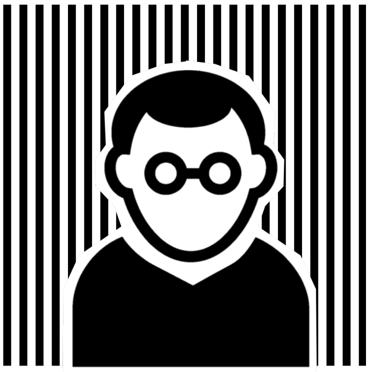
DT1130 Spektrala transformer
Tentamen 190107

Tentamen består av fem uppgifter där varje uppgift maximalt ger 4 p. Normalt gäller följande betygsgränser: **E: 9 p, D: 11.5 p, C: 14 p, B: 16 p, A: 18 p**
Tillåtna hjälpmedel: räknare, formelblad (bifogat)

Lycka till!

1



Gösta ska uppdatera en profilbild. Bilden han vill använda har 600 x 600 pixlar men max storlek för profilbilden är 200 x 200. Han bestämmer sig för att skriva ett program som minskar bildens storlek genom att plocka ut var tredje pixel, rad- och kolumnvis dvs

$$J(x, y) = I(3x, 3y)$$

där I är den gamla bilden och J är den nya, förminskade bilden. När han tittar på den förminskade bilden upptäcker han till sin förvåning att den randiga tapet som syns i bakgrunden har ändrats - det är mycket färre ränder på den nya bilden än på orignalet! Hjälp Gösta att reda ut vad som hänt.

Vi kan anta att tapetmönstret i originalbilden endast beror på x-koordinaten och att intensiteten varierar enligt funktionen

$$i_{tapet}(x) = \sin \frac{5\pi x}{8}$$

där enheten för x är pixlar.

a Hur många tapetränder syns i originalbild respektive förminskad bild, totalt sett räknat från bildens högra till vänstra kant? (2p)

b Vad kallas fenomenet som gösta råkat ut för? (1p)

c Hur skulle gösta ha gjort när han förminskade bilden för att undvika detta? (1p)

2

I figur 1 ser du ett antal pol-nollställesdiagram och ett antal frekvenssvar. Para ihop pol-nollställesdiagrammen med de frekvenssvar som härrör från samma system, med tydlig motivering! Observera att det blir ett pol-nollställesdiagram över.

(1p/ korrekt motiverat par)

3

Betrakta en tvåpolsresonator med poler i $Re^{\pm j\theta}$ där radien $R = \frac{1}{\sqrt{2}}$ och vinkeln $\theta = \pi/4$.

a Ange resonatorns överföringsfunktion $H(z)$

(1p)

b Ange resonatorns filterekvation (dvs ett uttryck på formen $y(n) = ...$)

(1p)

c Studera resonatorns impulssvar $h(n)$, antingen numeriskt (med hjälp av filterekvationen) eller analytiskt (med hjälp av Z-transform).

Vid vilket värde på n blir $h(n)$ negativt första gången? *(2p)*

4

En tvådimensionell filterkärna av stolek 3×3 ges av

$$h(x, y) = 4 - 2x^2 - 2y^2 + x^2y^2$$

där $-1 \leq x \leq 1$ och $-1 \leq y \leq 1$

a Visa att $h(x, y)$ är linjärt separerbar i $x-$ och y *(2p)*

b Beskriv hur separerbarhet kan utnyttjas för att göra faltningen snabbare och ange hur stor vinsten blir procentuellt, räknat i antal multiplikationer, vid faltning med en 3×3-kärna. *(2p)*

