



KTH Datavetenskap
och kommunikation

Spektrala transformer

Laboration: Vokalsyntes

1 Introduktion

I denna laboration är målsättningen att du ska få handgripliga erfarter av digital filtrering. Du ska implementera en enkel men användbar typ av återkopplat filter - tvåpolsresonator, och använda den för att simulera mänskligt tal.

2 En modell av vokalproduktion

Produktion av vokalljud i mänskligt tal kan grovt sett beskrivas så här: luft pressas ur lungorna mellan stämbanden, vilket ger upphov till ljudpulser, som fortplantas genom talröret (strupen, munhålan, näshålan, etc). Talröret formar ljudet olika beroende på form och läge hos tunga, käke, läppar mm, och producerar den slutliga ljudsignalen. Stämbanden kan därför ses som en *ljudkälla*, och talröret som ett *filter*. Denna populära beskrivningsmodell kallas *källa-filter-modellen* och den ska du tillämpa i denna laboration. Det filter som talröret utgör har en komplicerad överföringsfunktion, som innehåller ett antal resonansfrekvenser som kallas *formanter*. För vokaler kan filtret effektivt approximeras med ett antal tvåpolsresonatorer som parallellkopplas, där varje formant motsvaras av en resonator. Vanligtvis räcker det med att modellera de första formanterna, vars centrumfrekvenser brukar betecknas F_1 , F_2 , F_3 och F_4 , med tillhörande bandbredder B_1 , B_2 , B_3 och B_4 .

3 Utförande

Du ska utföra ett antal uppgifter med hjälp av Matlab, och svara på de frågor som ställs i peket. Laborationen utföres självständigt, antingen enskilt eller i grupp om två. Börja med att hämta `lab-reson.zip` från kurshemsidan. Denna fil innehåller några hjälpfunktioner och andra filer som du kommer att behöva. Packa upp arkivet i din hemkatalog och använd detta som "working directory" i matlab.

I matlabkoden i peket betecknas alltid arrayer och matriser med versaler, medan skalärer betecknas med gemener. Vidare används genomgående konventionen att skriva all kod med `skrivmaskinsstil`. Den kod som efterfrågas (och ska redovisas) ska oftast vara i form av matlabfunktioner. Ofta beskrivs även hur du kan testa att funktionen gör vad den ska. Du gör klokt i att skriva dessa tester i m-filer (behöver *ej* vara funktioner) för att göra själva testandet rationellt och konsekvent.

Var noga med att dokumentera vad du gör och spara kod, plottar, bilder, ljudfiler och vad det kan vara till redovisningstillfället, då du ska redogöra för hur du kommit fram till dina resultat.

3.1 Tvåpolsresonatoren

En tvåpolsresonator är en enkel typ av återkopplat filter som beskrivs av överföringsfunktionen

$$H(z) = \frac{b_0}{(1 - Re^{j\theta}z^{-1})(1 - Re^{-j\theta}z^{-1})}$$

som har två poler i det komplexa planet på avståndet R från origo, och med vinklarna $\pm\theta$. (Dessutom har den två nollställen i origo, men dessa påverkar inte frekvenssvaret.)

Tvåpolsresonatoren har ett frekvenssvar som består av en topp, vars läge i frekvensled bestäms av polvinkeln θ , och toppens "spetsighet" bestäms av polradien R (ju närmare R är 1, desto spetsigare topp).

Impulssvaret hos en tvåpolsresonator kan beskrivas som en dämpad svängning, där graden av dämpning bestäms av R - ju mindre R desto mera dämpning får man. Specialfallet när $R = 1$, dvs polerna ligger på enhetscirkeln, ger en helt odämpad svängning dvs en oscillator. $R > 1$ ger ett instabilt filter.

I denna laboration vill vi styra resonatoren i termer av resonansfrekvens f och bandbredd ψ . Polernas parametrar R och θ kan approximativt bestämmas ur frekvensen f och bandbredden ψ enligt följande¹:

$$R \approx 1 - \psi/2$$

$$\theta \approx 2\pi f$$

där f och ψ är resonansfrekvens resp. bandbredd som fraktion av samplingsfrekvensen, dvs $f = 0.5$ motsvarar nykvistfrekvensen.

Förberedelseuppgift (utan matlab)

Skriv om uttrycket för överföringsfunktionen $H(z)$ ovan på formen

$$H(z) = \frac{b_0}{a_0 + a_1z^{-1} + a_2z^{-2}}$$

och bestäm koefficienterna a_k uttryckt i frekvens f och bandbredd ψ .

