komprimeringsmetod är rekonstruerat signal aldrig densamma som den ursprungliga signalen. Metoden är lämplig för applikationer när man har begränsat lagringsutrymme och kapaciteten i överföringskanalen

Förlustalgoritmer är generellt mer effektiva än förlustfria algoritmer men den rekonstruerade signalen kommer alltid att vara annorlunda än den ursprungliga. Kvaliteten på den komprimerade signalen är alltså en kompromiss mellan mängden av information som raderas från originalsignalen och önskad storlek av en komprimerad signal.

Förändringar i signalen för samtal och ljud som fortfarande är acceptabla utan att man märker någon skillnad gentemot originalsignalen kallas för knappt märkbar snedvridning (*just-noticeable distortion -jnd*). För stillbilder och video kallas detta för knappt märkbar skillnad (*just-noticeable difference - jnd*).

Komprimeringsförhållande som avspeglar effektiviteten av komprimeringsmetoden ges av:

$$C_r = \frac{N_n}{N_k}$$

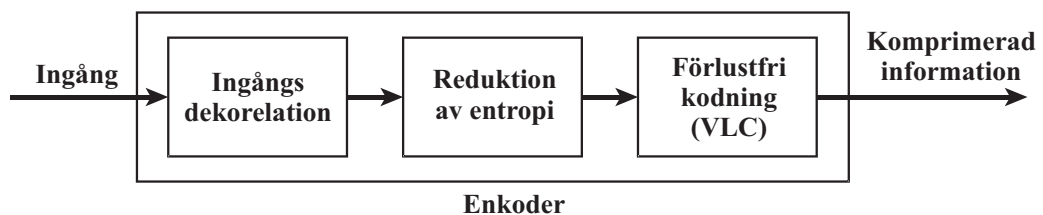
där N_n är antalet bitar av okomprimerad signal och N_k är antalet bitar av komprimerad signal.

Programvara som utför kodning av originalsignalen till en komprimerad form och avkodning av komprimerade signalen till sin ursprungliga form kallas kodek (förkortning från **koder-dekoder**).

Komprimeringsmodell

Komprimering processen består av tre grundläggande faser:

1. Indata dekorrelation
2. Reduktion av entropi
3. Förlustfri kodning



Allmän komprimeringsschema

Indata dekorrelation tar bort dubbel information från signalen. Detta kan göras på olika sätt eller i olika domäner:

- Tidsdomän (eller utrymmedomän)
- Parametriskdomän
- Signaledbrytning

Metoder baserade på linjär prediktion i tids- (eller utrymme-) domän, förutsätter att ett angränsande prov vid en tidpunkt (eller utrymme) är korrelerade (de liknar). Åtkomst genom nedbrytning av signalen bryter signalen till subband där separat övervakning av energi tillåts. Över till parametriskdomän går vi via specifika transformeringar så att vi kan få de karakteristiska parametrar för kodning och överföring av indata.

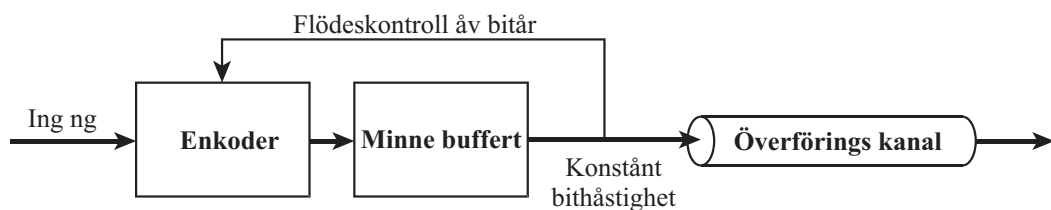
Kvantisering av signalprocess minskar entropin. Även om kvantisering alltid orsakar förlust av information reduceras denna förlust till ett minimum av tillräckligt exakt kvantisering.

Det sista steget i kodningprocessen är en förlustfri kodning även känd som entropi kodning. Under denna process ges ett kort kodord till de kombinationer av bitar (bytes) som statistiskt sett förekommer oftare i datamängden. Kombinationer av bitar (bytes) som statistiskt sett är mindre vanliga tilldelas ett längre kodord. Sådan kodning kallas också variabel längdkodning (*variable length coding - VLC*) eftersom kodorden har olika längd. Den mest effektiva algoritmen för att utföra förlustfri VLC-entropikodning är Huffmans kodning.

Det finns två lägen för överföring av komprimerad information genom överföringskanalen:

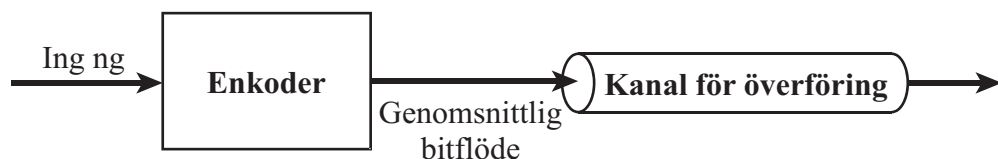
- Konstant datahastighet /bitrate/
- Variabel datahastighet /bitrate/

När man använder konstant datahastighet (bitrate) lagras utdata från enkoder i en buffert som säkrar konstant dataflöde till överföringskanalen. Dataflöde påverkar enkodaren genom att den levererar samma bithastighet. På detta sätt kan enkodaren påverka kvantiseringssteg och därmed kvaliteten på den rekonstruerade signalen.



Kodaren med konstant datahastigheten (bitrate)

När man använder sig av variabel datahastighet behövs inte buffert och enkodaren kan använda kvantisering steg enligt kraven från den kodade insignalen. Den rekonstruerade signalens kvalitet är oförändrad.



Kodaren med variabel datahastighet /bitrate/

Även om det kan tyckas att en variabel datahastigheten är det bästa valet gäller detta endast i fallet då överföringshastigheten i överföringskanalen är tillräckligt hög för det högsta dataflöde som en kodare med variabel datahastigheten kan producera. Dock om man har bara begränsad överföringshastighet i en överföringskanal så för att upprätthålla ett kontinuerligt flöde av data kommer den bättre kodaren vara den med konstant datahastigheten.

5.1 Komprimering av audiosignaler

För komprimering av allmänna audiosignaler finns det ett stort antal standardkodningar och komprimeringsmetoder. Många av dessa fokuserar sig på specifika typer av ljud (t.ex. tal) eller egenskaper (beräkningskomplexitet, förseningar, etc.)



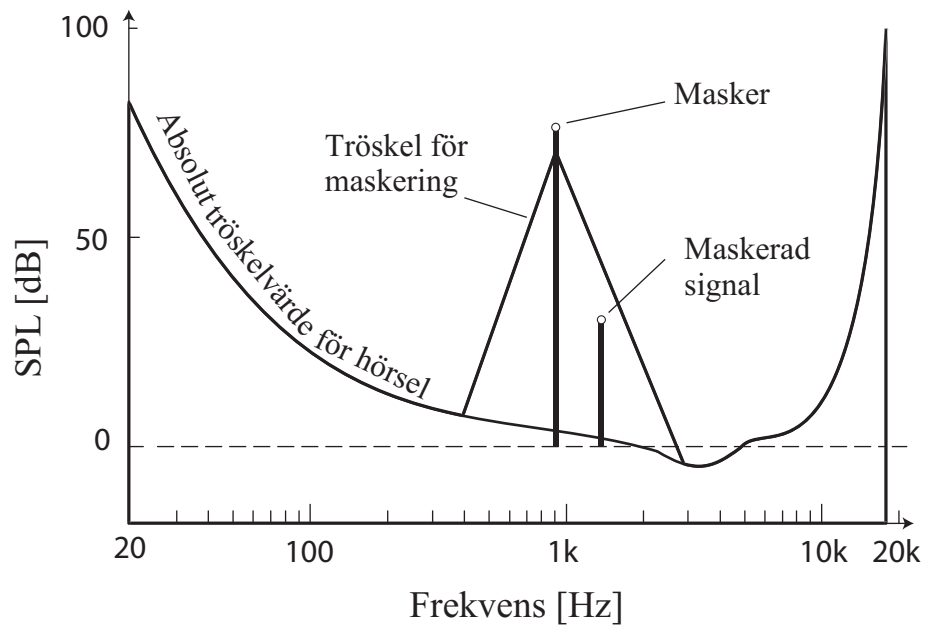
Samplingsfrekvens beskriver hur många stickprover erhölls från signalen varje sekund. Allmänt sett ju högre samplingsfrekvens desto bättre kvalitet, mer exakt och mera trogen inspelning. Bland de mest använda samplingsfrekvenser är: 8 kHz, 16 kHz och 22,5 kHz, 32 kHz, 44,1 kHz eller 48 kHz för varje ljudkanal.

Auditiv (hörsel) maskering – Hörsel maskering

Auditiv maskering är ett fenomen som yttrar sig som ett resultat av brister i mänskliga hörselsystemet. Det mänskliga örat kan inte uppfatta ljudets olika intensiteter som uttrycks med hjälp av så kallad absolut tröskelvärde för uppfattning av ljudet. Dessutom täcker ofta över ett högtljud det närbelägna tystare ljudet. Detta kan ske båda i tids- och frekvensdomän och därför indelas auditiv maskering i:

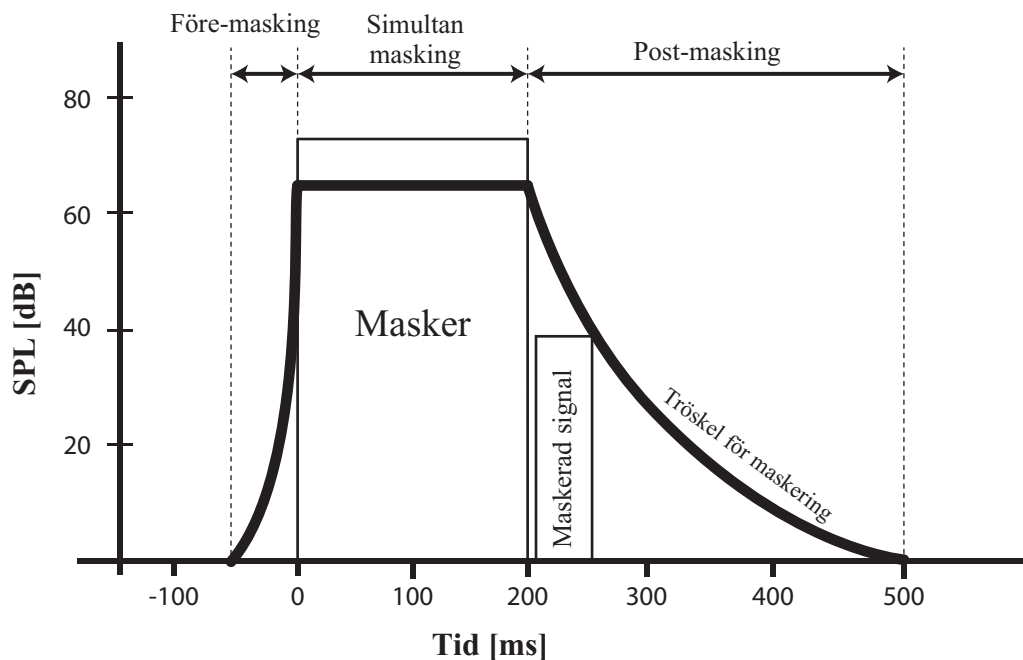
- temporör (icke-simultan) maskering
- frekvens (simultan) maskering

Ett hög ljud kallas för en maskerare. Om två ljud förekommer samtidigt kan en så kallad simultan maskering uppstå. En maskerare skapar en maskeringströskel under vilken man inte kan uppfatta några andra ljud. När signalen i närheten av en maskerare inte har tillräckligt intensitet och faller under tröskeln kommer den att maskeras. Följande illustration visar hur en maskerare kan övertäcka en tyst signal i frekvensdomän. Kombinationen av flera tröskelvärde-maskeringar flera masker och absolut tröskel för uppfattning skapar den globala maskeringströskel som kan förändras med tiden. Processen för att fastställa maskeringströskel kallas för psykoakustisk analys.



Maskering i frekvensdomän. När maskerintensiteten är högre än signalintensiteten och signaler är tillräckligt nära varandra kan en signal inte höras och man hör då bara en maskerare.

Under tiden för icke-simultan maskering kan en maskerare övertäcka signal som inträffar strax före (premaskering) eller strax efter (postmaskering). Intensiteten i en maskerare måste dock vara mycket högre än signalstyrkan.



Maskering i tidsdomän

För närvarande har frekvensmaskering undersökts tillräckligt exakt och har utbredd användning i många audiokodek. Å andra sidan har tidsmaskeringen ännu inte undersökts så ingående beroende på dess relativt korta varaktighet. Postmaskeringen tar i genomsnitt mer än 300 ms efter slutmasker och

premaskeringen tar 50 ms eller ännu mindre. Dessa tider är för korta för att noggrant analyseras eftersom kodek oftast arbetar med en bildrutalängd av minst 20 ms och då lämpar sig premaskering för bara 2 till 3 bildrutor.

MPEG

För närvarande kommer de flesta audiokodek genom ett arbet i Expertgruppen för rörliga (dynamiska) bilder - *Motion Picture Experts Group (MPEG)*, som är en del av den Internationella standardiseringsorganisationen (*International Standards Organization - ISO*). Under sin existens tog gruppen fram flera ljudformat som används över hela världen.

Som senare blir mer tydligt är en kodek från MPEG-gruppen baserad på förlustkodning vilket innebär modifiering av den ursprungliga audiosignalen och den rekonstruerade signalen är aldrig helt identisk med originalet.

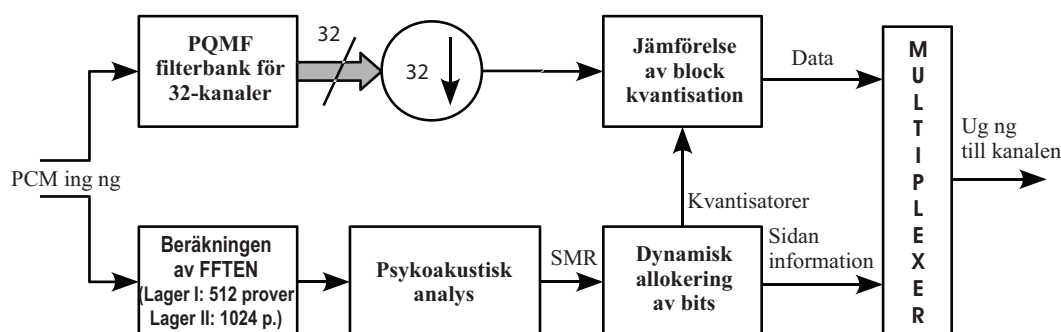
MPEG-1



MPEG-1-standard representerar en flexibel kodningsteknik som använder flera metoder t.ex. subband kodning, analys genom bankfilter, transform kodning, entropikodning och psychoakustisk analys.

Dessa standarder arbetar inom samplingsfrekvenser 32, 44.1, 48 kHz eller 16 bitar (prov och utdataflöde) som varierar från 32 till 192 kbit/s per kanal. Standardläge erbjuder 4 kanalerskodning: mono, stereo, dual mono, och förenade- stereon (endast platform III).

Standarduppbyggnaden innehåller 3 skikt som skiljs sig i beräkningskomplexitet, dröjsmål och utgångskvalitet. Skikt I (mp1) och II (mp2) liknar varandra och skiljer sig endast i några detaljer. Båda använder snabb Fourierstransformation (*fast Fourier transform - FFT*) men skikt I använder fönsterstorlek av 512 stickprover medan skikt II använder fönster med 1024 sticksprover. Upplösning i subbandskvantisering i skikt I stöds maximal av 15 bitar/prov och i skikt II 16 bitar/prov. Även om dessa skillnader tycks vara minimala visade det sig att skikt II ger samma eller ännu högre utgångskvalitet i bitflöde 128 kbit/s än skikt I med bitflöde 192kbit/s per ljudkanal.