5

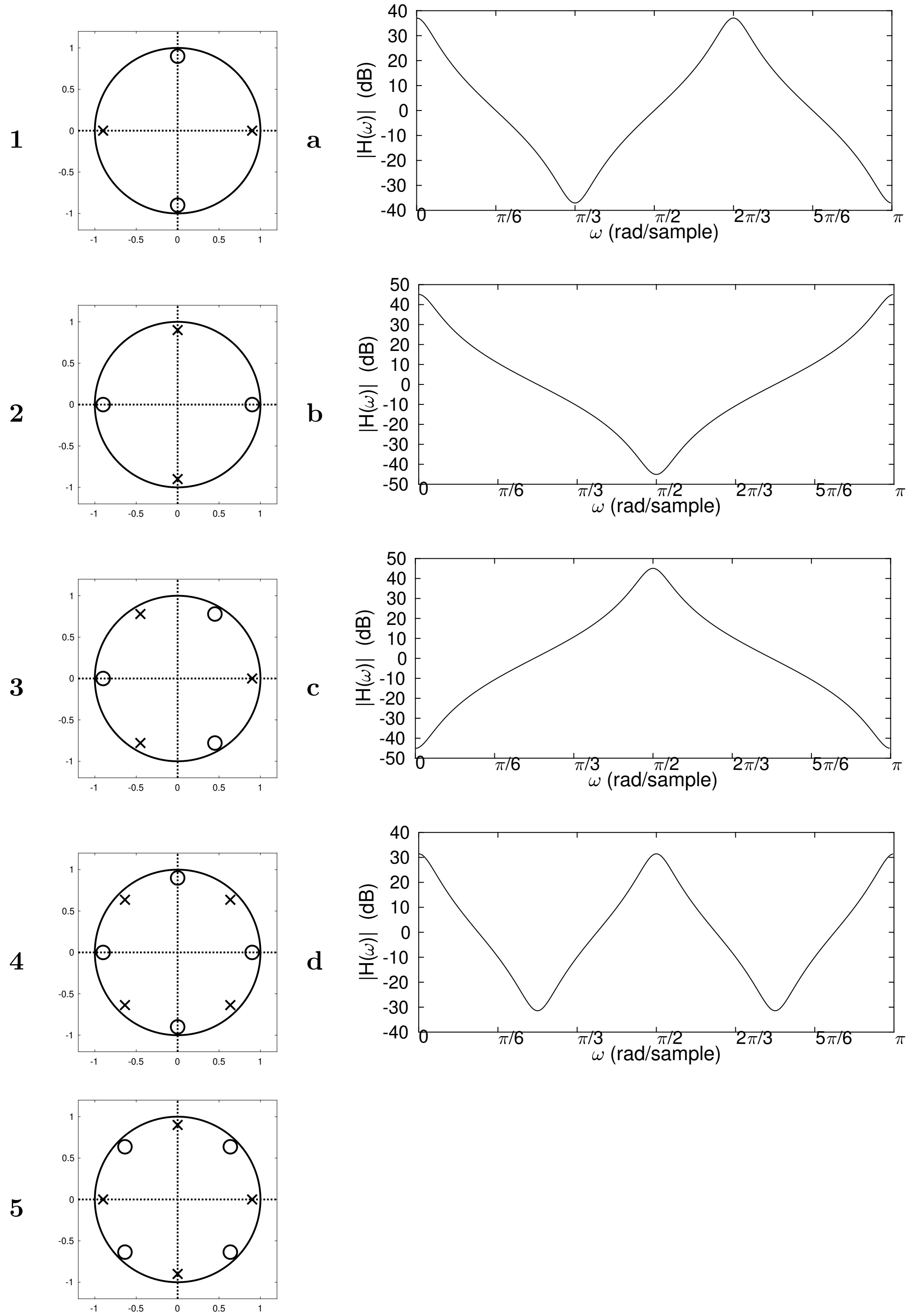
Musiksöktjänsten *Shazam* kan, givet några sekunder musik uppspelat från en okänd högtalare, i okänd akustisk miljö, upptaget med mobiltelefonmikrofon, och med diverse potentiella störningar (t.ex. röster i förgrunden) med förvånansvärt hög träffsäkerhet hitta rätt matchning i en databas med miljontals låtar.

Principen för hur det fungerar beskrivs i en konferensartikel från 2003 [1] (givetvis har algoritmerna som används i tjänsten utvecklats efter detta men artikeln beskriver en fullt fungerande och reproducerbar metod).