Studerar man $H(z)$ ser man att täljarens enda koefficient b_0 kommer att tjäna som någon sorts global förstärkningsfaktor. Detta är speciellt viktigt om resonansfrekvensen ska vara variabel, är det ofta önskvärt att filtrets totala förstärkning är oberoende av resonansfrekvensen. Vi kan låta b_0 ta hand om detta åt oss, genom att välja den så att filtrets förstärkning vid resonansfrekvensen alltid är ett. Detta fås då

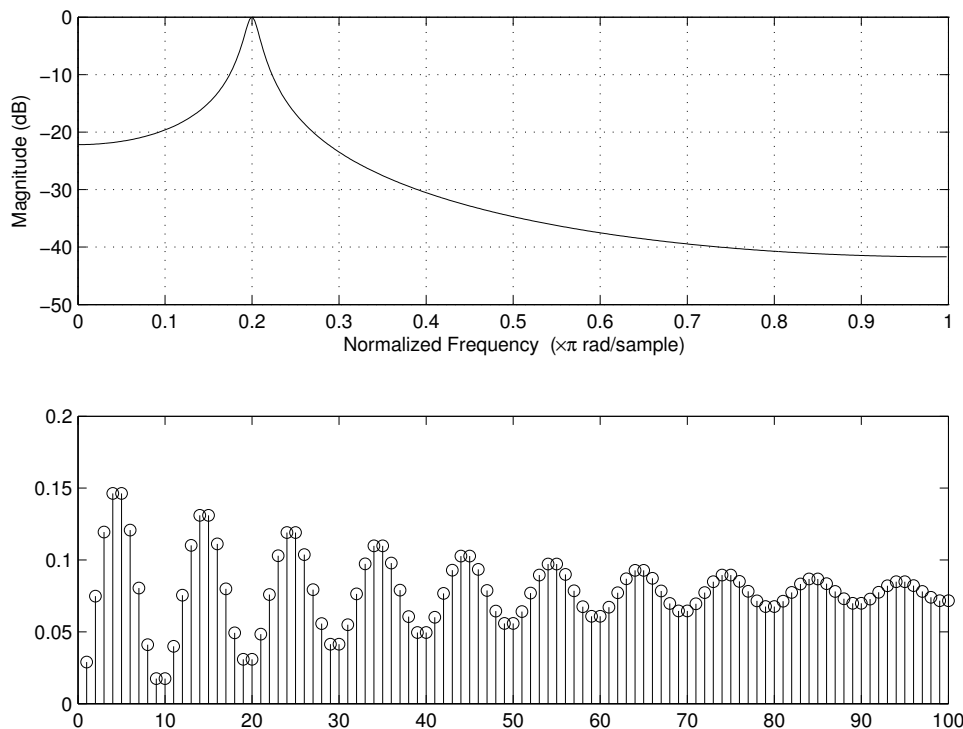
$$b_0 = (1 - R^2) \sin \theta$$

Verifiera gärna att det stämmer med egna uträkningar!

Uppgift 1: statiska vokaler

I den första uppgiften ska du använda det inbyggda matlabkommandot `filter` för att skapa ett statiskt tvåpolsfilter enligt ovanstående överföringsfunktion. Som inparametrar till `filter`

¹ dessa approximationer gäller när $R \approx 1$, dvs för relativt skarpa resonanser, vilket kan antas i denna laboration



Figur 1. Frekvenssvar och impulssvar för tvåpolsresonatorn då $f = 0.1$ och $\psi = 0.05$.

anger man två vektorer B och A som innehåller koefficienterna $b_0, b_1, \dots$ resp. $a_0, a_1, \dots$, samt en vektor X - indatasekvensen som ska filtreras. För mer info, se `help filter`.

Innan du ger dig på att göra vokalljud är det bra att verifiera att filtret gör vad det ska. Ett bra sätt att göra detta är att plotta frekvenssvar och impulssvar för filtret. Frekvenssvaret får du med kommandot `freqz(B,A)`. Impulssvaret fås genom att mata in en impuls (se tipsruta nedan). I figur 1 kan du se hur dessa plottar bör se ut för $f = 0.1$ och $\psi = 0.05$. Pröva även med andra värden på f och ψ tills du förstår hur saker hänger ihop.