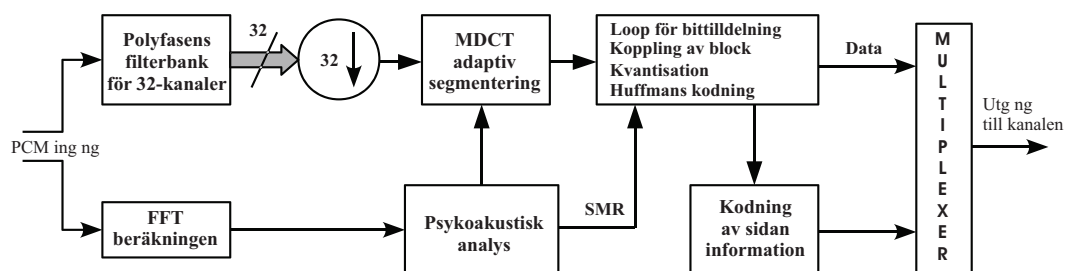
Allmän schema för enkoder MPEG-1 platformen I och II

Komprimeringen i båda platformar arbetar med PCM signalingång som indelas till 32 subdomäner. Under indelningen utförs FFT vars utdata passerar genom psychoakustisk analys och anger jnd. Beroende på maskeringströskeln för varje delband fastställs det mest lämpliga kvantiseringssteget så att önskad dataflöde och maskeringsnivå kommer att uppnås. Vid avslutningen blir utdata från enkodare kodade med Huffmansentropi kodning.

- + Även om MPEG-1 platform II ger acceptabelt resultat är det dominerande formatet MPEG-1 platform III (*MPEG-1 layer III*) vida använd och känt under förkortningen **mp3**. Baserat på platformar I och II innehåller den en hel del ny teknik som leder till lägre bitflöde (cirka 64 kbit/s per kanal) men med samma kvaliteten som dess föregångare.

Algoritmen arbetar med fönster av 1152 stickprover vilken delas i två partiklar (granules) med 576 prover. Var och en av dessa partiklar går genom en hybridfilterbank (ett system av bandpassfilter som används för att dela indata till subband där varje subband kan sedan hanteras separat) för att öka den frekventiella resolutionen. Varje delband transformeras till ett frekventiellt domän med hjälp av modifierad diskret kosinusomvandling (*Modified discrete cosine transform -MDCT*). Sedan kommer delbandet att upprepande tilldelas bitar och kvantisering. Under varje upprepning sker analys genom syntesprocessen processen för att bestämma den kvantiserade brusnivån.

- + Modifierad diskret kosinustransformation (**MDCT**) är baserat på diskret Fourierstransformation (DFT) men är utformad för signaler vars block av stickprover överlappar varandra. Insignalen sönderfaller (transformeras) i cosinusfunktioner. Jämfört med Fourierstransformation vars utdata är en uppsättning av komplexa nummer, är utdata från MDCT uppsättning av verkliga nummer som kännetecknar cosinusfunktionen. Dessutom har utdata från DFT samma antal koefficienter som antalet prover av insignalen medan MDCT har pga vare dess arbete med övertäckning från utdata hälften av antalet koefficienter.



Generellt schema för enkoder MPEG-1 nivå III

Den ursprungliga platformen III fick två förbättringar, **MP3pro** och **mp3surround**. Mp3pro innehåller en teknik som kallas replikering av spektral band (*Spectral Band Replication - SBR*), som används i de lägre data flödena för att avlägsna de ursprungliga höga frekvenserna. Dessa är sedan möjliga att rekonstruera med hjälp av ytterligare information från en komprimerad signal.

mp3 surround kan koda 5.1-kanalsljud (5 fullt band och en låg frekvens bas-kanal) till två kanaler för mp3. Av dessa är det möjligt att rekonstruera 5.1-kanalsljud med hjälp av sidinformation. När en dekoder inte stödjer MP3pro eller mp3 surround ignoreras då sidinformation och filen spelas av som en vanlig mp3.

MPEG-2

MPEG-2s standard är en formell efterträdare av MPEG-1. Denna innehåller 2 lägen, den ena bakåtkompatibel med MPEG-1 (Backward Compatible, MPEG-2 BC) och den andra, bakåt inkompatibla (Non-Backward Compatible, MPEG-2 NBC) som tar fram nya metoder och teknik för kodning.

De enda förändringarna i MPEG-2 BC jämfört med MPEG-1 är stödet för lägre samplingsfrekvenser (LSF) och flerkanaligt kodning liknande den som finns i mp3 surround. MPEG-2 NBC arbetar med avancerad audio kodning (AAC-Advanced Audio Coding) och har tagits fram som ett verktyg för effektiv kodning. Om mer hjälpmedel används erhålls bättre komprimering med bevarad outputkvalitet. Högre beräkningskomplexitet och förseningar är dock det pris man måste betala. Till skillnad från MPEG-1 använder inte MPEG-2 NBC en hybridfilterbank för att analysera en signal. Den använder endast MDCT i kombination med nya fönsterfunktioner. MPEG-2AAC blev standard för MPEG-4.

MPEG-4 AAC

MPEG-4 AAC försöker ta mer dominerand position för mp3-format genom att komma med samplingsfrekvenser från 8 till 96 kHz, 1 till 48 ljudkanaler plus 15 bas och 15 datakanaler med en upplösning på 8, 16, 24 nebo 32 samplingsbitar. Formatet AAC med låg komplexitet (Low Complexity-LC-AAC) är den ursprungliga MPEG-2 AAC och är lämplig för talkodning i dataflödet på 8-12 kbit/s. AAC formatet med hög verkningsgrad (High Efficiency (HE AAC) ger stöd för teknologin av SBS (version 1) och parametrisk stereokanal mode (version 2), som är baserat på den gemensamma profilen för stereo MPEG-1 plattform III.

Ogg Vorbis



Audiokodek Vorbis är en av de mest framgångsrika kodeks med öppen källkod. Sedan år 2000 när kodek standardiserades, blev den direkt konkurrent till mp3 formatet i MPEG-gruppen. Denna audiokodek stödjer frekvenssamplar från 8 till 192 kHz upp till 255 kanaler. Bitflödet av utdata är variabelt.

Kodning processer skiljer sig från MPEG-standard. För det första omvandlas signalen med hjälp av MDCT. Därefter bestäms frekvensdomän genom en grov uppskattning ett spektralt kuvert (en kurva som förbinder alla amplitudkoefficienter i ett frekvensspektrum) med hjälp av splitlinjärfunktion som kallas "floor". Skillnaden mellan "floor" och spektrum är kodad sedan med vektor för flerövergångskvantisering (multiple pass) kvantisering.

Ogg Vorbis har högre fordran på minneskapacitet, jämfört med mp3 eftersom dess „header“ innehåller kodningstabell för entropikodning (mp3 har en fast tabell) samt

inställning för dekodern. Ändå är det ett mycket lämpligt format för komprimering av ljudsignaler, och ger samma eller bättre ljudkvalitet vid samma bitsflöde som kodek mp3.

Windows Media Audio

Windows Media Audio (WMA) är en skyddad kodek skapad av Microsoft som svar på mp3 licensieringsbegränsningar. Det finns flera kodek versioner: WMA 9 är en direkt konkurrent till mp3 med stöd för stickprover med frekvens upp till 48 kHz med 16 bitar/prov och utdata flöde från 64 till 192 kbit/s med stöd för CBR och VBR.

WMA 10 Professional utökar i en duell med MPEG-4 AAC kodeks funktioner genom tillägg av provfrekvensen 96 kHz med 24 bitar/prov för 7.1 kanalerna. När enheten inte kan spela av 7.1 inspelning degraderas signalen automatisk (samplingsfrekvens, antalet bitar per prov, en minskning av antalet kanaler) till för enheten lämplig nivå.



WMA 10 tillhandahåller också ett system för komprimering av tal kallad WMA 10 Voice som ger bitflöde från 4 till 20 kbit/s. Dess attraktion ligger i möjligheten att dynamiskt byta mellan tal och en standardversioner av kodek när signalen som bearbetas är alltför komplex. Utöver detta ger WMA 10 bland annat ett förlustfritt läge vilket kan minska storleken på den ursprungliga PCM-signal från att halveras till bara en tredjedel.



Kodek WINDOWS MEDIA AUDIO 10 Professionell ger högre subjektiva kvalitet vid 64 kbit/s jämfört med MPEG-4 AAC v2 vid 70% av jämförelser.

5.2 Komprimering av tal

Med tanke på att tal är i grunden en audiosignal har det några specifika egenskaper som tillåter att använda mer radikal komprimeringsteknik än för allmänt ljud. Först och främst anses talsignalen vara medlet för att sprida information. Denna information behöver inte vara efter rekonstruktionen exakt som i originalet för att vara förståeligt. Detta ger en möjlighet att avlägsna några av egenskaperna för ljudet. T.ex. urval prover från ett vanligt telefonsamtal i en frekvens på 8 kHz (jämfört med standard ljudsamplingsfrekvens på 44.1 kHz) vilket betyder att endast 4 kHz-bandbredd tas i anspråk.

För det andra är talsignalen enkel i jämförelse med exempelvis en inspelning av ett rockband då vanligen denna fångar upp en enda talare utan några musikinstrument. För att uppnå det renaste möjliga talet kan man dessutom använda algoritmer i syfte att dämpa buller och bakgrundsljud.

De metoder som används för komprimering av tal kan delas in i följande grupper:

- Vågformigkodning
 - i tids domän
 - i frekvens domän
- Vokoders
 - linjär förutsägbarkodning
 - formantkodning

För kvalitetsbedömning beträffande olika algoritmer för talbearbetning finns det olika grader av begriplighet. Dessa beskriver hur en tal är förståeligt och begriplig. I resonemanget bör ingå de olika egenskaper, t.ex. ljudvolym, icke-linjärdistorsion, bakgrunds brus, eko m.fl. För jämförelse av begripligheten, finns Det finns huvudsakligen två skalor då begripligheten jämförs: index för taltransmission (Speech Transition Index - STI) och den gemensamma begriplighetsskala (Common Intelligibility Scale - CIS), som har ett intervall från 0 (sämsta) till 1 (bästa), eller från 0% upp till 100%. Allmänt sett är ett tal förståeligt när algoritmen når en värde på skala av minst 0,5 eller 50%.

Tidsdomän

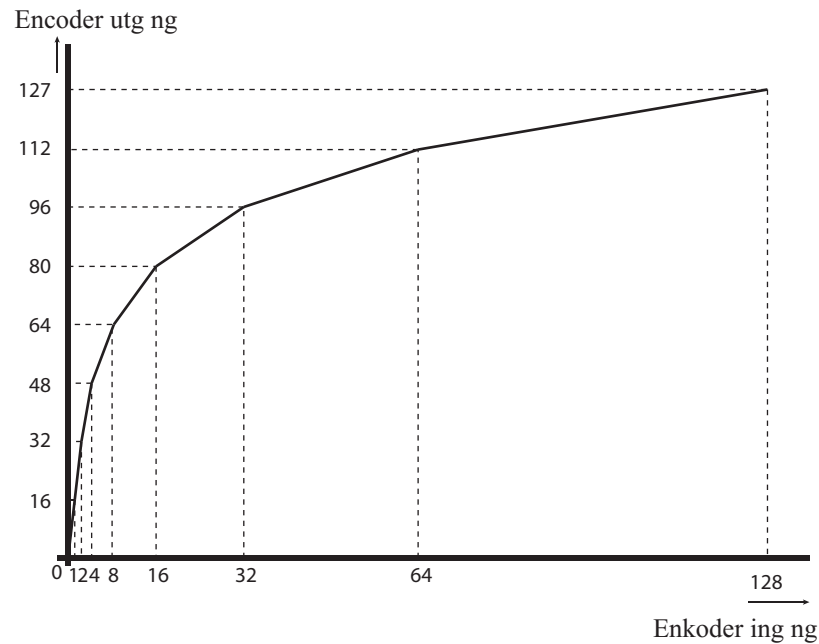
Vågformigkodning i tidsdomän utförs av PCM-teknik. Medan linjär PCM använder samma avstånd mellan kvantiseringssnivåer använder icke-linjär PCM en icke-linjär kvantiseringsskala eller dess modifikation i form av dynamiskkomprimering av insignalen på sändarsidan och dess utvidgning på mottagarsidan.

Rekommendation från G. 711 definierar två kompressionsegenskaper: μ -lagen (μ -law i USA och Japan) och A-lagen (A-law i Europe). T.ex. A-lagens karakteristik ges med relation:

$$F(x) = \text{sgn}(x) \begin{cases} \frac{A|x|}{1 + \ln(A)}, & |x| < \frac{1}{A} \\ \frac{1 + \ln(A|x|)}{1 + \ln(A)}, & \frac{1}{A} \leq |x| \leq 1 \end{cases}$$

där $\text{sgn}(x) = \pm 1$ för positivt eller negativt värde av x och A är parameter för komprimering.

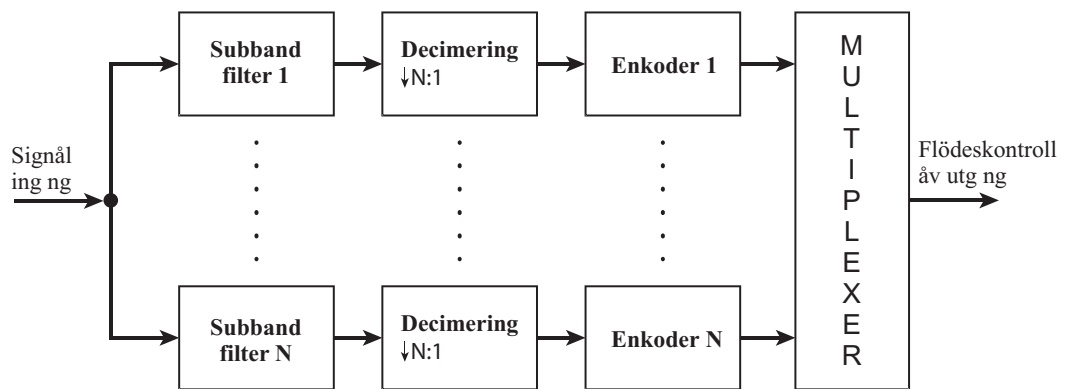
Vanligen $A=87,7$.



Ett exempel på A-lags compansionkurvan. Högre frekvenser (representeras av högre nummer på den horisontella axeln) är kodade med färre antal värden än de lägre frekvenser.

Frekvensdomän

I frekvensdomän används subbandsmetoden och adaptiv transformkodning. Vid subbandskodning (Subband Coding - SBC) delas en talsignal med hjälp av en bandpassage uppsättning (filterbank) upp i ett antal frekvensband och signal minskas så att antal stickprover reduceras. Varje band kodas sedan självständigt, oftast med hjälp av ADPCM-metoden som möjliggör flexibel kvantisering och tilldelning av bitar.

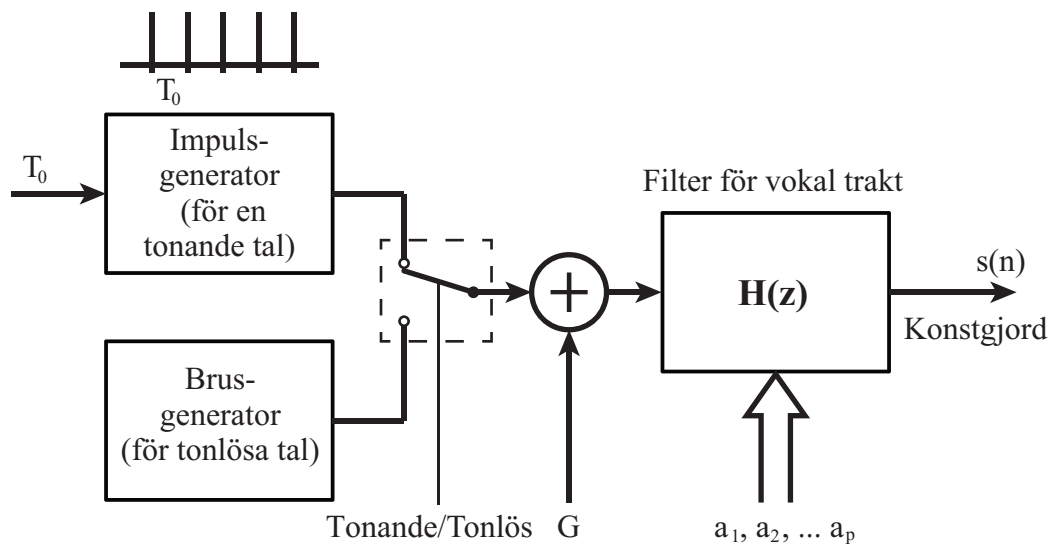


Exempel på en filterbank

I stället för ADPCM kan alternativa metoder användas. T.ex. teknik baserad på adaptiv transformationkodning (*Adaptive Transform Coding* - **ATC**). Här omvandlas signalen till ett frekventiel domän med hjälp av FFT, eller delas i subband med en annan omvandling. Sedan tilldelas bitar dynamisk till stickprover och till var och en av subband efter deras behov.

