KTH Datavetenskap
och kommunikation

2F1120 Spektrala transformer för Media
Tentamen 060602

Tentamen består av fem uppgifter där varje uppgift maximalt ger 4 p. Normalt gäller följande betygsgränser: **3:10 p, 4: 14 p, 5: 18 p**
Tillåtna hjälpmedel: räknare, formelblad

1

Stina ska göra en ny ringsignal till sin mobiltelefon. Ljudet som Stina vill använda, en sekvens ur en Finsk monsterhårdrocksschlager, är inspelat med 48 kHz. Stinas telefon klarar dock bara ljudfiler med samplingsfrekvensen 8 kHz. Efter lite funderande skriver Stina följande rader i matlab:

```
X=wavread('halellujah.wav')
for n=1:length(X)/6
    Y(n)=X(n*6);
end
wavwrite(Y,8000,'halellujah_8kHz.wav');
```

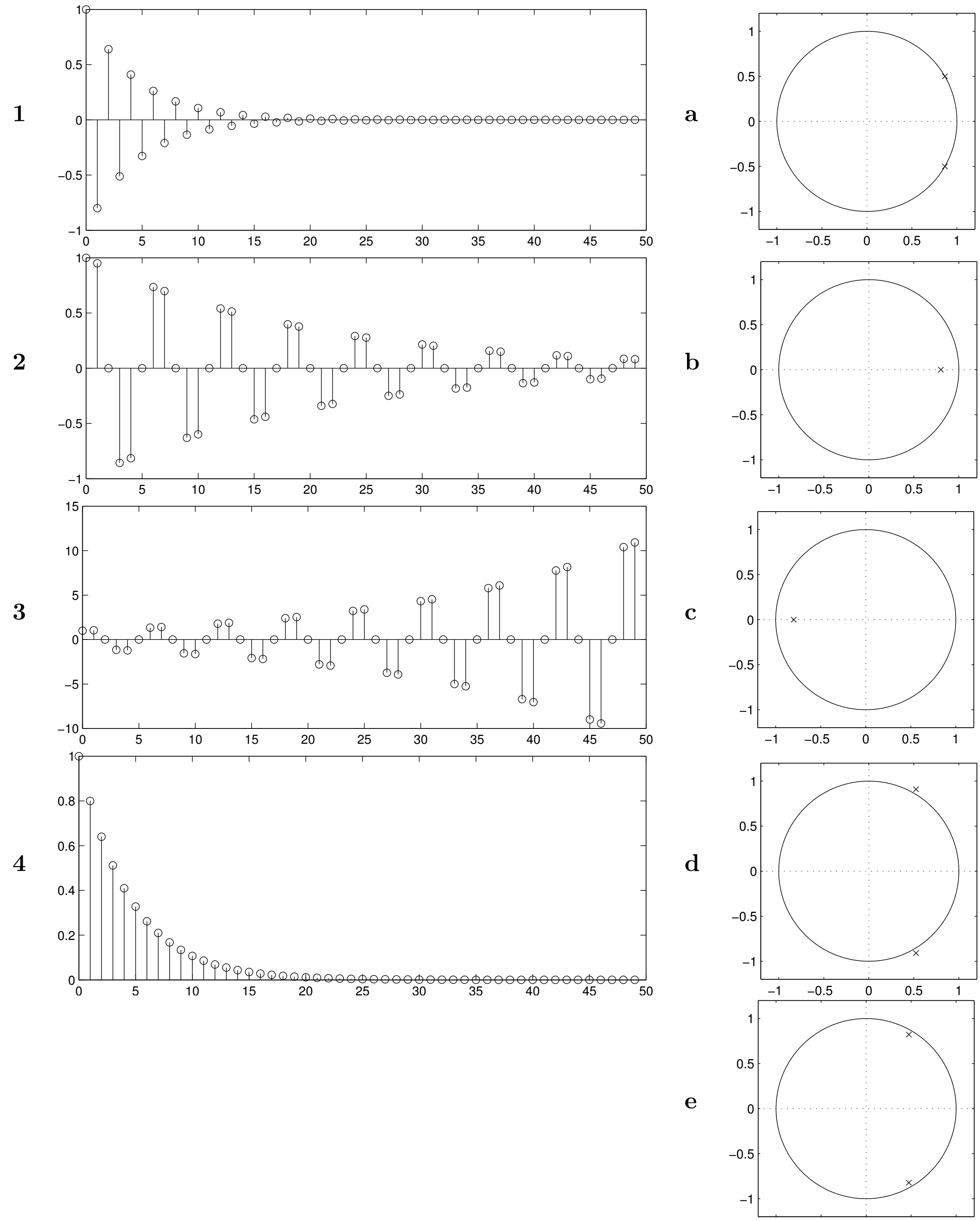
Hon provlyssnar det nya ljudet och det låter fruktansvärt - inte alls som hon tänkt sig...

