

# MUS4831 - Lydanalyse

## Signalrepresentasjoner

5 september 2013

Tre hovedområder:

- Tidsdomenet
- Frekvensdomenet
- Perseptuelle modeller

Først litt repetisjon (om tidsdomenet) fra MUS2801 Lydlære 1.

# Repetisjon fra Lydlære: Periodisitet

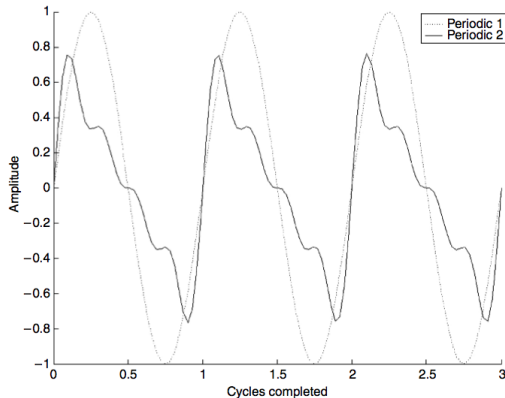
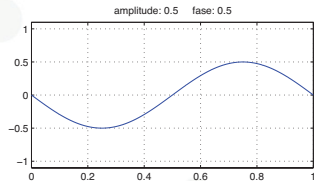
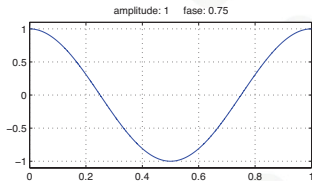
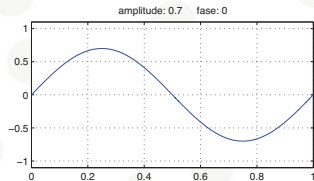
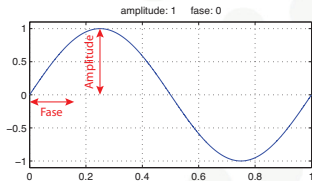


Figure 1.2 Periodic waveforms.

- Her har begge de to bølgeformene samme frekvens
- Frekvens: Antall svingninger per sekund
- Test selv noen enkle tidsdomeneoperasjoner, f.eks. med *Grapher* (Mac)

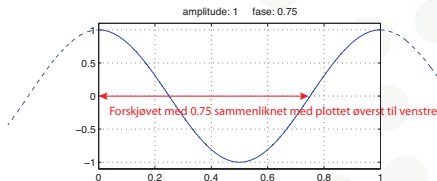
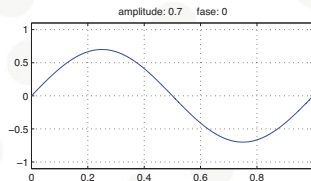
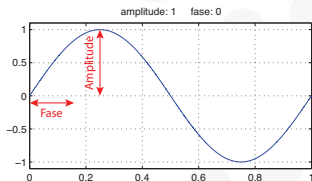
# Repetisjon fra Lydlære: Amplitude og Fase

sinustoner gjengis med *amplitude* (også kalt magnitude) og *fase*:



# Repetisjon fra Lydlære: Amplitude og Fase

sinustoner gjengis med *amplitude* (også kalt *magnitude*) og *fase*:



# Repetisjon fra Lydlære: Grunnfrekvens og Overtoner

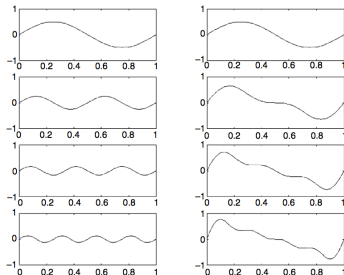
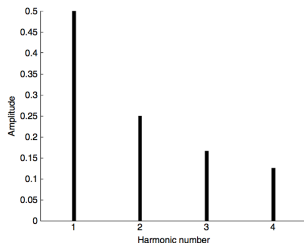
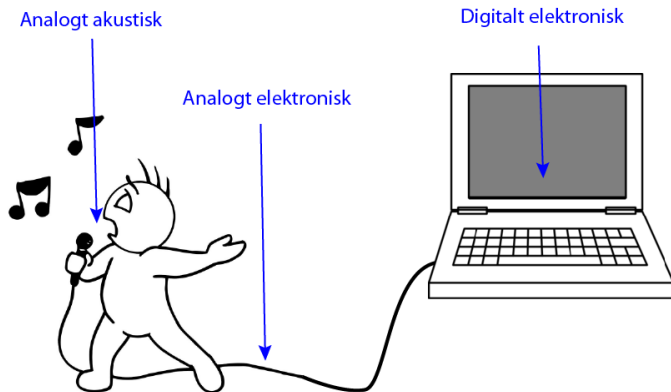


Figure 1.3 Sinusoidal components of a periodic sound. The left column shows each individual sine component (they vary in their amplitude). The right column gives the mix so far at each stage, as the sines are added together down the page.