Förutsägbar linjärokodning

Naturlig mänsklig tal kan förstås som återgivning av talarens vokala domän på excitationssignal i vårt fall på luft utandad från lungorna. Utsignalen är modulerad med egenskapsförändringar i den vokala domänen (stämbanden, munhålan, tänder etc.). Om vi tittar på denna process från för signalanalysperspektiv kan vi representera utsignalen med hjälp av excitationssignal och med filter som representerar domän för röst med tidsmässiga parameterförändringar. Dessa koefficienter beräknas ungefär varje 10-30/ms. För beskrivningen av koefficientens funktioner i vokal domän finns många metoder. Den vanligaste är en metod baserad på linjär förutsägbarelse. Det hörs att dess namn att metoden baseras på linjär förutsägbarkodning (*Linear Prediction Coding* - **LPC**).



Det allmänna schema för LPC dekode

LPC-koefficienter minimerar kvadratiska avvikelser mellan ursprungliga och förutsägbara talprover. Som man kan se består talmodellgeneratoren LPC av två delar:

- Domän för vokal excitation
- Domän för vokal filter

Excitation inom vokal domän använder sig av pulsgenerator och brusgenerator mellan vilka är det möjligt att växla fram och tillbaka beroende på röstklngen. Excitationssignalen är sedan förstärkt med förstärkaren (G) till önskad nivå.



Stämbanden vibrerar, darrar med en grundläggande frekvens f_0 , från vilken kan härledas den grundläggande stämbandsperioden T_0 . Ju högre frekvens, desto högre är höjden i talet.

Beroende på om stämbanden används eller inte används, delas talljudet i:

- **tonande** ljud produceras med vibrationer av stämbanden över vilka luft passerar från lungorna t.ex. „a“, „v“, „z“. Alla vokaler är tonande.
- **tonlösa** ljud produceras endast med hjälp av brussliknande luftflöde och stämbanden är inte inkopplade, till ex. „s“, „c“, „f“.

Filter för vokal domän ges av linjär digital filter med det slutliga återgivningen (FIR filter) vars överföringsfunktionen kännetecknas av relationen:

$$H(z) = \frac{G}{1 + \sum_{i=1}^p a_i \cdot z^{-i}} = \frac{S(z)}{E(z)}$$

där a_i är filterkoefficienter och p är filtertillstånd. Om $S(z)$ representerar utdataprovet och $E(z)$ är excitation. Vid nästa prov $s(n)$ får man en linjärkombination av tidigare prov med excitation $G \cdot e(n)$:

$$s(n) = G \cdot e(n) - \sum_{i=1}^p a_i \cdot z^{-1} = G \cdot e(n) - a_1 \cdot s(n-1) - \dots - a_p \cdot s(n-p)$$

För användning av LPC-talgenerator behöver vi ange dessa parametrar:

- röstsegment
- basperioden T_0
- filterparametrar (förstärker G och koefficienter a_i)

+ Bitrate för talsignal kodade med LPC varierar från 1,2 till 2,4 kbit/s och har en tydlighet med omkring 80-85%.

— Rekonstruerade talsignal har en maskinklang som orsakas av två skäl:

1. Det är svårt att segmentera tal exakt till de tonande och tonlösa segment eftersom intonation minglar i naturliga tal.
2. Basperioden (som utmärker den talandes röst) ändras i naturliga tal oftare än ruttalängden och ändringarna är inte periodiska.

För att undanröja ofullkomlighet i LPC-metoden finns förfaranden som kodar skillnaden mellan den ursprungliga signalen och signalen som genereras genom LPC.

Linjär prediction med uppvaknande residue (*Residually Excited Linear Prediction* - **RELP**) sänder direkt skillnaden mellan en original och en renkonstruerad signal. På mottagarens sida används för signalrekonstruktion LPC-koefficienter och för mer exakt rekonstruktion läggs residum till.

En efterträdare till RELP-algoritm är *Code Excited Linear Prediction* - **CELP**. Algoritmen bygger på syntesanalys princip som optimerar sedan den syntetiserade signalen i en sluten perceptionsslinga. En sökning i kodboken görs och den väljer den lämpligaste excitationsfunktion. Tillsammans med LPC-koefficienter överförs bara en position i kodboken. Excitationsfunktionen kan kodas med hjälp av kvantiseringsvektor.

— CELP-metoden uppnår bitrate från 4 till 8 kbit/s. Dess nackdelar är den relativa beräkningskomplexitet och förseningar i nivå med 35 ms.

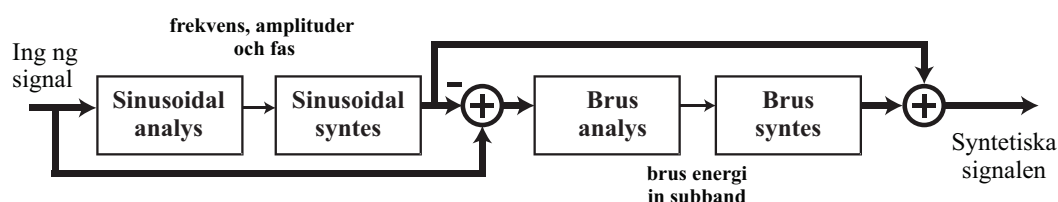


CELP modifierad med en liten fördröjning (Low Delay CELP, LD-CELP) minskar tidsfördröjningen till 2 ms i bitflödet f av 16 kbit/s och blev en del av ITU-T -organisationens standard G.728. Nästa kodek baserat på CELP är Speex,

utvecklat av Xiph.Org, som skapade kodek Ogg Vorbis. Speex har en öppen källkod.

Sinusodialkodning

Sinusodialkodning bygger på antagandet att varje ljudsignal är en kombination av deterministiska och stokastiska komponenter. Man kan presentera deterministiska mappar med harmoniska funktioner (sinus, cosinus) och stokastiska mappar med hjälp av brus eller genom att använda en annan parametrisering. Principiellt schema för sådan enkodare presenteras nedan. Sinusoider är ihopkopplade med tidsskiftande frekvenser som man förväntar sig bildar en enda kontinuerlig ton.



Allmän schema för sinusodial enkoder

Denna modell, kan inte hantera snabba ljudförändringar på ett lämpligt sätt och därför har en tredje komponent lagts till en så kallad transient som modellerar snabba förändringar i signalen. Detta leder till en sinusoid+ transient+brus-modell (STN).

Annan utökning av SN-basmodeller är Harmonisk+separat linje+brus-modell (HILN). I detta utförande är sinuskurvan indelad i två grupper; en harmonisk och en individuell. I den harmoniska delen är sinuskurvor representerade som multiplar av harmoniska basfrekvenser och endast dessa multiplar sparas. Sedan kodas enskilda sinusoider och återstoden (residuum) behandlas som brus.

+ Från sinuskodning förväntas bra bearbetning av enkla signaler som vanligtvis består av harmoniska ljud t. ex. tal. Denna SVOPC-teknik användes av den första SKYPE-kodek som med 20 kbit/s var av god kvalitet och dessutom resistent mot paketförluster.

— Beräkningskomplexitet med SVOPC-kodek ledde dock till skapande av en ny kodek med förkortningen SILK som är baserat på LPC.

Baserat på SILK-kodek med tillägg av kodekegenskaper från **CELT** (*Constrained Energy Lapped Transform*) var i september 2012 en ny kodek med namn Opus standardiserad. Denna kodek med bra prestanda är i stånd att dra nytta av SILK-kodek på låga frekvenser med en liten fördröjning och av CELT-kodek vid högre frekvenser och kan på begäran växla mellan dem. SILK-kodek är mycket lämplig för talkodning och allmän ljud samt även för online-applikationer som VoIP och live streaming.

5.3 Komprimering av stillbilder



Syftet med komprimering av stillbilder är reduktion av onödig information i bilden för att minska bredden i den erforderliga överföringsbanden och i lagringsutrymmet. På samma sätt som audiosignaler finns det även stillbilder med förlust- resp förlustfri algoritmer. Dessa används beroende på användnings domän.



När man analyserar bilderna används automatiskt tvådimensionella transformationsversioner.



Då man sparar stillbilder på datorn sparas dessa i olika färgdomän. Nedan presenteras de vanligaste.

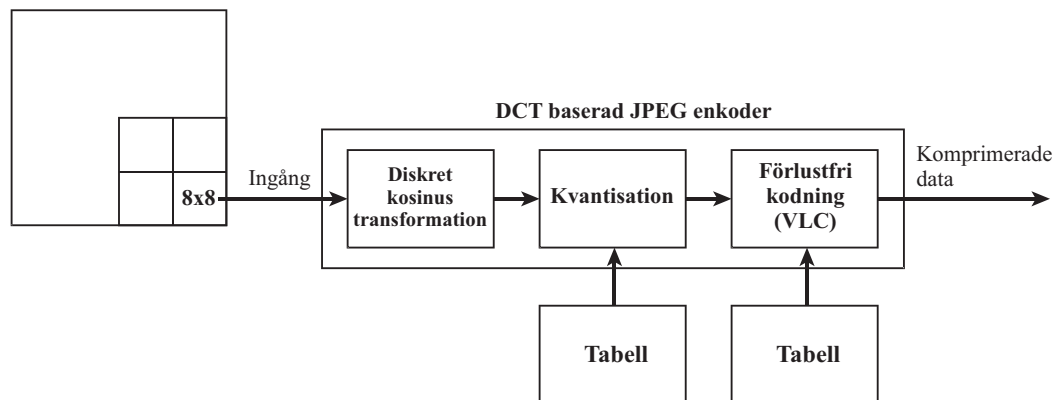
De utvalda färgdomäner

Förkortning	Betydelse	Förklaring
RGB	Red, Green, Blue (röd, grön, blå)	Varje pixel bestäms med hjälp av en kombination av ljus från 3 färger. Kombinationen av de högsta nivåerna från alla 3 färger ger vit färg. Denna används för skapande av bilder med hjälp av ljus.
RGBA	Red, Green, Blue, Alpha (röd, grön, blå, alfa)	Samma betydelse som i RGB. Tilläggskanal Alfa beskriver transparency.
YCbCr eller YUV	Ljusstyrka (Y), Blå färgläggning (U), Röd färgläggning (V)	Ljusstyrkan skiftar från svart till vit. Blå och röd färgläggning beräknas från RGB källa. Även om YCbCr märks som YUV, existerar sinsemellan skillnader vid beräkning av färgmappar.
CMYK	Cyan, Magenta, Yellow, Black (cyan, magenta, gul, svart)	Varje bildpunkt bestäms genom en kombinationen av fyra färger. Kombinationen av de högsta nivåerna av alla färger ger svart färg. Denna används främst i tryck.

JPEG

JPEG-standarden är en av de mest förlustanvända bildformat för att lagra bilder. Den är uppkallad efter den förenade expertgruppen för stillbilder (Joint Picture Experts Group). Formatet lanserades 1986. Detta format når upp en komprimeringsgrad på 10:1 med en knappt märkbar förlust i kvaliteten.

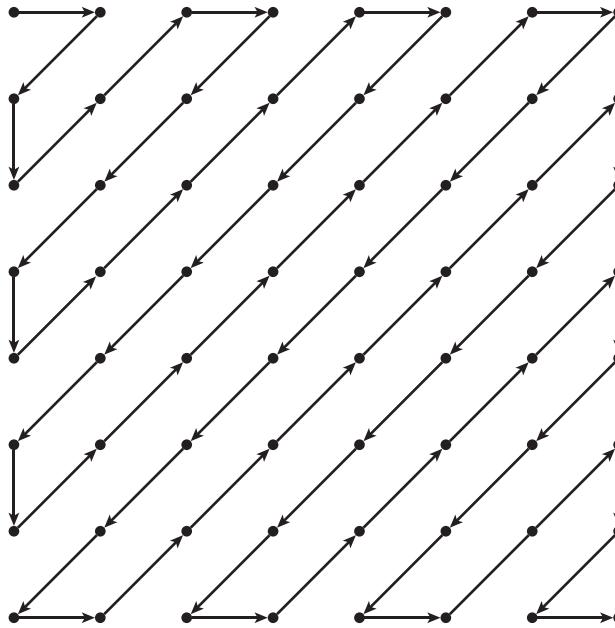
JPEG-algoritmen är baserad på en tvådimensionell diskretkosinustransformation (DCT). Indatabilden omvandlas till färgutrymme $YCbCr$ som har bättre egenskaper än RGB. Bilden delas sedan in i icke-överlappande ruta i en storlek på 8×8 punkter omvandlas med hjälp av DCT. De erhållna koefficienterna är kvantiserade samtidigt som några mindre betydande koefficienter tas bort (detta är typiskt för förlustkomprimering). Koefficienterna ordningssorteras sedan till en endimensionellsekvens och kodas utan någon förlust. Bildkomponenterna (Y , C_B a C_R) kodas gradvis.



Allmän schema för JPEG enkoder. Indata är en ruta av 8×8 punkter av ljusstyrka eller färgmappen.

Kvantisering är nyckeln till kompression i JPEG-algoritmen. Kvantisering är icke-linjär eftersom det mänskliga ögat är mer känsliga för förändringar i de låga frekvenserna. För att säkerställa denna kvalitet resp justerbart komprimeringsförhållande har man införts kvalitetsfaktorn q_f . Denna varierar i ett intervall mellan 1 och 100 och justerar kvantiseringsmatrisen.

Till sortering av koefficienterna används så kallad “zig-zag” avläsning som börjar vid det övre vänstra hörnet. Om varje 8×8 ruta kodas gemensamt kallas förfarandet för JPEG-baslinjekodning. En annan metod är att koda de första vänstra övre hörnen av alla bildrutor och därefter följande koefficienter från alla rutor. Sådant förfarande kallas progressiv JPEG. Dess fördel är den gradvisa rekonstruktionen av bilden redan under nedladdningen. JPEG erbjuder också ett hierarkisk system där bilden är kodad med hjälp av pyramidmetoden i skikt (layers). Varje pixel i ett högre layer uppstår genom användning av en specifik operation på en 2×2 rutor som ligger direkt under den. Dekodern kan avkoda varje skikt (layer) separat och detta möjliggör att använda bilden i olika upplösningar.



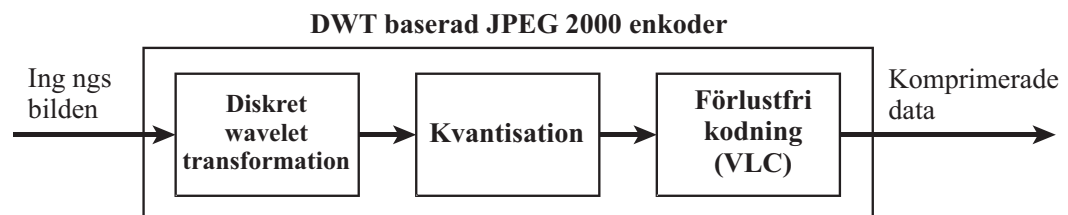
Sortering av koefficienter i en endimensionellvektor med hjälp av "zig-zag" avläsning



JPEG-standarden stöder också förlustfri kodning med hjälp av prediktionskodning och VLC-förlustfrikodning som utesluter DCT och justering av spektrummet. Typiskt kompressionsförhållande är kring 2:1.

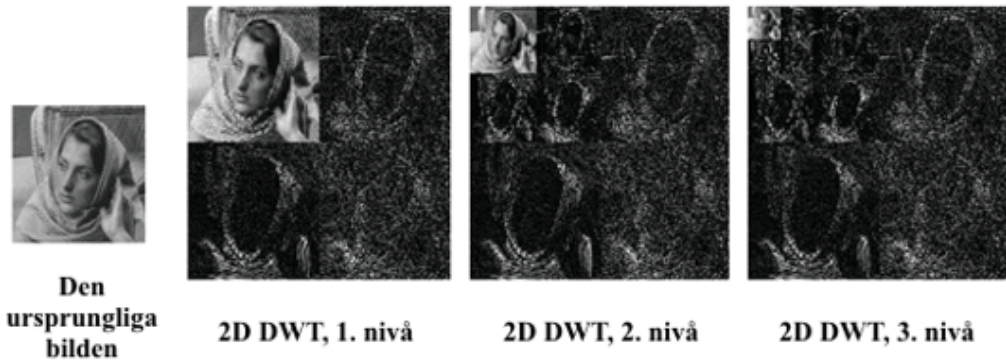
JPEG 2000

JPEG 2000 försöker ersätta en icke-perfekt men fortfarande väldigt populär JPEG-format. Detta utökar funktionerna i dess föregångare, förbättrar kvalitet/komprimering-förhållandet, samt tillåter en skalbar förlust- och förlustfrikomprimering. Bland andra förbättringar är domänkodning av intresse. I denna kods viktiga delar av bilden mer exakt än resten bilden.