a Förklara för Stina vad det troliga problemet är, och exemplifiera genom att beskriva vad som händer med en ton från en distorderad elgitarr som har frekvensen 3500 Hz, med hjälp av en frekvensgraderad spektrumskiss av signalen före och efter nedsamplingen! Tonen kan approximeras med en fyrkantvåg, som är bandbegränsad till 12 kHz. *(2p)*

b Beskriv vilken operation hon behöver lägga till i sitt program för att få ett bra resultat, och gör motsvarande frekvensgraderade spektrumskisser före och efter operationen i fråga. Ge även förslag på hur detta kan åstadkommas i matlab (i ord, ej kod). *(2p)*

2

I figur 1 ser du impulssvaren från ett antal från tvåpoliga resonatorer. Du ser även ett antal pol-plottar som beskriver dessa resonatorer. Para ihop varje impulssvar med rätt pol-plot! Observera att det blir en polplott över. Motivera kort varje par. *(1p/ korrekt par)*



Figur 1.

3

I ett digitalt kommunikationssystem har det uppkommit ett eko som är en fördröjd version av den ursprungliga signalen, så att den resulterande signalen beskrivs av $y(n) = x(n) + 0.5x(n-3)$

a Betrakta den resulterande signalen (original + eko) som utsignalen från ett filter, och beskriv detta filters överföringsfunktion $H(z)$ och skissa $|H(\omega)|$ för $0 \leq \omega \leq \pi$. *(2p)*

b Konstruera ett filter $G(z)$ som släcker ut ekot, och visa att det fungerar genom att filtrera en impuls-signal genom de båda filtren kaskadkopplade - resultatet ska bli en impuls. *(2p)*

4

a Beskriv principen för transformbaserad bildkodning, typ JPEG. Hur fungerar det, och varför? *(3p)*

b Förklara med utgångspunkt i föregående svar varför foton lämpar sig bättre för JPEG-komprimering än t.ex. datorgenererade linjeritningar och text. *(1p)*

5

Betrakta ett sk. rullande medelvärdesfilter av längden p, som beskrivs av filterekvationen

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

Det visar sig att detta återkopplingsfria filter kan implementeras på ett mycket mer effektivt sätt om man tar hjälp av återkoppling:

$$y(n) = y(n-1) + x(n) - x(n-p-1)$$

a Visa att de två implementationerna är likvärdiga, och beräkna överföringsfunktionen $H(z)$ för den icke-återkopplade respektive återkopplade formen av filtret. *(2p)*

b Förklara i ord hur den återkopplade lösningen fungerar! *(2p)*

Lösningar

1

a

Problemet är att det uppkommer vikningsdistorsion, eftersom alla frekvenser ovanför halva samplingsfrekvensen 4000 Hz kommer att vikas. Elgitarrtonen kan skrivas som en fourier-serie. Fourierseriekoefficienterna för en fyrkantvåg (se formelblad) är $c_k = \frac{-2j}{k\pi}$ för udda k, och 0 för jämna k, vi får deltoner vid f_0, $3f_0$, $5f_0$ osv.

$f_0 = 3500$ kHz innebär två komponenter under 12 kHz: 3500 och 10500. Den första ligger under $f_s/2$ och samplas ok, men den andra kommer vikas ned till 2500.

b

Hon bör lägga in ett anti-vikningsfilter med en stoppfrekvens strax under $f_s/2 = 4000$ Hz innan nedsamplingen, för att säkerställa att inga komponenter över $f_s/2$ finns med i signalen. Efter filtrering kommer bara komponenten vid 3500 Hz att finnas kvar.

Detta filter kan implementeras på en mängd sätt. Enklast kanske genom att använda något av matlabs inbyggda filterberäkningsprogram (t.ex. `butter()`) för att få fram filterkoefficienter, och sedan filtrera ljudvektorn med hjälp av (`filter()`).