Med hjälp av ditt tvåpolsfilter ska nu syntetisera några statiska vokaler. Använd två parallellkopplade filter för att simulera de två första formanterna F_1 och F_2 . I tabell 1 finner du typvärden på formantfrekvenserna för de Svenska vokaler. Som källfunktion (dvs "stämband") kan du använda ljudfilen `source.wav`. Läs in filen med kommandot `[X,fs] = wavread('source.wav')` och filtrera med frekvenser ur tabellen, för några olika vokaler som du väljer. Bandbredderna kan du sätta enligt följande: $B_1 = 50Hz$, $B_2 = 75Hz$, ($B_3 = 100Hz$, $B_4 = 150Hz$). Lyssna, och spara ljuden till fil och titta gärna på spektrum i exempelvis *WaveSurfer*². Notera att frekvenser och bandbredder naturligtvis måste räknas om till normerad frekvens, dvs delas med samplingsfrekvensen.

Matlab-tips

En insignal med en impuls skapas enkelt med

```
X = [1 zeros(1,n-1)]
```

där n är önskad sekvenslängd. För plottningen kan det vara bra att även skapa en vektor med tidsvärden:

²Laddas hem från www.speech.kth.se/wavesurfer

Tabell 1. Formantfrekvenser i Hz för Svenska vokaler för en mansröst (efter Fant). För en kvinnoröst ligger formanterna ca 20% högre.

Vokal	exempel	F_1	F_2	F_3	F_4
O:	rot	300	600	2350	3250
O	ro ^t t	350	700	2600	3200
Å:	rå	400	700	2450	3250
Å	rå ^t t	500	850	2550	3250
A:	bar	600	950	2550	3300
A	ba ^r r	750	1250	2500	3350
I:	bil	250	2200	3150	3750
I	Bi ^l	350	2150	2750	3500
E:	deg	350	2250	2850	3550
E/Ä	väg ^g	500	1900	2550	3350
Ä3	he ^r r	650	1700	2500	3450
Y:	ny	250	2050	2700	3300
Y	ny ^t t	300	2000	2400	3250
Ö:	föd	400	1750	2300	3350
Ö	fö ^d d	550	1550	2450	3300
Ö3	fö ^r	550	1150	2450	3250
U:	duk	300	1650	2250	2250
U	puc ^k	450	1050	2300	3300

`T = 0:n-1`

. Plotta t.ex. med `stem(Y)` eller `plot(T,X,'o',T,Y,'x')` - den senare visar både insignal och utsignal.

Matlab-tips

För att läsa in en ljudfil som en vektor använder du funktionen `wavread()`.

För att lyssna på en matlab-vektorn `Y` som ett ljud med samplingsfrekvensen 16000 Hz kan du använda kommandot

`soundsc(Y,16000)`

För att skriva samma ljud som en wave-fil använder du

`wavwrite(Y,16000,'mitt_ljud.wav')`

3.2 Tidsvariabelt filter

I talsignalen (och i alla andra någorlunda intressanta ljud) så varierar frekvensinnehållet kontinuerligt över tiden. Detta innebär att om man ska modellera tal med ett filter, måste filterparametrarna kunna variera. Problemet i detta fall är att matlabs **filter** inte tillåter variabla filterkoefficienter, så vill man göra detta måste man antingen filtrera små bitar (typ 10 millisekunder) åt gången, eller så får man göra sin egen filterfunktion. Det senare är mer lärorikt och inte alls så svårt som det kanske låter.

Uppgift 2: en tidsvariabel resonator

Skriv en funktion som implementerar en tidsvariabel tvåpolsresonator - ett formantfilter. Funktionen ska följa nedanstående prototyp (dvs ha samma in/utparametrar):

```
function Y = formantfilter(F,BW,X)
%
% formantfilter - time-varying two-pole resonator filter
%
% input:
%   F (1xN row vector)    - Time-varying resonance frequency (normalized by fs)
%   BW (real number)      - Constant bandwidth (normalized by fs)
%   X (1xN row vector)    - Input sequence
% output:
%   Y (1xN row vector)    - Output sequence
```

Funktionen ska alltså ta en insekvens $x(n)$ och filtrera den enligt filterekvationen

$$y(n) = b_0x(n) - a_1y(n-1) - a_2y(n-2)$$

där b_0 och a_n är desamma som de du fick fram i förberedelseuppgiften. Den viktiga skillnaden nu är att de ändras hela tiden, eftersom de beror av F som är en vektor av samma längd som X . Bandbredden kan vara konstant i denna laboration.