Allmänt schema av enkoder för JPEG 2000

Transformerande funktion ändras från DCT till 1D Diskret Wavelet Transformation (DWT). Den ursprungliga bilden omvandlas med hjälp av wavelet, kvantiseras och entropikodas. Den största skillnaden mellan DCT och Wavelet-transformation består i det faktum att DWT delar bildrutor till underrutor som delas vidare in i nya underrutor osv.



Ett exempel på en bildnedbrytning med 2D DTW. Märk hur detaljer är indelade i rutor. Varje ruta är formad med DTW transformation i horisontell, vertikal och diagonal riktning. Den lägre rutanivån komponeras in i den högre rutanivån.



Vad är en wavelet? Wavelet är en del av en funktion i form av en vågliknande del av en funktion som till skillnad från sinusfunktionen går från oändlighet till oändlighet som har en fix början, amplituden och ett slut. Beroende på analyserad signal kan den ha anta olika former.

Vad är en wavelettransformation? Wavelettransformationen söker i huvudsak likheter med wavelet i delar av analyserad signal. Om wavelet har en början och ett slut, då kan vi "expandera den till vilken grad som helst". Om vi använder väldefinierade steg, får vi ett multilevel wavletspektrum.

DTW transformationen ger två stickprover vilka fås genom filtrering via nedre eller övre passagetypen. För en framgångsrik signalrekonstruktion behöver man endast ha stickprov från det höga passagefiltret samt varje upplösningens nivå. De föreställer detaljer som vi måste lägga till för en uppskattning av signalen på lägre nivå för att kunna rekonstruera en högre nivå.

GIF

Formatet **GIF** (*Graphics Interchange Format*) är känd som en populär internetfilformat för stillbilder. Det kom till 1987 och är ett bitmappsformat med stöd för 8-bitars färgpalett, genomsynlighet (transparency) och goda komprimeringssförhållande. På grunden av den begränsade färgpaletten (255 färger) gör den begränsad nytta då hög återgivningstrohet som till exempel i fotografier eftersträvas. Den är dock lämpligt för bilder med ett begränsat antal färger t.ex. logotyper med skarpa-kanter och minimala färgövergångar. Den andra versionen av formatet som kom till 1989 stödjer bättre genomsynligheten.

För stillbildkomprimering använder GIF en Lempel-Ziv-Welch (LZW)-algoritm som tilldelar en bitsekvens från färgordlistan till färger i färgpaletten.



Även om det finns mer avancerade algoritmer idag som t.ex PNG behåller GIF fortfarande sin popularitet på grund av sitt stöd för animering genom att kunna stapla flera bilder ovanpå varandra. Denna egenskap har använts för att skapa verklighetstroga 24-bitarsfärgbilder och animeringar genom att placera tre 8-

bitars rutor ovanpå varandra, där varje bildruta innehåller en del av 24-bitarsfärgpalett.

PNG

Bitmappsbildformat *Portable Network Graphics* (bärbar nätverksgrafik) kom till för att ersätta GIF-formatet som lider tekniska begränsningar samt licenstvång. PNG infördes 1996 och år 2004 blev också den godkänd som ISO/IEC-internationella standard. **PNG** stöder RGB- och RGBA färgutrymme med 8 bitars/färg (24-bitars RGB- eller 32-bitars RGBA).

PNG-formatet är mycket flexibel tack vare sin struktur som liknar en behållare. Bilden skapas "bit för bit" vilket tillåter spridningen av bildinformation och stöder skiktningen och dataflöde.

Förlustfrikomprimering fungerar i två steg:

- Prekompression (förfiltering)
- Kompression

Under den inledande komprimeringen där mängden av bilddata minskas med en metod liknande DPCM där värdet av en pixel lagras som en skillnad mellan den pixel och en pixel till vänster, pixel ovanför den eller ovanför till vänster eller deras kombination. För varje pixellinje är det möjligt att använda ett annorlunda filter. Sedan komprimeras värdena på nytt med DEFLATE-algoritm som tar bort strängar av dubletter och med användning av referenser samt Huffmans schemakodning av de enskilda datablocken som bearbetas var för sig i stället för bearbetning av hela bilden.



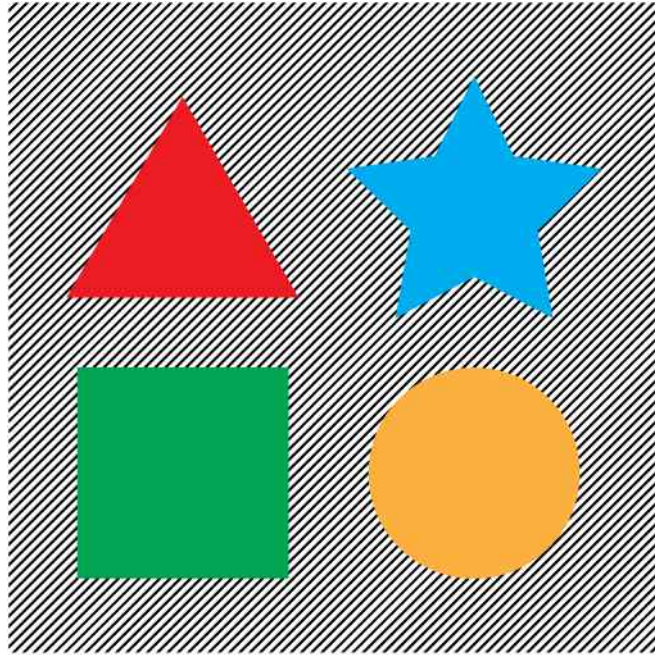
Det första förslaget till formatnamnet var "PING is not GIF" (PING är inte GIF).

—

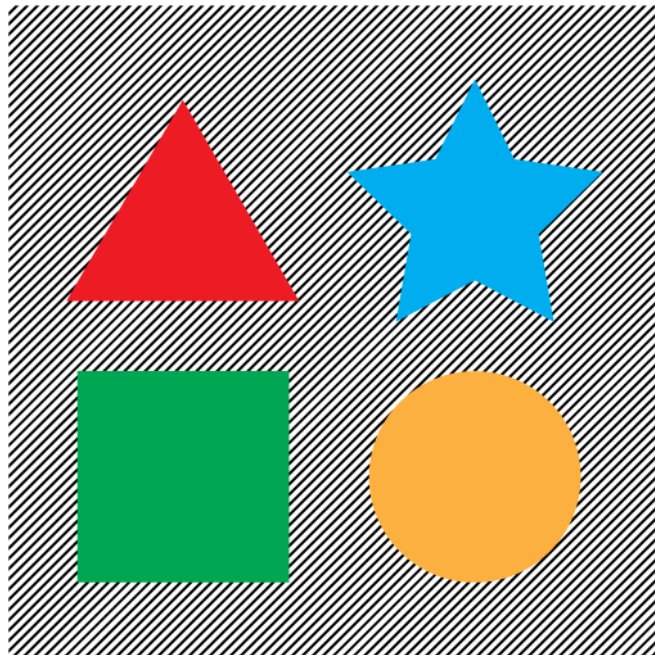
Jämfört med JPEG-formatet skapar PNG större filer med subtila färgövergångar från fotobilder.

+

JPEG har problem med hanteringen av skarpa övergångar och kanter såsom text, linje eller grafik samt med stora domäner med samma färg där det skapar konstigheter. PNG har bättre komprimering varefter det inte uppstår konstigheter vilket gör den idealisk för användning på webben.



Jämförelsen av bilder kodade i JPEG (filen ovan) och PNG (filen nedan). JPEGs bildversion uppvisar synliga brister i kanterna. Storlek på JPEG file är 61 kbit medan PNG file har bara 11 kbit.



WebP

Det yngsta bildformatet WebP lanserades av Googles forskningslaboratorier under 2010. Formatet presenterades som en ny öppen standard som vill konkurrera med det fortfarande populära JPEG-formatet. WebP kombinerar de bästa egenskaperna av JPEG (bra bearbetning av helfärgsgrafik), JPEG 2000

(förlust- och icke-förlustkomprimering), PNG (översikt i förlust- samt även i icke-förlustäge) och GIF (animationsstöd).

Förlustalgoritmen baseras på förfarandet i VP8-videoformatet. Komprimering är baserad på prediktion (förutsägelsen) av tre rutor ovanpå varandra och en till vänster från det analyserade ruta med användning av något av de fyra lägen: horisontella, vertikala, DC (en färg) och TrueMotion. Dålig rutaprediktion eller icke-rutaprediktion komprimeras sedan i 4x4 pixelruta med CDT- eller Walshs-Hadamartsomvandling. Utdata kodas med entropikodning.

Förutom konventionell teknik som ordlistan och Huffmans kodning, använder förlustfrialgoritmen avancerade metoder såsom olika entropikoder för de olika färgkanalerna eller temporärt minne av nyligen använda färger.



WebP i jämförelse med andra bildformat som JPEG och PNG vinner i sin fokus över sina konkurrenter med minst 20%. WebP stöds för närvarande av Linux och Windows med hjälp av plugins (insticksmoduler) har också stöd från sökmotorerna Firefox, Chrome och Opera.

5.4 Komprimering av video

Video eller rörliga bilder är numera dagens synonym för multimedia. Videomängden ökar snabbt och jämfört med andra typer av media tar upp mest utrymme. Med ökande möjligheter för individuella anordningar som används vid skapande och uppspelning av videor ökar mängden av överförda data för varje dag. Det är därför nödvändigt att minska video storleken för att minska kostnaderna för överföring och lagring av data.

Videosekvenser består av separata bilder eller bilder som är i princip möjliga att koda som stillbilder. Emellertid när sekvensen av snabbt föränderliga bilder visar hur samma objekt ändrar positionen i tid vilket innebär att en rad på varandra följande bilder liknar varandra med endast minimala förändringar kan man koda bara dessa förändringar. På detta sättet kan vi uppnå betydligt mer effektiv komprimering.

Videokodning metoden kan delas in i två huvudkategorier:

- Kodning i tidsdomän (vågformadkodning)
- Kodning som bygger modellering

Kodningmetoden i tidsdomän använder transformationskodning med intra-frame prediktionsbildruta som används i många format. **Kodningsmetod som bygger på modellering** används i begränsade bandbreddsapplikationer (med begränsad överföring) såsom i videotelefoni med dataflöde upp till 64 kbit/s. Metoden använder sig av det typiska utseende i scenerna vid videotelefoni med icke-förändringsbar innehåll, minimal rörelse och minskad frekvensen för bildtagningar.

Interbildsprognos och rörelsekompensation

För att minska tidsredundans mellan två bildrutor används interbildsprognos med rörelsekompensation som fungerar i två steg:

- Rörelseuppskattning
- Rörelsekompensation

Under uppskattning av rörelsen skapas en rörelsevektor som beskriver den relativa rörelsen av bildruta från en tidigare bild till den aktuella bildrutan. När positionen är förutsägbar behöver man bara överföra en rörelsevektor. För rörelseuppskattningar används två algoritmer:

- Iterativa (upprepade) algoritmer (pel recursive algorithms)
- Algoritmer med blocksökning (block matching algorithms)

Iterativa algoritmer med försöker att iterativt (upprepande) minimera prediktionsfel. Eftersom dessa är beroende av lokala statistiska avstånd kan de inte användas för uppskattningar av större avstånd i och är därmed lämpliga för

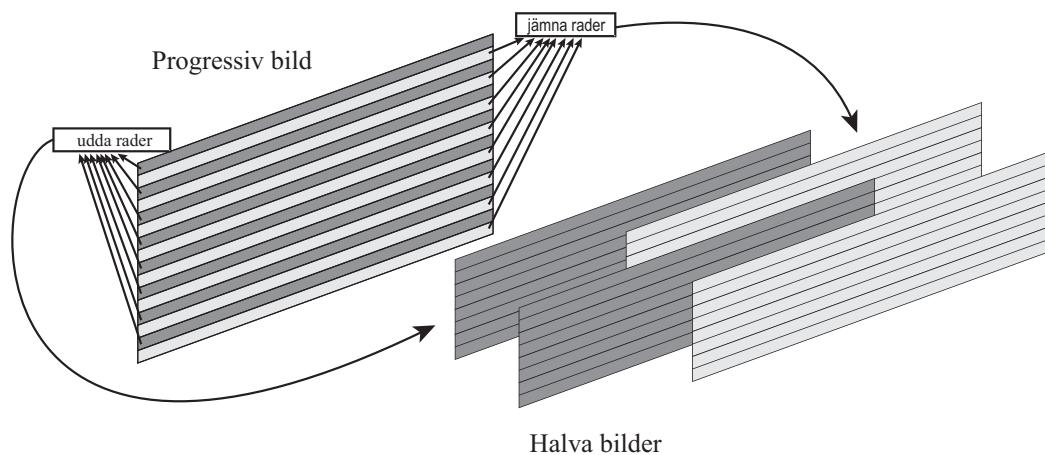
långsamma video såsom t.ex. videotelefon. Algoritmer med bildsökning antar att alla delar av en bildruta är på väg i samma riktning. Den aktuella bildrutan indelas i rutor för att sedan hitta i de föregående rutor de närmast matchande bildrutor.

Algoritmer för rörelsekompensation drar nytta av rörelsevektorer för varje bildförflyttning från föregående bildruta till en ny position i den aktuella bildrutan. Detta resulterar i en förväntad bild som kodas och sedan förs över.

Videokomprimeringsteknik

Interlaced (sammanflätad) video

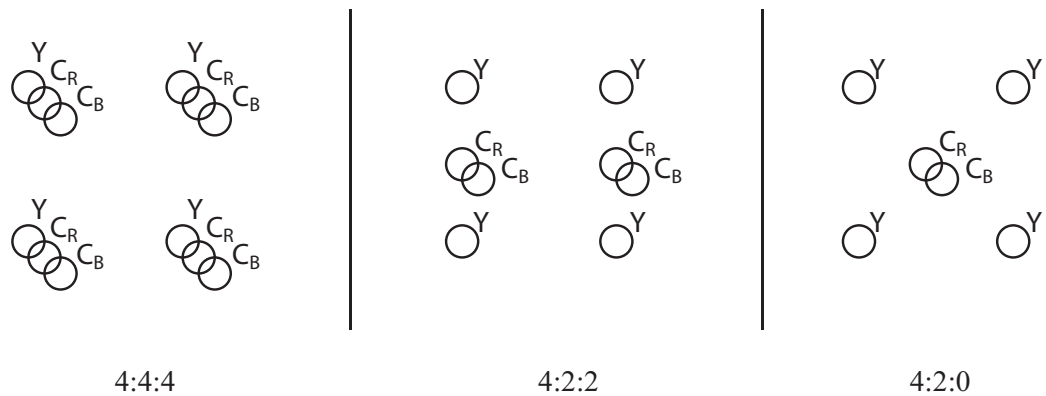
Standard (progressiv eller interlaced) video består av 25 bilder per sekund i Europa) och 29,97 i USA. Interlaced video består av dubbla antalet halvbildrutor per sekund (50 i Europa eller 59,94 i USA). Varje halvbildruta innehåller antingen jämnt eller udda del av varje videobildruta, som avkodas och spelas upp interlaced ordning.



Skillnaden mellan progressiv och interlaced video. Hela bilden är uppdelad i udda (översta) och jämna (nedersta) bildfält. Dessa halvbilder lagras sedan växelvis i en sekvens. Videon får därefter dubbel bildrutehastigheten i sekunden

Färgning av delsampling

En annan kodningsteknik är färgad delsampling. Varje pixel i en videobildruta är kodad med hjälp av YUV-färgrymden som består av tre subpixlar: en för ljusstyrka (luminans) och två för färgen (chroma). En kvadrat med fyra sådana pixlar kallas 4:4:4. Eftersom det mänskliga ögat är mer känsligt för ljusstyrka /luminans/ än för färgförening tillåter detta en minskning av antalet färgpixlar till hälften eller t.o.m. en fjärdedel. I det första fallet, fyra luminanspixlar motsvaras av två färgpixlar (4:2:2) I det senare fallet är det bara två, eller en och en halv färgpixlar (4:2:0 eller 4:1:1).



Exempel på färgning av delsampling. Den vänstar bilden visar att varje luminans subpixel har ett eget par krominans subpixlar. I mitten två luminans subpixlar delar på ett par krominans subpixlar. På den högra bilden finns det för fyra luminans sub-pixlar tillgängliga bara två chromatiska.