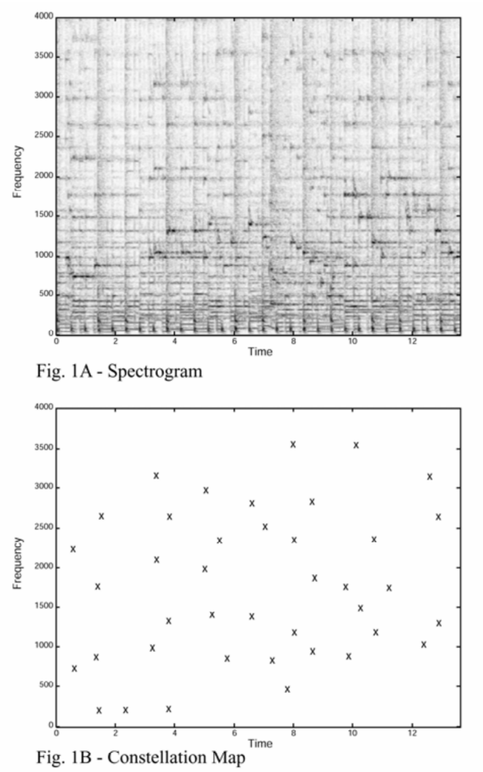
Metoden bygger på musikaliska ”fingeravtryck” som extraheras från all musik i databasen. Samma fingeravtryck beräknas på det inspelade ljudet i telefonen när en sökning ska göras, och en effektiv sökalgoritm hittar den bästa matchningen.

Fingeravtrycken beräknas med hjälp av ett spektrogram (dvs en tvådimensionell representation (bild) av ljudet där x-axeln representerar tid och y-axeln frekvens och gråskalenivån representerar energi - ju mörkare desto starkare).

[1] Wang, A. (2003, October). An Industrial Strength Audio Search Algorithm. In Ismir (Vol. 2003, pp. 7-13).



Figur 1. Vilken av pol-nollställesdiagramen till vänster hör ihop med vilket frekvenssvar till höger? (poler = ×, nollställen = ∘)



Figur 2. Ur artikeln (Wang, 2003): de första två stegen i extraktion av fingeravtryck i Shazams algoritm.

Spektrogrammet görs sedan om till vad författarna kallar en *stjärnkarta* (constellation map) genom att välja ut de hösta/tydligaste topparna, där varje punkt kan beskrivas av sina två koordinater i tid och frekvens (energin/gråskalenivån sparas inte), se figur 2

Dessa punkter grupperas tillsammans med en speciell algoritm och bildar då fingeravtrycken som används i sökningen.

a Antag att ljudsignalen har samplingsfrekvensen $f_s = 8000Hz$ och varje kolumn i spektrogrammet innehåller ett frekvensspektrum som beräknas med hjälp av DFT där fönsterlängden $N = 1024$. Antag vidare att för beräkning av varje ny kolumn så flyttas fönstret i ljudsignalen framåt med 50% av fönsterlängden (dvs varje fönster överlappar med 50% med det föregående). Vad blir upplösningen i frekvensled repsektive tidsled (dvs avståndet i Hz mellan två närliggande punkter i spektrogrammet i y-led, respektive avståndet i sekunder mellan två närliggande punkter i x-led)?

(2p)

b Att göra om spektrogrammet till en stjärnkarta har flera fördelar - det gör bland annat att fingeravtrycken tar mycket mindre lagringsutrymme och att sökningen kan göras mycket mer effektivt.

Men det gör också matchningen mer robust mot variationer i kanalen (olika högtalare, rum och mikrofoner). Varför, tror du? Diskutera!

(2p)

Lösningar

1

a

Intensiteten i tapetmönstret varierar enligt en sinusfunktion med vinkelfrekvens ω radianer per pixel:

$$i_{tapet} = \sin\frac{5\pi x}{8} = \sin\omega x \to \omega = \frac{5\pi}{8}$$

Periodicieteten T, dvs hur tätt det är mellan ränderna i antal pixlar, ges av

$$T = \frac{1}{f} = \frac{2/pi}{\omega} = 2\pi\frac{8}{5\pi} = \frac{16}{5}$$

och antal ränder totalt i originalbilden $N_{orig} = \frac{600}{16/5} \approx 187$

I den förminskade bilden byts x mot $3x$, vilket ger $\omega = \frac{15\pi}{8}$

Eftersom $\omega > \pi$ får vi vikning till $2\pi - \frac{15\pi}{8} = \frac{\pi}{8}$

Periodicieteten $T = \frac{2\pi}{\pi/8} = 16$ vilket ger att antal ränder i den förminskade bilden som $N_{forminsk} = \frac{200}{16} \approx 12$

b

Fenomenet kallas aliasing eller vikning.

c

För att undika aliasing vi förminskning behöver han lågpassfiltrera bilden först, för att få bort alla frekvenser högre än halva samplingsfrekvensen efter nedsampling. I praktiken kan han sätta varje pixel i den förminskade bilden till medelvärdet av motsvarnde 3×3-block i ursprungsbilden.