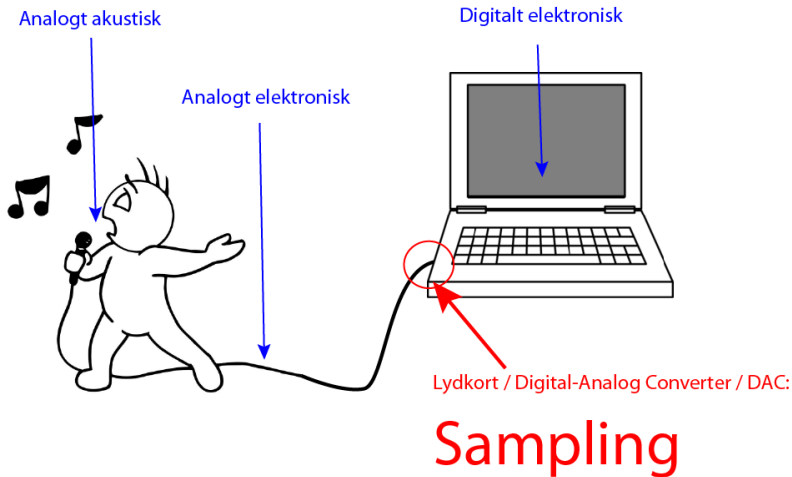


Harmoniske toner består som regel av en ganske tydelig grunnfrekvens, med overtoner som har frekvens lik  $2, 3, 4, 5, \dots$  ganger grunnfrekvensen.