På samma sätt som för stillbilder stödjer video en komprimeringsteknik baserad på hybridmetoden vilken kombinerar kodning i tidsdomän med transformationskodning (t.ex. diskretkosinus kodning såsom DCT eller diskret Wavelet- transformation såsom DWT). Prediktionsbildrutor skapade i en process med rörelsekompensation dras från den aktuella bildrutan och på så sätt skapas en felaktig bild. Denna kodas sedan ruta för ruta med DCT, transformationskoefficienter kvantiseras, ordnas med „zig-zag“ uppräknings och kodas med hjälp av variabel längdkod (VLC). Sedan skickas vidare bara VLC-sekvensen.

På mottagare sidan utförs rekonstruktion med ett omvänt förfarande dvs omvänt VLC med sedan omvänt kvantisation och därefter IDCT.

Detta beskrivna förfarandet används med små skillnader i kodstandarderna MPEG och H26x.

MPEG

Expertgruppen för den rörliga bilden (Motion Picture Experts Group: MPEG) har inrättats av organisationerna ISO och IEC för att etablera standard för video- och ljudkomprimering. Detta resulterade i följande olika standard som har utformats med fokus på olika applikationer:

- MPEG-1
- MPEG-2
- MPEG-4

MPEG-1 fokuseras på interaktiva mediasystem baserade på CD-ROM media. MPEG-2 utökar funktionerna i MPEG-1 för att gälla även för digital-TV och högupplösnings-TV (HDTV). MPEG-4 är inriktad på multimediaapplikationer med mycket låga bitflöden.

MPEG-1

Standarden utvecklades för att koda videosignaler med tillräcklig kvalitet vid bithastigheter 1,4 Mbit/s. MPEG-1 stöder snabbspolning, bläddring framåt och att pausa bilden. Även om videobildrutan har en typisk storlek på 352x288 pixlar (CIF) stöder kodek bildrutestorlekar på upp till 720x576 pixlar med 30 bilder per sekund och överföringshastigheten på 1,86 Mbit/s.

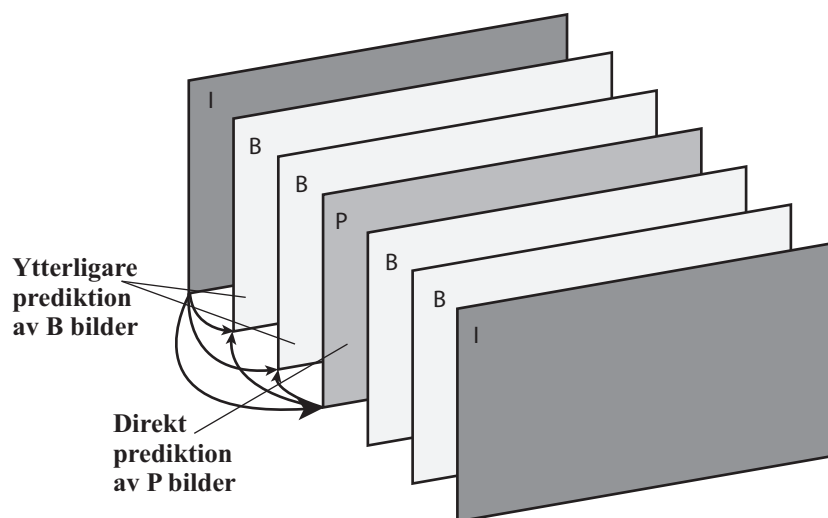
Intra-frame kodning i MPEG-1 baseras på intra-frame prediktion och DCT kodning. Det finns tre typer av makroblock (bestående av fyra luminansuppblock och två krominans block) som företräder de tre typer av bilder:

- **I bilder** – Intra bilrutor – bildrutorkodning
- **P bilder** – Förutsedda bildrutor– direkt prediktionsprognos med intra-frame kodning
- **B bilder** – Dubbelriktade bildrutor – två vägs prediktion/interpolering

“I”-bilder i rutor av 8x8 är DCT-kodade, kvantiserade, zig-zag sorterade och därefter entropikodade. Det görs ingen uppskattning av rörelsen så bilden beter sig som fotobild är oberoende av de andra bilder och fungerar som stopp under snabb spolning bakåt och framåt.

“P”-bilder är kodade med intra-frame prognosen och jämförs med föregående bild. Detta skapar en felaktig bild (representerad genom en rörelsevektor), som delas in i makroblock med storlek 16x16-pixlar. Dessa kodas med DCT, kvantiseras och entropikodas, som ”I” bilderna. Dessa kodade bilder innehåller ännu inte all bildinformation, eftersom de är beroende av de föregående bilderna. De tjänar bara som referensbilder vid prognosen inte för snabb uppspelning.

“B”-bilder erhålls genom att använda endast vanliga och omvända prognoser från “I” eller “P” bilder. “B” bilder används vanligtvis som utfyllning eller att lägga till detaljer i snabba scener mellan “I” och “P” bilder då de innehåller samma information som “P” bilder. Från “B” bilder görs inga prognoser.



Sortering av I-, P- och B-bilder och deras sinsemellan beroende av varandra

Bilddrutor kan kombineras på ett flexibelt sätt i enlighet med programkraven. Sekvensen IBBBBBB-erbjuder utmärkt tillgång till bilder, snabb uppspelning och flytt framåt, men den låga komprimeringen av "I"-bilder ställer krav på överföringen. Därför används vanligtvis, sekvensen IBBPBBPBBPBB(I) (som kallas för Group of pictures - the GOP), och "I"-bilder endast används ungefär två gånger per sekund.

MPEG-2

MPEG-2 är en förlängning från den tidigare standarden (med bakåtkompatibilitet) med möjlighet till en sammanlänkad (interlaced) video, en förbättrad maximal bildupplösning, TV-video kvalitet vid bithastigheter från 4 till 8 Mbit/s och HDTV-video kvalitet vid 20 Mbit/s.

Avkodarens egenskaper erhålls genom profiler och nivåer. Varje nivå definierar den parameteruppsättning som avgör målet för videoapplikationen medan profiler bestämmer komplexiteten i algoritmerna som används. Följande tabeller beskriver i detalj nivåer och profiler.

Översikt över nivåerna i MPEG-2

Nivå	HIGH (hög)	HIGH 1440 (hög 1440)	MAIN (central)	LOW (låg)
Parametrar	1920x1152 pix. 60 sn/s 80 Mbit/s	1440x1152 pix. 60 sn/s 60 Mbit/s	720x576 pix. 30 sn/s 15 Mbit/s	352x288 pix. 30 sn/s 4 Mbit/s

Profilöversikt i MPEG-2

Profil	Algoritmer
High (hög)	Alla funktioner för Spatial Scalable-profilen och mera för läge med 3-layer rumslik skalning och SNR-skalning Färgmodell YUV 4:2:2 för krävande uppgifter
Spatial Scalable (rumslik skalning)	Alla funktioner för profilen SNR Scalable, och mera för kodning i läge 2-layer rumslik skalning Färgmodell YUV 4:2:0
SNR Scalable (SNR skalning)	Alla funktioner för profilen Main, mera för kodning med 2-layer SNR- (signal-to-noise ratio) skalning Färgmodell YUV 4:2:0
Main (central)	Utan skalning, kodning av översatt video Tillfällig tillgång till bilder, prognosläge med B bilder Färgmodell YUV 4:2:0
Simple (enkelt)	Stöder alla funktion för Main-profilen utöver prognosläge med B bilder Färgmodell YUV 4:2:0

Videoskalning

Videoskalningen möjliggör för avkodare (dekodare) att spela upp video med låga överföringshastigheter när dessa inte kan spela upp video med hög överföringshastighet. Avkodaren tar emot video av låg kvalitet och tillägsinformation möjliggör förbättring av videokvaliten. Med hjälp av SNR-skalning är DCT-koefficienter grovtskalade och detta tar fram video med låg överföringshastighet. Skillnaden mellan grov kvantisering och verkliga värden kvantiseras på nytt med hjälp av finare kvantisering och denna information överförs separat. Detta gör att kvaliteten hos „video-on-demand“ förbättras. Vid en rumslig skalning är video först kodad med lägre upplösningar och högre upplösning åstadkoms med hjälp av tillägsdata. När avkodningsenheten inte stöder högre upplösning utelämnas tillägsdata och enheten avkodar endast låguppläsningvideo. Tidsskalning fungerar på samma sätt.

Detta skapar en video med ett lågt antal bilder och med hjälp av kompletterande data möjliggör video återuppbyggnad med ett högre antal bildrutor. Rumslig och tidsmässig skalning kan kombineras, vilket leder variabiliteten i video koding, t.ex. vid stöd för HDTV och TV i standardupplösning.



Organisationer ISO och ITU-T skapade tillsammans MPEG-2. ITU-T gav denna standard betäckningen H. 262.

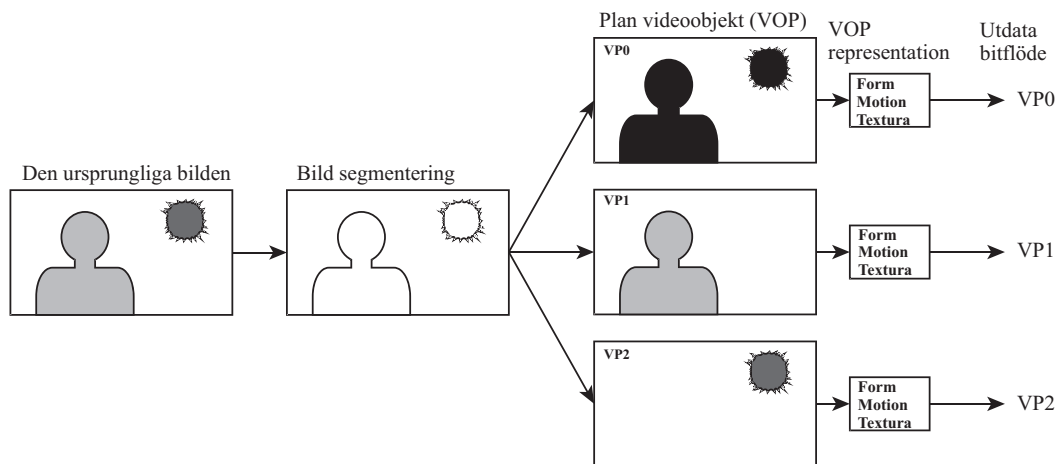
MPEG-4

Denna standard har utvecklats till stöd för mycket låg upp till 64 kbit/s överföringshastighet. Dess syfte är att stödja video på Internet, på mobila enheter och nätverk samt främja interaktivitet med objekt på scenen. Detta krävde förbättringarna av komprimeringsmetoder som drar nytta av videokodning i vanliga standardvideo och syntetiska (objektmodeller) video.

MPEG-4 finns i två versioner. Den första av kallas Del 2 och används av många kodeks inklusive DivX, XviD, Nero Digital m.fl. Den andra versionen kallas Del 10 eller också MPEG-4 AVC/ H.264 Advanced Video Coding och används i x264 kodeks Quicktime och i HD-videomedia som Blu-ray Disc.

Kodning av vanliga video görs genom kontakt och kodning av videoobjekplan (video object plan-VOP). Varje VOP innehåller information om strukturen och textkonsistens av objekt på scenen.

VOP-sekvens representerar samma objekt och kallas videoobjekt (VO). Varje video objekt kan kodas med ett olika dataflöden som tillåter flexibel tilldelning av bitflöde och andra uppgifter om objektet såsom skalning, rotation, förändringar av ljusstyrka och färg.



Exempel på användning av videoobjekt

Videoobjekt bestämmer sin form av en binär mask eller av en mask i nyanser av grått. Kodning av rörelsen bygger på liknande principer som används i de tidigare MPEG-standarder men tillämpas på videoplanobjektet vilket skapar IVOP-, PVOP- och BVOP- bilder. Rumslig redundans tas bort med DCT och temporal redundans med rörelsekomensationen. Textkonsistens på videoobjekter kodas på nytt med DCT-modifikationen kallad Shape Adaptive (SA) DCT. Som alternativt kan användas kodning med hjälp av SA-DWT (Diskret Wavelet Omvandling).

Syntetisk video möjliggör skapande av konstgjorda objekt för att sedan klistra in dem mellan befintliga videoobjekt på scenen. Huvudsälet till detta är att aktivera animering av ansiktet i multimedia applikationer

Standarder H.261 och H.263

H.261-standarden kom ut 1990 för att användas i videotelefoni och videokonferenser med låga överföringshastigheter från 64 kbit/s och upp till 1920 kbit/s med en liten fördröjning. H-261 förenar de olika TV-standarderna med olika antal linjer och halvratofrekvenser (PAL och SECAM med 625 linjer med 50 Hz, 60 Hz NTSC 525 linjer). Kodek använder CIF-upplösning (352x288) och QCIF (176 x 144) där en resolution används för videokonferenser med flera deltagare och den andra för videotelefoni där överförs oftast endast ansikte och axlarna av personen.

CIF a QCIF upplösningar består av grupper av rutor (GOB): CIF från GOB 1-12 och QCIF från GOB 1, 3 och 5. Varje GOB består av 33 makroblock bestående av 6 block: 4 för ljusstyrka och två för färg (kroma- C_R , C_B), vardera bestående av 8x8 pixlar.

H.261-kodek använder bara "I"-bilder (märkta som referensbilder) och "P"-bilder som erhålls med hjälp av prediktion av rörelse från "I"-bilder eller föregående "P"-bilder. Standarden använder inte "B"-bilder.

Kodningsalgoritmen använder en hybrid blockkodning med mellanbildsprogno, rörelse ersättning och transform kodning baserat på DCT som liknar MPEG-1-

kodning. Efter avlägsnande av den rumsliga och tidsmässiga redundansen förvandlas varje ruta med DCT, kvantiseras, sorteras med zig-zag räkning till vektor och kodas med Huffmans förlustfrikodning. Utöver detta används också iterativfilter som jämnar ut skillnader mellan rutor av bildprediktion vilket förbättrar mellanbildsprognos.

Iterativ filter fungerar med en sekvens av bilder och tar bort bildrutans artefakter som härrör från DCT-transformeringarna av varje ruta. Dess roll är att jämna ut skarpa kanter mellan bildrutor. Utjämningsförelkommer flera gånger i en loop tills den når den önskade gränsen. Även om bearbetning i slingan kan ta mer tid än avkodningen själv kan det vara möjligt efter uppskattningsprocessen av rörelsen att koda en mindre rörelsevektor.

H. 263-standard ger effektivare kodning jämfört med H.261. Tack vare användningen av vissa metoder från MPEG-1 har kodek sänkt bithastighet med omkring 50% samtidigt som upprätthåller samma subjektiva kvalitet. Jämfört med H.261 ger H.263-standard ett bredare stöd för videoformat (SQCIF, 4CIF och 16CIF), en bättre uppskattning av rörelsedetektorn, en modifierad -VLC kodning samt den introducerar PB bilder.

Rörelseuppskattningen i H.263 fungerar med halvpixel (halv-pel) förutsägelse. Medan rörelsevektorer i H.261 arbetar med heltal, är vektorer i H.263 representerade med en noggrannhet av 0,5. Dessutom beräknas rörelsevektorn för makroblock genom att jämföras med rörelsevektorer i angränsande makroblock (genom beräkning av medianen): Därefter överförs endast skillnaden mellan den beräknade och den faktiska rörelsevektor (förfarandet kallas för en median förutsägelse).

Rutiner för PB bildruter arbetar på liknande sätt som i MPEG-1 kodek. "P"-bildruta erhålls från "I" -bildruta eller från "P"- bildruta. "B"-bildruta erhålls från en två-vägs prediktion från närliggande bildruter. Skillnaden mellan MPEG-1 och H.263 är att "B"-bildruta i H.263 lagras direkt i "P"-bildruta vilket skapar en PB-bildruta. Detta är ett fynd särskilt för video med låg bitflöde /bitrate/.



En utveckling av H.263 till H.263+ standard ger mer motståndskraft gentemot överföringsfel, dynamisk scenupplösning upplösning och bildskalning.

MPEG-4 AVC/H.264

Senaste version av MPEG-4 känd under namnet avancerad videokodning (Advanced Video Coding, AVC) är den mest använda standard idag. Den hanteras gemensamt av ISO och ITU-T och är särskilt lämplig för videokompression i high-definition läget.

Till skillnad från tidigare versioner från dessa två organisationer ger MPEG-4 en hel del förbättringar såsom högre upplösning av färginformation en skalbar videokodning och flerperspektiv videokodning som gör att man kan koda video i flera vinklar, och därmed få (3D)-stereoskopisk video.