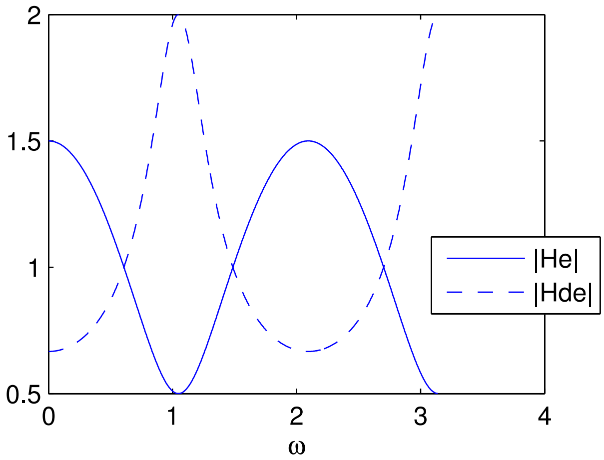
2

1 - c	vartannat sampel pos/neg betyder att frekvensen, dvs polvinkeln, är π.
2 - e	Dämpad svängning med perioden 6 sampel, dvs polvinkel $2\pi/6 = \pi/3$. Detta stämmer för d och e. Dämpningen innebär dock att polradien är mindre än 1, dvs. e.
3 - d	Samma frekvens som ovan, $2\pi/6 = \pi/3$. Men framförallt är systemet instabilt, dvs polradie > 1, dvs d.
4 - b	Ingen svängning, endast dämpning, dvs polerna ligger på positiva reella axeln < 1, dvs b.

3

a

Filterekvationen ger direkt att $H(z) = 1 + 0.5z^{-3}$. $|H(\omega)| = |1 + 0.5e^{-j3\omega}|$. Funktionen uppnår sitt maximum 1.5 då termen $e^{-j3\omega}$ är 1, dvs $-j3\omega = jN2\pi$, dvs $\omega = 0\&2\pi/3$. Minimum inträder då exponetialtemren är -1, dvs $-j3\omega = j(\pi + N2\pi)$, dvs $\omega = \pi/3\&\pi$, se figur (heldragen linje)



b

Inversen av $H(z)$ ger

$$G(z) = 1/H(z) = 1/1 + 0.5z^{-3}$$

Detta ger i sin tur filterekvationen för inversfiltret

$$w(n) = y(n) - 0.5w(n-3)$$

där $y(n)$ är utsignalen från första filtret, och $w(n)$ är utsignalen från inversfiltret. Låt $x(n)$ vara en impulssekvens [1 0 0 0 0 0 ...] och sätt in i filterekvationerna:

n	$x(n)$	$y(n)$	$w(n)$
0	1	1+0	1-0
1	0	0+0	0-0
2	0	0+0	0-0
3	0	0+.5	.5-.5
4	0	0+0	0-0
5	0	0+0	0-0
⋮	⋮	⋮	⋮

$w(n)$ blir en impuls.

4

a

JPEG-kodningen innebär att bilden först delas in i block, om 8x8 pixlar. Dessa block transformeras sedan med DCT, vilket resulterar i en koefficientmatris, där koefficienter som motsvarar låga spatiala frekvenser återfinns i ena hörnet och höga frekvenser i andra. Kodningen bygger på att den lågfrekventa informationen är perceptuellt viktigare, dvs den högfrekventa informationen inte behöver representeras med lika många bitar, och mycket information kan kastas bort. Detta görs genom att kvantisera koefficientmatrisen på ett sådant sätt att högfrekvenskoefficienter kvantiseras hårdare än de "låga". (Många koefficienter blir dessutom i praktiken noll vid kvantiseringen och dessa kan packas extra effektivt med traditionella metoder).

b

Vid JPEG-kodnign av linjeritningar och andra bilder som innehåller mycket skarpa kanter så uppkommer artefakter i bilden, som beror på att skarpa kanter kräver mycket högfrekvent information för att representeras korrekt. Om bilden dessutom bara innehåller ett fåtal färger (t.ex. bara svart och vitt) så blir dessa artefakter extra tydliga.

5

a

Skriv ut summan som

$$y(n) = x(n) + x(n-1) + x(n-2) + \ldots + x(n-p+1) + x(n-p)$$

Detta uttryck gäller ju för alla n, vi kan t.ex. stoppa in $n-1$ och få

$$y(n-1) = x(n-1) + x(n-2) + x(n-3) + \ldots + x(n-p) + x(n-p-1)$$

Nu ser vi att det finns flera termer som är gemensamma ($x(n-1)$ t.o.m. $x(n-p-1)$

Om man subtraherar andra uttrycket från det första, kommer dessa att ta ut varandra och vi får

$$y(n) - y(n-1) = x(n) - x(n-p-1)$$

vilket ger

$$y(n) = y(n-1) + x(n) - x(n-p-1)$$

Q.E.D.

b

Den enkla tolkningen är att när vi flyttar vårt medelvärdesfönster ett steg, så utnyttjar vi det faktum att det nya medlevärdet är *nästan* samma som det förra - istället för att räkna om alltihop så lägger vi bara till den punkt som just kommit in i vårt rullande medelvärdesfönster (dvs $x(n)$)) och drar bort den som just åkt ut (dvs $x(n-p-1)$).