# Repetisjon fra Lydlære: Analog → Digital

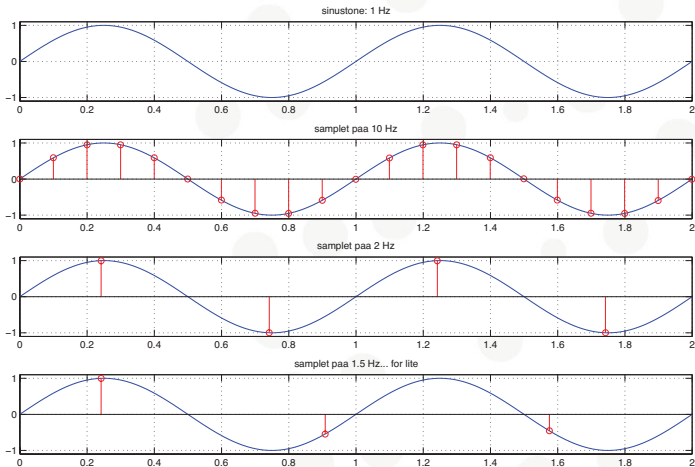


# Repetisjon fra Lydlære: Analog → Digital

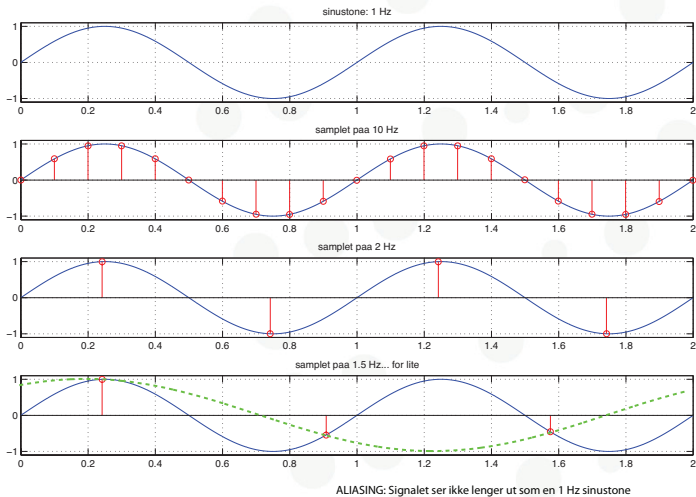




# Repetisjon fra Lydlære: Sampling



# Repetisjon fra Lydlære: Sampling



# Repetisjon fra Lydlære: Nyquist

## Nyquist–Shannons samplingsteorem (litt omskrevet):

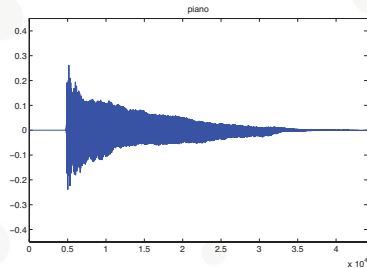
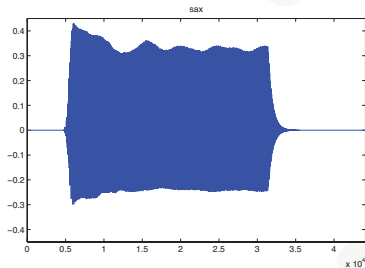
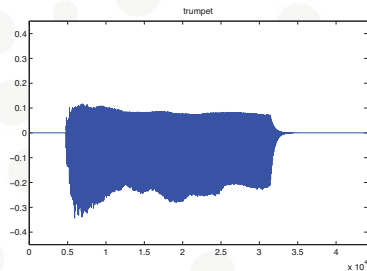
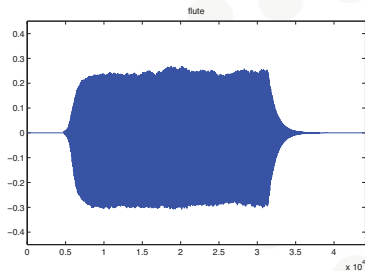
Hvis et signal inneholder frekvenskomponenter på opptil  $X$  Hz, må vi minimum bruke en samplingsrate på 2 ganger  $X$  Hz.

CD samplingsfrekvens: 44100 Hz

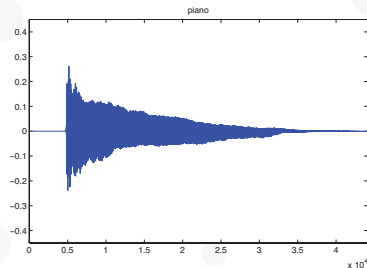
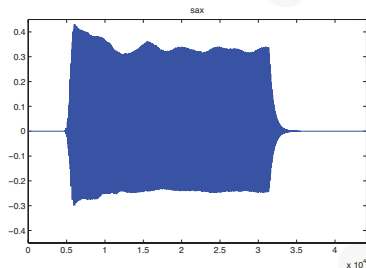
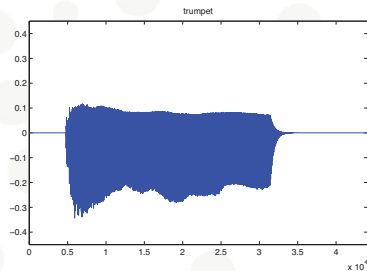
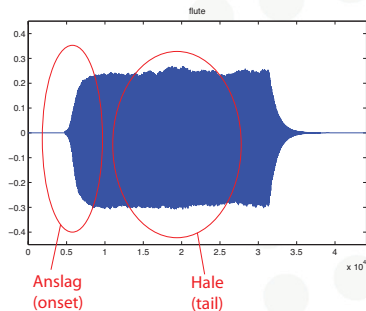
høyeste gjengivbare frekvens: 22050 Hz (Nyquistfrekvensen)

Menneskelig hørsel: ca 20 – 20000 Hz

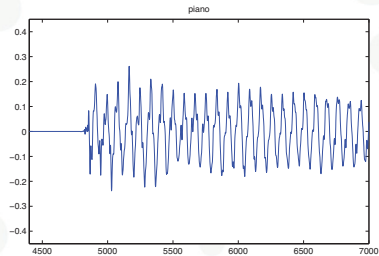
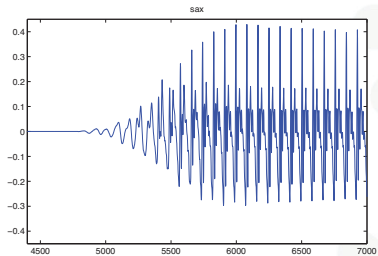
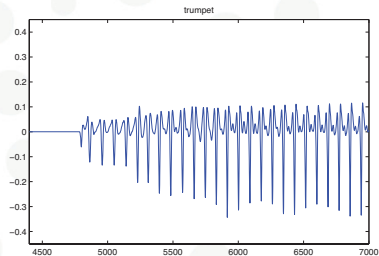
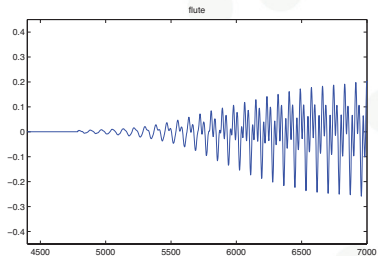
# Noen “naturlige” (fra sampler) lyder i tidsdomenet