Från en enda makroblock kan härledas flera vektorer gentemot olika referensbilder. Algoritmen för rörelse kompenstation fungerar med en

fjärdedelspixel noggrannhet (jämfört med halvpixel noggrannhet i H. 263) vilket ger högre precision för rörelsevektorer. DCT-transformation har förbättrats och justeras för att ge entydig avkodning. De släta domäner kan dessutom använda en sekundär Hadamardstransformation vilken ytterligare förbättrar komprimeringsförhållandet.

Utöver detta infördes förlustfrikodning av makroblock vilken tillåter en perfekt representation av de markerade domäner i bilden. Standarden fungerar i två lägen: PCM-makroblock eller i en förbättrad mer effektiv förlustfri makroblock. Entropikodning använder nya algoritmer för kodning av syntaxelement och kvantiserade värden från omvandlingsskoefficienter: Context-Adaptive Binary Arithmetic Coding och Context-Adaptive Variable-Lenght Coding finns för mer effektivt kodning än i tidigare standarder.

Ett antal andra förbättringar bevarar samma subjektiva videokvalitet som de tidigare standarderna men vid hälften av bitflöde eller ännu mindre vilket är uppenbart särskilt vid höga bitflöden och i video med hög upplösning.

Liksom standard MPEG-2, stöder MPEG-4 AVC/H.264 kodningsprofiler för användning i en mängd olika målprogram och nivåer som definierar önskad prestanda för dekodern.

WebM

WebM är en kodek för ljud och video med öppen källkod från Google designad för att användas ihop med HTML5- video. Det är en multimedia container baserat på Matroska och innehåller kodatljud med standarden Ogg Vorbis och kodadvideo med kodek VP8.

VP8-kodek är utvecklad av On2 Technologies och efter att företaget köptes upp av Google 2010 och släpptes den ut med öppen källkod. Även om VP8 användarteknik redan infördes i MPEG och H.26x innehåller den ytterligare förbättringar, som gör det möjligt att upprätthålla en hög subjektiv kvalitet genom att minska dess beräkningskomplexiteten. Några av de kommer att presenteras här.

Kodek använder så kallad konstruerade bildruta referens som fungerar som en referensbild för att kompensera rörelsen av flera bilder. Utseendet på de konstruerade referensbilderna anges inte så dessa ligger kvar till designers urval. Iterativ filtreringsprocess som tar bort en bildrutakonstigheter efter minskad rumsredundans (genom DCT-transformation) kan använda olika nummer för varje bildruta för olika antal bildrutor i en sekvens. Entropikodning använder som mest binär aritmetiskkodning som är individuell anpassad till varje bild.



Förutom de ovan nämnda kodstandard finns det ett antal andra videoformat, till exempel Windows Media Video (WMV) från Microsoft eller Ogg Theora, baserad på den tidigare VP3 kodek från On2 Technologies och ytterligare andra standard som inte täcks av ISO- eller ITU-T organisationer.

6 Multimedia bearbetning Multimedia processing Multimediala bearbetningar

6.1 Talsyntes



Med talsyntes avses skapandet av tal som låter som tal människor använder sig av. Detta görs med hjälp av en maskin känd som synthesizer ofta kallad bara synt.

Det finns flera typer av syntar men målet för var och en av dem är i princip detsamma: att återge ett tal så tydlig som möjligt och i den mest naturliga formen.

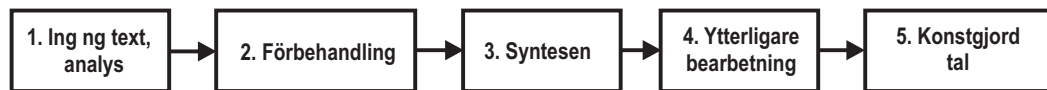
Det finns fyra grundläggande metoder:

- Syntes med segmenter (t.ex. diphone syntes där små i förväg inspelade stickprover av mänskligt språk är förenade så att de skapar meningsfulla ord, se nedan).
- Formantssyntes (hela tal genereras av en dator och inte från mänskliga stickprover. Denna syntes är enkelt att genomföra men låter artificiellt).
- Artikulationssyntes (en strategi som bygger på en modell av den mänskliga vokalorganen och verkliga artikulations processer. Denna syntes är inte allmänt använd).
- HMM-syntes (bygger på en matematisk modell och genererar tal baserad på kriterier av maximal sannolikhet).

På bilden visas en blockdiagram av allmänt förekommande syntar. Naturligtvis är diagrammet förenklat för våra behov och vissa element (till exempel feedback i vissa syntprogram) utelämnas. Dock består i stort sett varje synt av följande delar:

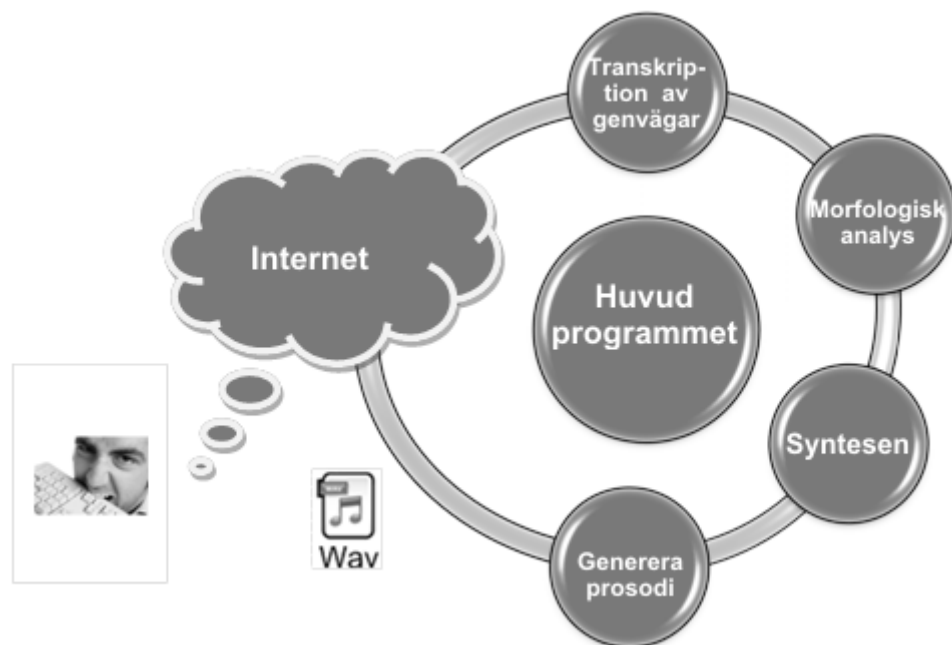
- Textinmatning
- Förbehandling – ord eller fraser är „översatta“ till en speciell form som datorn förstår: nummer och alla förkortningar skrivs över till ord i rätt format. Alla bokstäverna skrivs i ett speciellt alfabetet kallad SAMPA- alfabet. SAMPA (från engl. Speech Assessment Methods Phonetic Alphabet) är fonetiska översättningen som använder endast de utskrivbara ASCII tecken,
- Syntesen
- Efterbearbetning där den slutliga talsyntesen behövs fortfarande göras mer naturligt särskilt genom ändringar av tal takt (prosody modification)

Udata block som innehåller talsyntes



Blockdiagram över allmänna syntar (synthesizers)

För att uppnå det mest naturliga tal måste syntar (synthesizers) utföra komplexa uppgifter som förbehandling och efterbehandling (postprocessing). För att åstadkomma det mest perfekta talet bör system vara anpassningsbara och ha förmågan att lära sig. Ett sådant system skulle bestå av fyra grundläggande moduler: fonetisk översättning av ord, ange ordklass (huvudsakligen för germanska och slaviska språk som använder böjningsformer), fonetisk översättning av förkortningar och slutligen modul för ändringar av tal takt (prosodimodifikation).



Modulär arkitektur av en synt (synthesizer)

I följande exempel kommer vi att fokusera på synteser.

Diphone (difftong) talsyntesexempel

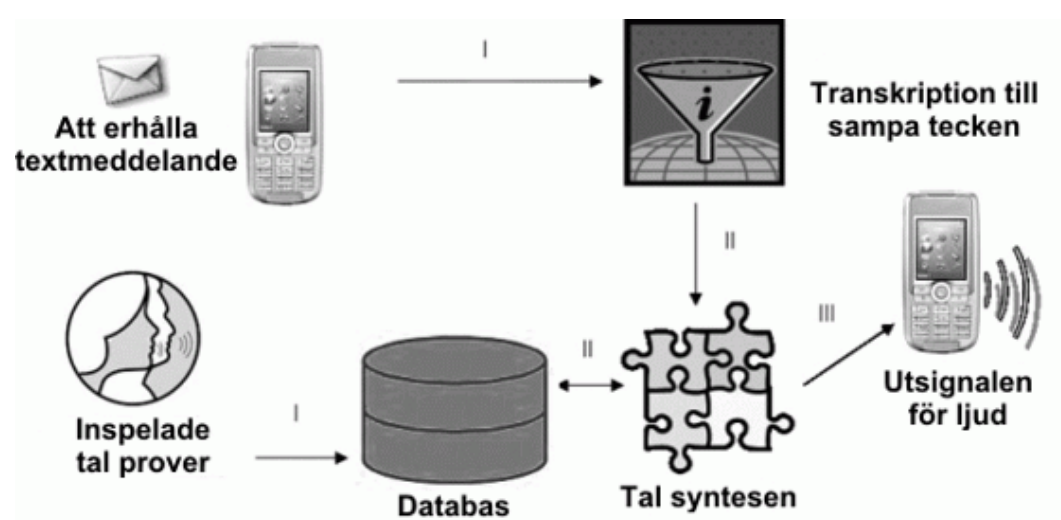
Följande exempel beskriver hur man använder difftongsynt (syntesizer). Den största fördelen med denna lösning är ett naturligt klingande röst och en liten databas. Slovakiska språket har t.ex. endast 1550 difftonger och detta gör användningen praktiskt (särskilt i jämförelse med andra typer av syntes, där databasstorleken är mycket större, som t.ex. vid korpussyntes).

En difftong är tillsammans med ett fonem (phoneme) de grundläggande enheterna i ett tal. Fonem är den minsta omfattande enheten av ljudelement i språket. En difftong består av två på varandra följande fonem. Gränserna för difftong går

mellan två fonem så dess längd i tidsdomän är ungefär densamma jämfört med fonem. Fördelen med diftonger jämfört med fonem, är att de bättre följer övergångar vid ljudändringar mellan ljud och har gränser i deras mitt där tidsförlopp är mest stabilt.

Allmänt är antalet diftonger lika med kvadraten på antal fonem i ett språk. Således får vi alla kombinationer. Faktum är att antalet diftonger är mindre eftersom vi inte kan uttala alla kombinationer eller därför att de inte finns i det aktuella språket. Det faktiska antalet diftonger finner vi bara med efter en grundlig studie av språket. Diftongsdatabas innehåller riktiga inspelningar av tal, tal som delas upp i mindre diftongdelar. Det finns flera alternativ hur man skapar och laddar upp en databas. Man väljer ord som innehåller alla diftonger från en ordlistan. Orden i ordlistan behöver inte ha någon mening eller kopplingen, betoningen ligger på det minsta möjliga antalet av inspelningar.

På bilden visas en design av en talsyntes. Principen för syntesen visas mycket förenklad.



Design av talsyntes

Inmatat text måste syntetiseras till tal. Först måste man skriva texten i SAMPA-alfabet. Alltså är första steget i syntesen att alla tecken översätts till SAMPA. I andra steget är den skriven i SAMPA på nytt och då enligt reglerna för uttal i det specifika språk (i vårt fall slovakiska). Sedan väljs för de motsvarande diftonger ut inspelningar från en databas och dessa kopplas ihop. Resultaten är en syntetiserad text.



Här är några exempel på användning av syntetiserad tal i praktiken: personliga assistenter, assistenter i mobiltelefoner för blinda, trafikgrafikon och navigering system, multimediala internetjänster, ett program för dokumentationen av incidenter eller "fröken ur"-tjänst. Under senare år har det kommit ut syntetiserade ordböcker på online (internet) som har blivit snabbt populära. Det finns också bokläsare med implementerad textsyntes i synnerhet för det engelska språket.

6.2 Bildigenkänning

Ett typiskt problem med bildigenkänningen är identifiering av ett specifikt objekt som man letar efter i bilden resp en viss egenskap som borde finnas där. Problemet är för närvarande löst vad gäller specifika objekt (t.ex. ansiktsigenkänning) men inte för allmänna objekt (t.ex. en lista över alla saker som finns på bilden). Av denna anledning finns det flera domäner av identifieringar och upptäckande såsom :



- Streckkoder – streckkod består av svarta linjer av olika tjocklek som representerar siffror. Bredden på dessa är relativ därför ska streckkoden vara skannade från olika avstånd och i olika storlekar.

2D-koder - även kända som QR-koder eller matriskoder är i princip en förlängning av streckkoder med en dimension. 2D-koder kan lagra mycket mer information än enkla streckkoder inklusive webbadresser (URL) texter eller siffror. De kan variera i storlek och täthet och har består av RS-koder som har förmågan att upptäcka och korrigera fel.

Optisk teckenläsning (Optical Character Recognition OCR) - baseras på igenkännande av tecken (bokstäver och siffror), tryckta bilder eller maskinskriven text. Denna process är i princip det specifika exemplet med mönsterigenkänning.

- Fingerprintsidentifiering — ett slags algoritm för att känna igen ett mönster som jämför grundläggande kännetecken för avtryck (valv, loopar och virvlar) med de ursprungligt sparade på den granskade bilden.
- Specifik detektion – bilden genomsöks för att identifiera de specifika förhållanden såsom rörelsedetektorn i säkerhetskameror.
- Ansiktsigenkänning – bilden skannas för att identifiera de särskilda tecken i ansiktet.
- Objekt erkännande – ett eller flera förutbestämda eller inlärda objekt är automatiskt igenkända. Exempel på detta är program som Nokia Lens eller Google Goggles som automatiskt visar beskrivningen av olika objekt på skärmen i realtid.

Varje igenkännande typ har en särskild egen algoritm som kan vara som såväl enkla som komplexa – t.ex. användning av statistiska metoder eller neurala nätverk. Som ett exempel tillhandahålls nedan en enkel algoritm för ansiktedetektion med punkter för panna och haka.

Igenkänning av ansiktpunkter

Igenkänning av ansiktpunkter är grundad på färgskiftningar i människohud och morfologisk karakteristik av mänsklig huvud.

Algoritmen för huddetektion är används för att skilja mellan de relevanta punkterna i det utförskade ansiktet. Den första punkten för korrekt detektion är att ta bort brus med hjälp av lågpasfilter (low band-pass) filter. Med denna filter undviker man oönskad resp felaktiga identifieringar. Nästa viktiga steg är att hitta kanter. Till detta används så kallad Sobel-filter. Efter dessa nödvändiga förberedelser är bilden förberedd för igenkänning åtföljt av igenkänning av enskilda delar av ansiktet (såsom ögon, näsa och mun, haka,...) baserat på kända egenskaper av dessa punkter.

Hakans igenkänning äger rum i området från underläppen och nedåt. Horisontell skanning slutar ungefär i en tredjedel av bilden eftersom vi utgår från att villkoren för bildinput är uppfyllda (dvs. att hakan kommer inte att vara där mer). Vi söker efter ett område som kännetecknas av relativt platthet med en åtföljande kraftig stegring som anger slutet på hakan.

Identifieringen bygger på en jämförelse av horisontal distans mellan två punkter. Båda punkterna ligger i kanten dvs. i utkanten av profilen (på profil-gränsen). Ypsilon-koordinater i den andra punkten är dock omkring 10 punkter större (10 pixlar – detta är det värde som anges i algoritm och det baseras som ett krav på den ingående bild). Om vågräta avståndet mellan två punkter (dvs. skillnaden på x-koordinater) är större än det förväntade avståndet då punkten är markerad som hakan.

Sökningsområdet för punkt på pannan definieras på bilden som den första tredjedelen i horisontell riktning och som den första fjärdedel i vertikal riktning. Identifiering sker på en bild med markerade kanter (från Sobel-filtret) och på en bilden som är resultatet av huddetektion (Skin detect). Söknings punkten ligger ungefär på det område där huden slutar och håret börjar.

Vid detektion av den punkt som definierar pannan utgår man först från Sobel-filter. Man måste hitta en punkt som ligger vid kanten. Sedan börjar man söka genom bilden. „Skin detect“ i horisontell riktning startar går från den punkt och genomförs i högerriktning.