2

Toppar i frekvenssvaret motsvarar poler, dalar motsvarar nollställen.

a

Toppar vid $\omega = 0, 2\pi/3$, dalar vid $\omega = \pi/3, \pi$. Detta stämmer på diagram **3**.

b

Toppar vid $\omega = 0, \pi$, dal vid $\omega = \pi/2$. Stämmer på diagram **1**.

c

Topp vid $\omega = \pi/2$, dalar vid $\omega = 0, \pi$. Diagram **2**.

d

Toppar vid $\omega = 0, \pi/2, \pi$. Dalar vid $\omega = \pi/4, 3\pi/4$. Måste vara **5**.

3

Två poler: $p_1 = \frac{1}{\sqrt{2}} e^{j\pi/4} = \frac{1}{2} + \frac{j}{2}$ och $p_2 = \frac{1}{\sqrt{2}} e^{-j\pi/4} = \frac{1}{2} - \frac{j}{2}$

a

Överföringsfunktionens nämnare ges direkt av polerna. Täljaren kan vi välja så vi får samma gradtal som nämnaren:

$$H(z) = \frac{z^2}{(z-p_1)(z-p_2)} = \frac{z^2}{(z - \frac{1+j}{2})(z - \frac{1-j}{2})} = \frac{z^2}{z^2 - z + \frac{1}{2}}$$

b

$$H(z) = \frac{Y(z)}{X(z)} \rightarrow Y(z)(z^2 - z + \frac{1}{2}) = X(z)z^2$$

division med z^2 på båda sidor ger

$$Y(z)(1 - z^{-1} + \frac{z^{-2}}{2}) = X(z)$$

Invers z-transform ger

$$y(n) - y(n-1) + \frac{1}{2}y(n-2) = x(n)$$
$$y(n) = x(n) + y(n-1) - \frac{1}{2}y(n-2)$$

c

Alt. 1: numerisk lösning:

Direkt beräkning av impulssvar genom instättning av $x(n) = \delta(n)$ i filterekvationen ger $h(n) = y(n)$ enligt

$h(0) : y(0) = x(0) + y(-1) - 0.5y(-2) = 1 + 0 - 0 = 1$
$h(1) : y(1) = x(1) + y(0) - 0.5y(-1) = 0 + 1 - 0 = 1$
$h(2) : y(2) = x(2) + y(1) - 0.5y(0) = 0 + 1 - 0.5 = 0.5$
$h(3) : y(3) = x(3) + y(2) - 0.5y(1) = 0 + 0.5 - 0.5 = 0$
$h(4) : y(4) = x(4) + y(3) - 0.5y(2) = 0 + 0 - 0.25 = -0.25$
Alltså: $h(n)$ blir negativt första gången vid $n = 4$ då $h(4) = -0.25$

Alt. 2: analytisk lösning:

För att få impulssvaret analytiskt behöver i inverstransformera $H(z)$ till tidsdomänen. För att kunna göra det behöver vi först skriva $H(z)$ på en form som finns i z-transformtabellen (i formelsamlingen), dvs vi vill som mest ha ett första ordningens polynom i nämnaren. Vi kan åstadkomma detta med partialbråksuppdelning. Ansätt:

$$H(z) = \frac{Az}{z - Re^{j\theta}} + \frac{Bz}{z - Re^{-j\theta}}$$

(R och θ enligt uppgiften). Vi ska nu försöka bestämma konstanterna A och B. Sätt på samma bråkstreck och förenkla:

$$H(z) = \frac{Az(z - Re^{-j\theta}) + Bz(z - Re^{j\theta})}{(z - Re^{j\theta})(z - Re^{-j\theta})}$$

$$= \frac{(A+B)z^2 - (ARe^{-j\theta} + BRe^{j\theta})z}{(z - Re^{j\theta})(z - Re^{-j\theta})}$$