*tips: denna uppgift löses bäst med en for-slinga (se **help for**) - något som man annars ofta kan klara sig utan i matlab om man använder matriser på bra sätt.*

Om du vill testa funktionen kan du filtrera en brussignal med ett filtersvep – låt f svepa från 0 till 0.5, t.ex. $F = (1:8000)/16000;$. En brussignal av samma storlek som F skapar du enkelt med $N=\text{rand}(\text{size}(F));$. Sätt bandbredden till 0.05. Lyssna på ljudet - det ska låta som någon sorts "swoosch"... Ett spektrogram över ljudet bör visa ett tydligt diagonalt stigande band.

Att redovisa: m-fil.

3.3 Icke-statiska vokaler

I naturligt tal så ändras talröret - och därmed formanterna - kontinuerligt, dvs de glider mellan olika vokaler. Frekvenserna i tabell 1 är en sorts riktvärden för normaltalt, men i praktiken handlar det om ett kontinuum (ofta refereerat till som *vokalrymden*).

Nu ska du syntetisera icke-statiska vokaler, sk diftonger med hjälp av den nya resonatorn. Resonatorerna parallellkopplas, precis som i första uppgiften, men använd här fyra istället för två. Det ger lite mer definierad diskant (ljudet kan bli lite för vasst - i så fall kan man minska den relativa nivån på resonator 3 och 4 genom att helt enkelt applicera en skalfaktor på resonatorns utsignal. 0.7 för formant 3 och 0.3 för formant 4 kan vara bra, prova dig gärna fram)

Uppgift 3: sammansättning

Skriv ett program som genererar diftonger, dvs glidningar från en vokal till en annan. Man ska kunna ange startvokal och slutvokal, samt längden för övergången. För att underlätta, så finns filen `vokaler.m` som innehåller formantfrekvenserna i tabellen ovan, i form av en variabel för varje vokal (titta i filen så förstår du).

För att göra en vektor som går från v_0 till v_1 i $N+1$ steg, kan man göra nåt i den här stilen:

```
q = 0:1/N:N;
v = v0 + q*(v1-v0);
```

Använd detta för att skapa styrvektorerna till formantfiltren. Bandbredden kan du sätta till samma som i uppgift 1 (den kan hållas konstant). Glöm inte att frekvenser & bandbredd måste normeras (delas med f_s) innan de går in till filtret!

Du behöver även en röstkälla - använd den bifogade funktionen `simplesource()` - det är en enkel implementation av en tidsvariabel röstkälla där varje röstpuls modelleras som en "spik". Denna källa måste också ha en styrparameter, F_0 dvs grundtonsfrekvensen. Denna kan modelleras enligt samma princip som formatfrekvenserna, dvs ange önskad start och slutfrekvens. F_0 för en mansröst varierar oftast inom intervallet $85 - 180\text{Hz}$, och för en kvinnoröst mellan $165 - 255\text{Hz}$.

Uppgift 4 (frivillig): kvinnor, barn och längre sekvenser

Utveckla diftong-synten så att du kan syntetisera vokalsekvenser av valfri längd, genom att foga samman diftonger. OBS, gör sammanfogningen direkt på styrvektorerna (och inte på ljudsignalen) för att undvika diskontinuiteter. Utnyttja definitionerna i `vokaler.m` - det kan vara smidigt att beskriva vokalsekvenser på formen $U = [\mathbf{vA} \ \mathbf{vI} \ \mathbf{vA} \ \mathbf{vO}]$ - detta ger en matris med *målvärden* för vokalsekvensen där varje kolumn motsvarar en tidpunkt och varje rad en formant. Motsvarande sekvenser kan göras för grundtonen F_0 och för durationen. Gör sedan en loop som sätter ihop kolumnerna parvis till diftonger.

Försök få din vokalsyntesmaskin att säga olika saker (fundera ut ord och meningar som bara innehåller vokaler - det finns fler än man tror) och laborera med olika F_0 -konturer - stigande, fallande, topp i mitten, etc. Kan du få det att låta som en fråga?

Försök även att få rösten att låta som en kvinna eller ett barn genom att applicera skalfaktorer på formantfrekvenserna. Riktvärden att testa kan vara +20% för kvinnoröst, +50% för barnröst. Du kan skala om grundtonskurvan (F_0) med större skalfaktorer. Vad händer om du sänker formanterna?

Att redovisa: m-fil, ljudfiler och kommentarer.

Matlab-tips

För att skala kolumnerna i en matris kan man skapa en diagonal matris med skalfaktorer, som man multiplicerar med. Exempel: `B=A*diag([2 1 .5])` skalar kolumnerna i A med 2, 1 respektive 0.5.