Färg sammansättning i huden vid punkterna på "Skin detect"-bilden har värdet 0 för svart färg, färgsammansättningen i andra delar av ansiktet har värdet 255 (vitt). Om vi finner den första punkten med ett värde på 255 i vårt sökområde (då hittades den delen av huvudet där håret börjar), finner vi skillnaden mellan x-koordinater. Om den är mindre än den fastställda gränsen, då är detta den punkt vi lettade efter dvs. pannan.



Igenkänning av hakan



Igenkänning av pannan

De spårade punkterana resulterar i en 3D-medell av ett mänskligt huvud som kan bli de-formerat och personifierat.

6.3 Ansiktsanimering

Det finns många metoder för animering och modellering av ett ansikte. Innan man väljer rätt sätt, är det nödvändigt att definiera kraven för en viss animation. Ett exempel på sådana krav för modellering och animering är följande:

- Måste animation vara kontinuerlig?
- Skall prestanda vara optimerad pro PC eller mobiltelefon?
- Måste animation synkroniseras med tal?
- Kvaliteten på animationen skall vara det bästa möjliga och den resulterande animeringen bör vara den som mest liknar verkligheten.



Det finns två grundläggande metoder för ansiktsanimering: tvådimensionella eller tredimensionella, och animationer i realtid eller efterhandberäknade animationer.

Animering i realtid ger användaren en möjlighet att interaktivt ingripa i animationen och då minska på den tid som behövs för beredning av animeringen. Nackdelen med denna metod är sämre bildkvalitet eftersom tiden och prestanda för beredning av bilden är begränsad. Kalkulation som behövs för beräkningen av varje bild får inte vara längre än cirka 0,05/s eftersom behovet ligger på minst 20 bilder per sekund. (20 FPS – frames per second). Kvalitet på 3D-animation är subjektiv bättre än kvaliteten på 2D-animation eftersom den resulterande animationen är mer naturligt.



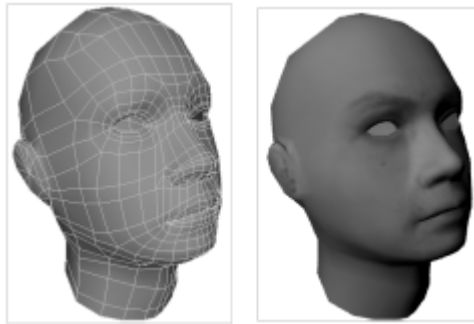
Det finns olika tekniker för modellering av ett mänskligt ansikte i ett fysiskt rum, t.ex. polygonal modellering, modellering med hjälp av parametriska ytor, modellering av mindre ansiktsdelar etc.

Inom ansiktsanimations domän är de mest kända metoder interpolering, parametrisering, musklersimulation etc.

Talande huvud i mobiltelefon

Detta avsnitt bärskiver talsyntes projekt ihop med ansiktsanimation i en mobiltelefon. Syftet med detta projekt var att skapa en multimedial kommunikationsystem som överför endast text och hjälpinformationer för ansiktsanimation och inte video och samtal. Resultatet blev en Java-applikation för mobiltelefoner vilken kunde läsa *korta textmeddelande* (SMS). Efter att man har tagit emot ett SMS visas på skärmen en animerad bild av talande huvud skapad från en bild på avsändaren (hans foto som har lagrats i telefonen vid tidigare tillfälle). Det efterföljande huvudet läser det meddelandet du fått.

I vårt exempel modellen och visemen som behövs för animation finns direkt i objektet (OBJ) vilket innebär att huvudet inte bara har linjer och konturer men också textur (beskrivning).



Neutral model och ansiktstextur (beskrivning)

Visemen är en speciellt deformerad ansiktemodell. Deformation är naturligtvis inte godtyckliga men ansiktet är deformerad så att det ser ut som om ansiktet uttalar ett specifikt fonem. Viseme modell har samma linjer, punkter och schemat som på ett icke-deformerad ansikte. Det som varierar är endast position för knutpunkterna i modellen. På grundval av denna förändring är det relativt enkelt att utföra interpolation av knutpunkterna och orienteringen på den yta som definieras av dessa med små anspråk på prestanda.

Den neutrala ansiktsmodellen avläses från filen (tillsammans med andra modeller och visemes). Interpolation utförs mellan modellerna. Animeringen utförs mellan olika modeller (en modell representerar en person med olika ansiktsuttryck). Animeringen skapas i realtid. Modellen är deformerad på basis av resultaten vid detektion av ansiktpunkter (som beskrivs i föregående kapitel).



Ansiktets visemes

Animation spelas upp samtidigt med syntetisk tal. En talsyntes som används i detta projekt är identisk med diftongsyntes som beskrivs i avsnittet Talsyntes. Det är viktigt att synkronisera animationen och det synkroniserade talet.



Aplikationsexempel på en mobiltelefon: personifierad modell (första) och allmänna modeller (andra och tredje).

6.4 Taligenkänning

Talsignalen produceras av mäniskoorgan för tal och representeras av vågor i luften. Dessa innehåller bland annat lexikala komponent som är utgör nyckel för taligenkänning. Lexikal information är kodad i den akustiska signalen som en sekvens av olika akustiska ljud. Varje språk har sin egen uppsättning av grundläggande ljud som är indelade i fonem. I slovakiska språket är t.ex deras antal 51. Ofta uttalade fonem inverkar på varandra både i tids- och frekvensdomän och skapar då de akustiskt olika ljud. Dessutom uttalas varje fonem olika av olika talare och detta faktum ger en karakteristisk informationen om talaren. Utöver detta existerar i den verkliga miljön additiva och konvolutionella brusljud som ytterligare försämrar situationen. Och för att göra saken ännu värre, varje språk innehåller ett stort ordförråd, vanligtvis flera hundra tusen ord som alla kan ha flera former (böjningsformer, ordtiming, etc.). Det är uppenbart att uppgiften är ganska komplicerad och datorkrävande. Många system har utvecklats så det finns grundläggande klassificering: system med små, medelstora och stora ordförråd, system som är beroende eller oberoende av talare, system baserade på fonem och/eller ord (möjliga är också subfoner eller fraser), system för kontinuerliga texter som för separata ord (diktamen) och så vidare.

I många år gjordes stora ansträngningar för att konstruera och sedan använda ASR (Automated Speech Recognition) som system inom domän såsom tillgång till information, dialog system etc. Först senare med mer utvecklade tekniska funktionerna börjar man skapa komplexa system såsom diktamen system eller automatisk tal transkription. Dessa avancerade system måste kunna fungera i realtid, vara oberoende av talaren, uppnå hög noggrannhet och främja ordböcker, som består av hundratusentals ord.

Metoder för extraktion av tal

Första steg i utformningen av en ASR-system är att avgöra vilken teknik som är bäst att använda sig av för urval av egenskaper. Inledningsvis bör det noteras att uppgiften är ännu inte helt löst och fortfarande görs en del ansträngningar på detta området. Syftet är att simulera det mänskliga hörselsystemet, att göra matematiska beskrivningar, förenkla den praktiska bearbetningen och i förekommande fall, anpassa det till en korrekt och enkel användning med de valda typerna av klassificeringsmetoder.

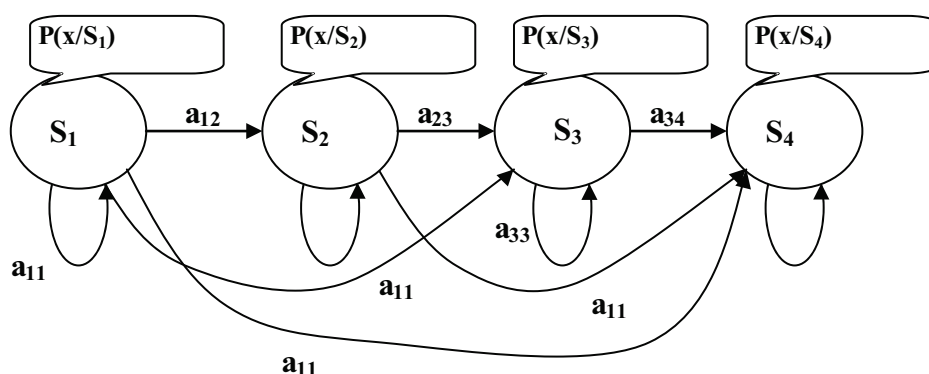
Bra funktion bör vara känsliga för förändringar i ljudet vilka ses som annorlunda hos människor och bör vara "döva" vad gäller de som är betydelselösa för våra hörselsystem. Det konstaterades att följande skillnader är hörbara: skillnader i formanternas position i spektrumet, formanternas varierande bredd samt och att signalintensitet i uppfattas som icke-linjärt.

Funktionerna bör också vara okänsliga för additiva och konvolutionella ljud eller åtminstone kunna hantera dem på ett sådant sätt att dessa störningar skall vara lätta att hitta och föra åt sidan.

Nuförtiden är de mest önskade akustiska egenskaperna MFCC och PLP vilka är utformade för att fånga positioner och snäva bredd som är akustiskt gripbara.

HMM-översikt

Dolda Markov-modeller är statistiskt modelleringsmetod för tal och mer exakt för dess delar (ord, stavelser, fonem, subfonemer, etc.). Den är baserad på konceptet av Markovs kedja som gör det beräkningsmässigt mycket effektivt även om den inte återspeglar tidsutvecklingen av naturliga tal. Således måste varje modell uppskattas med hjälp av vanligtvis mycket stor uppsättning av utbildningsexempel vilka innehåller flera inspelningar av samma ord och dess olika uttal. HMM definieras av a-priori sannolikheter (π) med start i särskild status, övergångsmatris (sannolighet för övergång mellan stånd, a_{ij}) och sannolikhetsfördelningen att generera observationsvektorer i en viss status, $P(x/s_i)$. Dessa sannolikhetsfördelningar är inte kända i förväg men de mest använda är en blandningar av flerdimensionella normala fördelningar (GMM). Fyra ståndsmodell från vänster till höger visas i bilden nedan.



En fyrastånds HMM-modell från vänster till höger

Sannolikheten att observera egenskaper för vektorsträngar i modell λ ges av:

$$P(\mathbf{x}_1 \dots \mathbf{x}_T | \lambda) = \sum_{i=1}^N \alpha_T(i), \text{ kde } \alpha_t(j) = \left[\sum_{i=1}^N \alpha_{t-1}(i) * a_{ij} \right] * P(\mathbf{x}_t / S_j) \quad j = 1, \dots, N$$

$$\alpha_1(j) = \pi_j P(\mathbf{x}_1 / S_j)$$

Därefter skall igenkännande utföras genom att välja HMM-modellen (λ) med den högsta sannolikheten. Den stora fördelen med HMM modeller är förmågan att sammanfoga flera modeller till en sträng så att framställningen av valfri längd kan byggas med hjälp av grundläggande modeller. I praktiska situationen kan bara sökvägen med maximal sannolikhet beräknas, t.ex. $P(x_1, x_2, St1, St2, \dots, x_T, StT)$ och genom tillbakaspårning bestäms sekvensen av en dold status (status kan relateras till en ordföljd) med användning av Viterbi-algoritm.

Förutom träningsstrategin är strukturen av HMM-modell också viktig. För talenhet, används vanligtvis modell från vänster till höger, den mest vanliga är sannolikt Bakisov-struktur. Det är allmänt accepterat att varje akustisk tillstånd är modellerat med tre HMM-status för att fånga början, mitten och slutdelen. Den grundläggande HMM-modellen kan modellera fonem, stavelser, ord eller meningar. Det är en kompromiss mellan riktigheten i modelleringen, datorflexibiliteten och tillgängligheten till utbildningsdata. Vissa oftast korta ord

kan modelleras med bara en enda HMM-modell. Bra balans kan uppnås med länkade triphoniemodell som används av de flesta befintliga system. Oavsett vilken utbildningskriteriet används finns det inga formler för optimala parametrar och då måste många iterationscykler tillämpas på ett kontrollerat sätt. Detta säkrar dock endast sökning av de lokala maxima. På grund av klassificeringsteori är HMM-modeller benägna att vara “overfitted” och därför måste en valideringsset användas för att upptäcka och förhindra detta fenomen.

6.5 Multimodal gränssnitt

Multimodala gränssnitt är mycket populära idag. Alla talar om dem – de är mycket naturliga och populära bland användare. Multimodala gränssnitt erbjuder lösningar på många problem med användargränssnittet samt tillåter också användning av nya tjänster och applikationer.

Multimodala gränssnitt representerar kombination av flera modaliteter eller interaktion med datorsystem. Multimodala gränssnitt används för identifiering av den talande med ansikteform och röst, eller utnyttjar igenkännande och verifierar användarens röst, ansikte etc. Dessutom används multimodala gränssnitt för att styra *set top box* (STB) med hjälp av rösten eller gester.

Ett exempel är projekt HBB-Next som är integrerat med multimodala gränssnitt. Projektets roll går ut på hjälpa med sammanslagning av radio- resp tv-sändningar med internet och utveckla denna tekniken för slutanvändarnas bästa. Projektet är att berika den klassiska TV:n med den sociala aspekten dvs implementera sociala nätverk, underlätta möjligheter att få tillgång till filmer från flera enheter (såsom TV, PC, mobil) och skapa länkar till filmer. Det skall också hjälpa med att blanda sända filmer, videklipp från internet och filmer gjorda av användaren.

HBB-Next är baserat på en modulär arkitektur. Moduler i **HBB** (*Hybrid Broadcast Broadband* – hybrid breddbands - TV) är utformade så att de kan samarbeta med varandra. Till exempel när en användare kommer in i rummet, upptäcker systemet detta och ställer sig automatiskt in i enlighet med dennes krav. När användaren sedan öppnar programmet AppStore låter systemet honom öppna, köpa och installera det valda programmet utan ytterligare autentisering. Varje av dessa åtgärder kunde annars kräva flera säkra användarautentiseringar. Detta är ett exempel på samarbete mellan ansiktsgenkänningsmodulerna som ger autentisering.

Identifiering av flera talande är baserad på det inspelade ljudet som kan innehålla flera personer samtalar. Denna uppgift kan delas in i flera kategorier enligt flera detaljer. I en kategori finns fall där en talande är känd och vi vill ta reda på om hans röst är i den inspelade signalen. Detta är möjligt om han var med vid testfasen på detektor. I händelse av att ingen av talarna är känd, måste man använda segmentering- och klustreringsteknik. Syftet med de flesta program av detta slag är, att de skall arbeta kontinuerligt : "avlyssna " inkommande ljud kodade i PCM-stickprover för att upptäcka röstförekomsten (*voice activity detection*, **VAD**), kunna hantera tystnad och buller i bakgrunden och identifiera överlappande talare. När programmet registrerar en längre röstinspelning måste det identifiera talare med noggrannhet. När en talande skall identifieras är målet att identifiera honom oavsett den semantiska aspekten i hans tal.

Varje system för identifiering av den talande, består av två huvuddelar. Den första är igenkännandet av talparametrar från den inspelade ljudsignalen, den andra är en klassificeringsmetod som har valts på grundval av parametrarna och som med angiven sannolikhet identifierar den talande. Dessa system är oftast utformade för ett särskilt ändamål och är ännu inte universala. Formgivaren till applikationen måste välja med hänsyn till ändamål med arbetsuppgiften rätt kombination av metoder för deras genomförande .

När det gäller igenkännande av talkommandon används ett system för igenkännande av isolerade ord. Den mest lyckade och mest använda är den som arbetar utifrån HMM-principen med hjälp av statistisk talmodellering särskilt den som använder basenhet för modellering av kontextberoende fonem. En bestämd grupp av kommandon tas fram för att sedan användas för modellering av hela ord och därmed nå en potentiellt bättre noggrannhet (bättre fångade koartikulationseffekter).

Allmänt finns det två kategorier för övervakning av en användares gester: betraktelsesätt och förfarande utifrån 3D-modellen. 3D-förfarande är baserat på att 3D-modellen jämför indataparametrar av en lem i en 2D-projektion på en 3D-lemmodell. Förfarandet bygger på bildegenskaperna som används för modellering av det visuella utseendet av lemmen och jämför den med de valda bildegenskaperna från videoingången.

I den andra proceduren beror resultatet på datorenhetens förmågan. Vid användande av RGB-kamera fokuseras metoden på övervakning av hudfärgen eller formen på de gestikulerande delar av kroppen. Förfarandet beror dock främst på ljusförhållanden samt förgrunds- och bakgrundsstabiliteten i referensenheten. Dessutom kan inte i utvärderings domän visas andra objekt med hudfärg eller i form av lemmar. En infraröd djupkamera använder sin egen IR-ljuskälla och är därför mer resistent mot ljusförhållandena på områden. Kameran kan dessutom tillhandahålla en detaljerad karta, en pseudo 3D-bild av områden vilket kan vara mycket användbart i övervakning av de gestikulerade delarna av kroppen till exempel av handen.