Detta uttryck måste blir identiskt med överföringsfunktionen i *a)* vilket ger följande ekvationer:
z^2-termen ger $A + B = 1$ z-termen ger $ARe^{-j\theta} + BRe^{j\theta} = 0$
Sätt in den första i den andra och förenkla:

$$ARe^{-j\theta} + (1 - A)Re^{j\theta} = 0$$

$$\frac{Re^{j\theta}}{Re^{j\theta} - Re^{-j\theta}} \qquad \rightarrow B = -\frac{Re^{-j\theta}}{Re^{j\theta} - Re^{-j\theta}}$$

Nu kan vi inverstransformera det uppdelade uttrycket för $H(z)$ med hjälp av 4:e z-transformen i formelsamlingen:

$$h(n) = A(Re^{j\theta})^n u(n) + B(Re^{-j\theta})^n u(n)$$

Sätt in A och B enl. ovan och förenkla med eulers formler:

$$h(n) = \frac{Re^{j\theta}(Re^{j\theta})^n - Re^{-j\theta}(Re^{-j\theta})^n}{R(e^{j\theta} - e^{-j\theta})} u(n) = \frac{R^n \sin(\theta(n+1))}{\sin\theta} u(n)$$

Med $R = 1/\sqrt(2)$ och $\theta = \pi/4$ blir det

$$h(n) = \frac{(\frac{1}{\sqrt{2}})^n \sin\frac{\pi(n+1)}{4}}{\frac{1}{\sqrt{2}}} u(n)$$

Detta uttryck blir negativt då $\sin\frac{\pi(n+1)}{4}$ är negativt, vilket sker första gången då $n = 4$.

4

a

För att en filterkärna $h(x, y)$ ska vara linjärt separerbar ska den kunna skrivas på formen $h(x, y) = h_1(x)h_2(y)$
Givet:

$$h(x, y) = 4 - 2x^2 - 2y^2 + x^2y^2$$

Faktorisera uttrycket:

$$h(x, y) = x^2(y^2 - 2) + 4 - 2y^2 = x^2(y^2 - 2) - 2(y^2 - 2) = (x^2 - 2)(y^2 - 2)$$

V.S.V

b

Om filterkärnan är separerbar kan faltningen göras i två steg, först i X-led med $h_1(x)$ (en $1 \times N$ -kärna), sedan i Y-led med $h_2(y)$ ($N \times 1$). Antalet multiplikationer per pixel blir då $2N$, istället för N^2 om kärnan inte är separerbar.

Man behöver alltså $\frac{2N}{N^2}$ gånger färre multiplikationer om kärnan är separerbar. För $N = 3$ innebär det en vinst på 33%.

5

a

Vid DFT kommer frekvensaxeln delas i N lika stora delar när högsta frekvensen är samplingsfrekvensen f_s. Alltså blir upplösningen i frekvensled

$$\Delta f = \frac{f_s}{N} = \frac{8000}{1024} \approx 7.8 \text{ Hz}$$

.

Upplösningen i tidsled bestäms av avståndet mellan två på varandra följande fönster längs tidsaxeln. Fönsterlängden N är 1024 och förflyttningen varje steg är 50% dvs $\frac{N}{2} = 512$ sampel. För att få det i sekunder delar vi med f_s:

$$\Delta t = \frac{N}{2f_s} = \frac{1024}{2 \cdot 8000} = 0.064 \text{ s}$$

.

b

Varje del av kanalen (högtalare, rum, mikrofon) kan ses som ett linjärt filter. Dessa kombineras (kaskadkopplas) för att bilda den faktiska kanalen. Detta filter påverkar energin i varje frekvens, vilket innebär att spektrumet vid en viss tidpunkt kan te sig väldigt olika i den ursprungliga signalen och i signalen som ska matchas. Dock kommer topparna i spektrumet i originalsignalen förbli toppar i den filtrerade signalen (förutsatt att frekvenssvaret är någorlunda kontinuerligt). Topparna är dessutom den del av spektrum som är minst känsligt för störningar eftersom det krävs starka ljud för att "överrösta" dessa.

Topparna är också robusta i tidsdomänen: när ett ljud spelas upp i ett rum uppkommer ekon pga rumsakustiken men det starkaste ljudet är oftast direktljudet dvs det med kortast fördröjning.