För närvarande finns det flera olika metoder hur man genomför ögonkontroll. Den viktigaste punkten är korrekt upptäckt av ögats pupillen. I HBB-Next projekt används den enklaste och mest naturliga metoden när RGB kameran drar nytta av Kinect-enheten. Principen är följande: sittande person framför skärmen utan att flytta på huvudet tittar på markerade punkter på skärmen. Programmet mäter avståndet mellan huvudet och bildskärmen. För ytterligare beräkningsbehov är det nödvändigt att ange storleken på bildskärmen och sedan kan man med hjälp av Pythagoras sats beräkna parametrarna för pupillens rörelse dvs deras vinkel och rörelse. Programmet anger den maximala räckvidden av pupillsrörlighet och denna leder användaren till att titta på skärmkanterna. Därefter beräknas vinklar för den avleda pupillen från dess utgångsplacering. Med hjälp av det kända huvudavståndet från skärmen kan man sedan räkna fram spridningen och vinkeln när pupillens position ändras.

Teknik för ansiktsigenkänning har beskrivits i detalj i kapitel Igenkännande av bild. I ett verkligt system definieras en lista över de kraven vilka ett system borde ha med:

- Systemet måste identifiera användare i rummet utifrån hans ansikte om denne hör till en grupp lokala användare.
- Systemet måste känna igen användare och jämföra hans profil med lagrade profiler utan tillgång till internet.

- Systemet kan i ett rum identifiera användare på grundval av hans ansikte även om användaren är okänd (finns inte bland lokala användare).
- Systemet kan känna igen användare även i mörka miljöer.

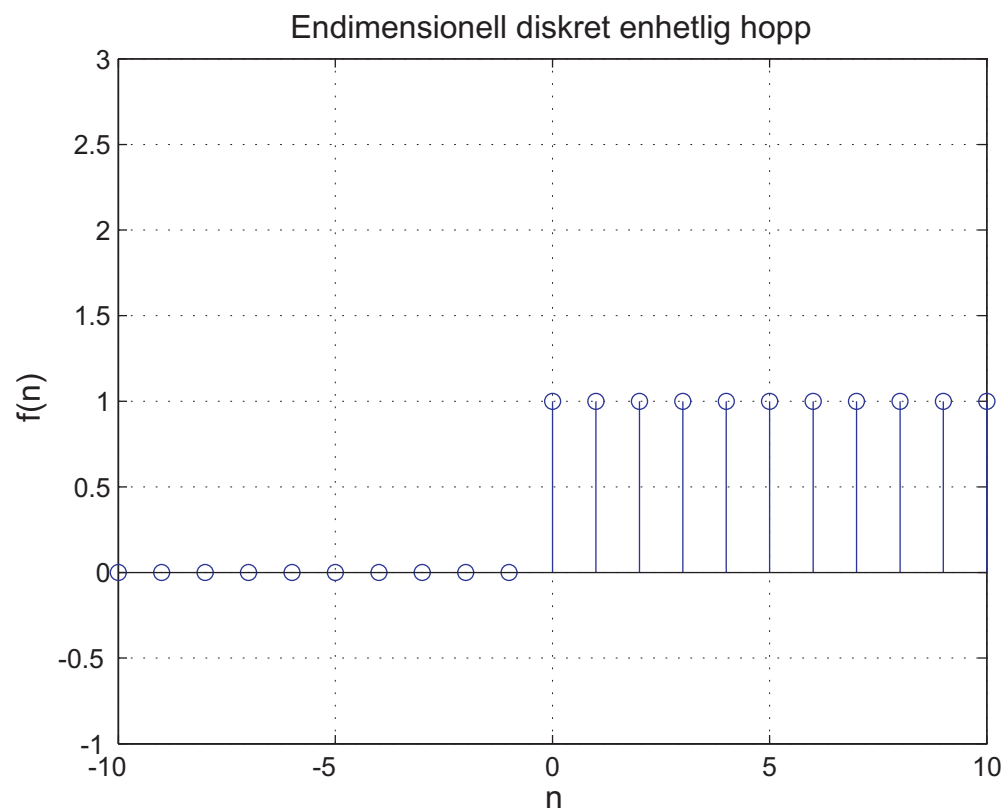
7 Exempel i Matlab

7.1 Exempel i Matlab



Exempel 1

Skriv i Matlab koden som återger endimensionell diskret enhetssteg.



Resultatet: Diskret enhetssteg

```
t=-10:20;      %definition för en signalperiods tid
step=heaviside(t); %heaviside(x) är
                funktion i Matlabu,
                deras värde är 0
                för x < 0, 1 för x > 0,
                och 0.5 för x = 0.

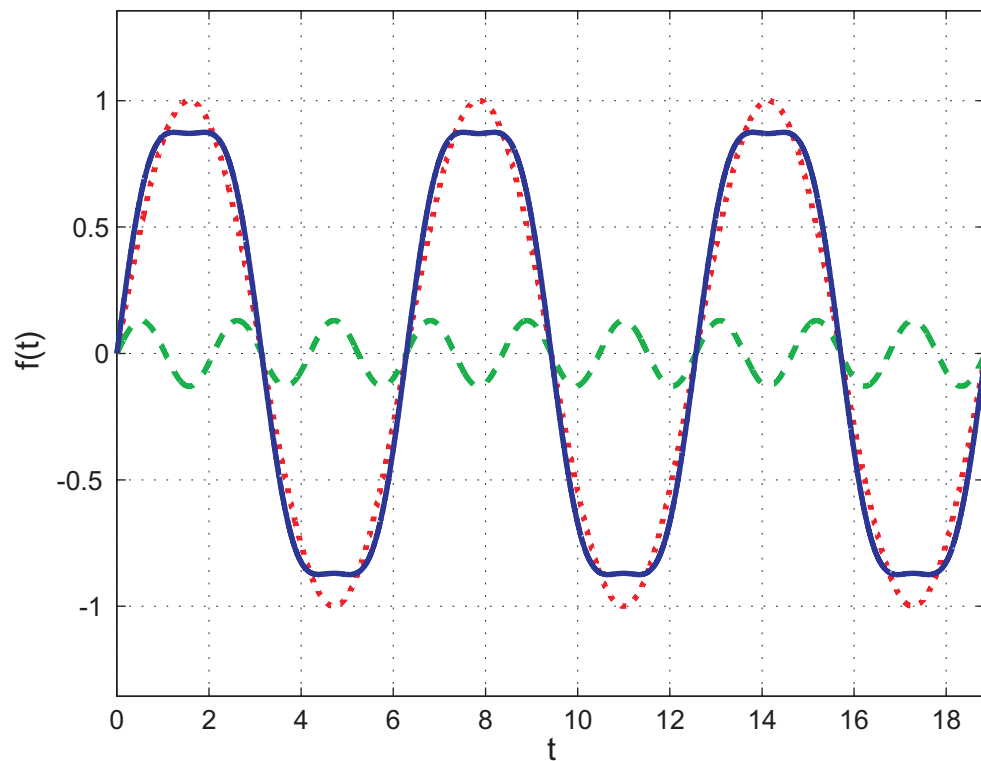
step(t==0)=1;
figure;        %order för bildutskrift
stem(t,step);
grid on;
xlabel('n');
ylabel('f(n)');
title('1D diskret enhetshopp');
axis([-10 10 -1 3]);
```



Exempel 2

Skriv i Matlab, koden som skapar en sinusformad signal.

Demoduleringen av antagen bruss signal och jämförelse med original



Resultatet: Två harmoniska signaler och deras summa

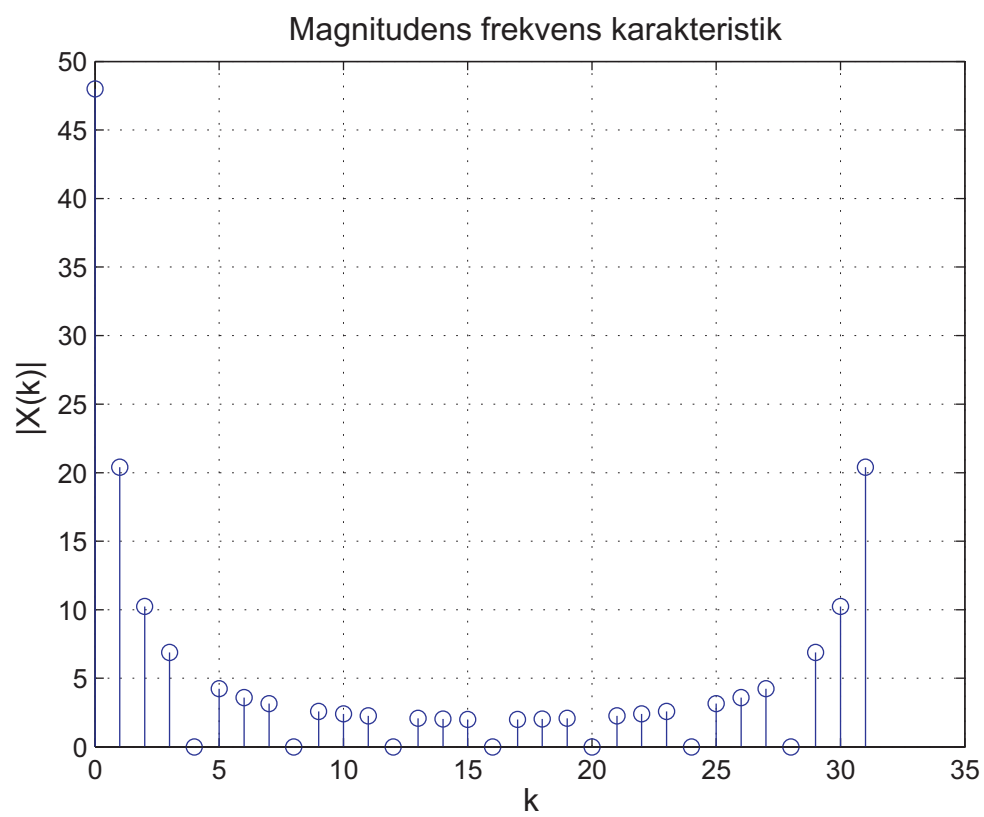
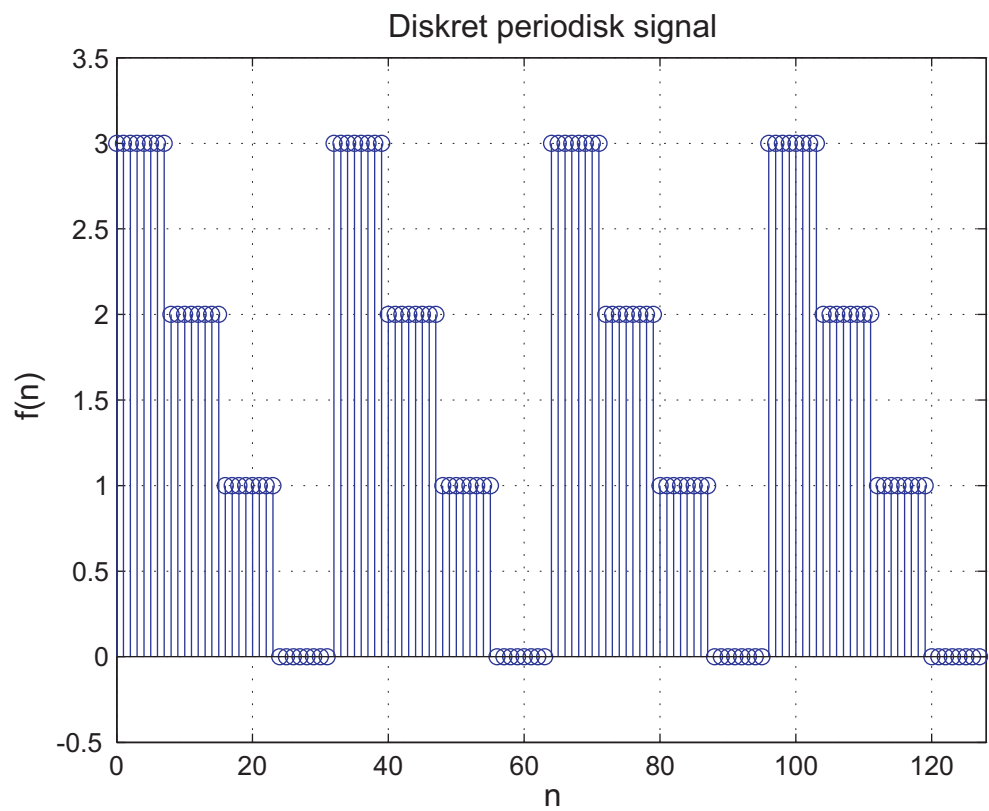
```
range=6*pi;           %tidsutsträckt signal
t=0:0.001:range;      %tidsenheter
A=[1 0.13];           %amplitudvektor
w=[1 3];               %frekvensvektor [Hz]
phi=[0 0];             %fasvektor
sig1=A(1)*sin(w(1)*t+phi(1)); % definition av
                        %varje
                        %signal

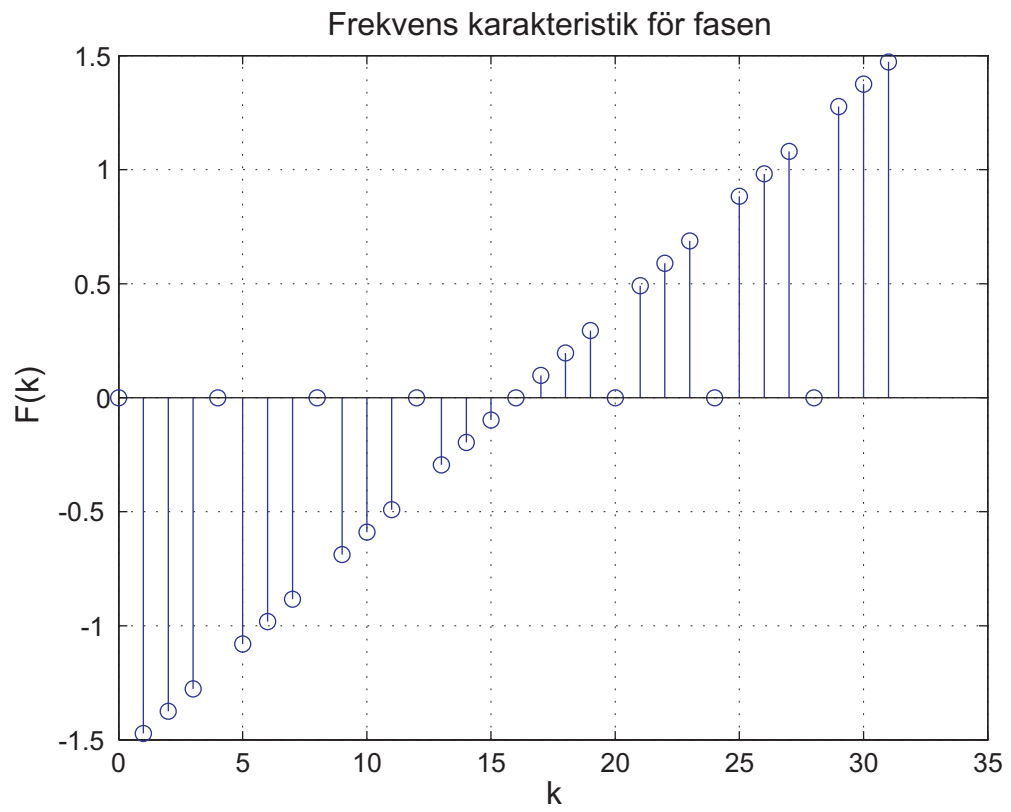
sig2=A(2)*sin(w(2)*t+phi(2));
signal=sig1+sig2;
figure;
plot(t,sig1,':r','LineWidth',2);
hold on;
plot(t,sig2,'--g','LineWidth',2);
hold on;
plot(t,signal,'LineWidth',2);
grid on;
axis([0 rozsah -1.2*sum(A) 1.2*sum(A)]);
xlabel('t [s]');
ylabel('f(t)');
title(Harmoniska signal a deras ihopslagning);
```



Exempel 3

Skriv i Matlab koden för att bestämma diskret Fourierstransformation och frekvenssegenskaper.





Ett exempel på överföringsfunktionen i LDKI-system, magnitud- och fasfrekvenssegenskaper