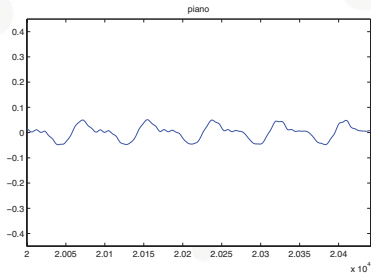
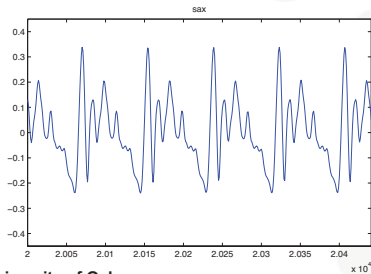
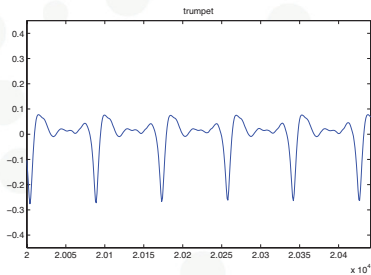
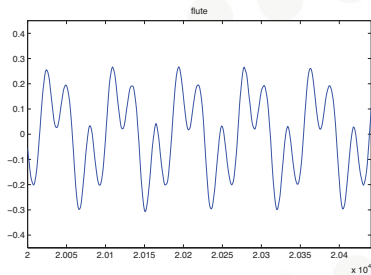
# Noen “naturlige” (fra sampler) lyder i tidsdomenet



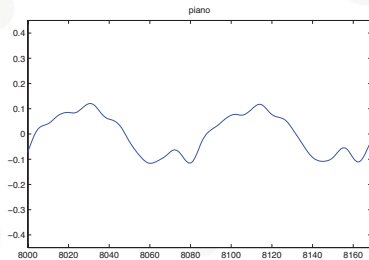
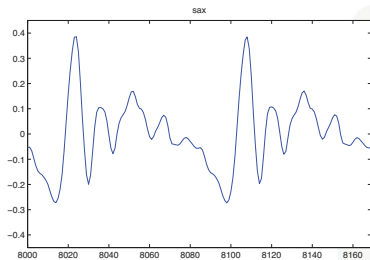
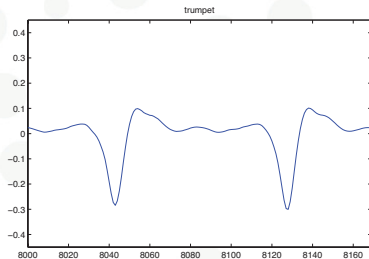
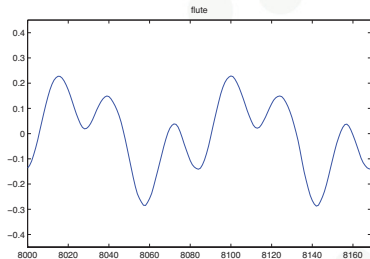
# Vi zoomer inn på anslaget



# Og på halen



# Vi zoomer helt inn på kun to perioder





# Fra tidsdomenet til frekvensdomenet: Fourier Transform

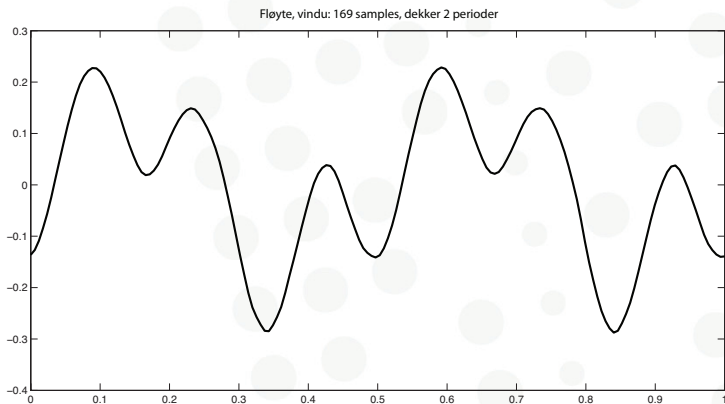
- Ethvert signal kan representeres som en sum av sinustoner med angitt amplitude og fase.
- Vi ser vanligvis på et relativt kort signal for å finne frekvensinnholdet i dette. Dersom vi skal analysere et lengre signal deler vi det opp i overlappende “vinduer”.

Flere varianter.

- Discrete Fourier Transform (DFT)
- Short-Time Fourier Transform (STFT)
- Fast Fourier Transform (FFT)
- og andre varianter som brukes på analoge signaler

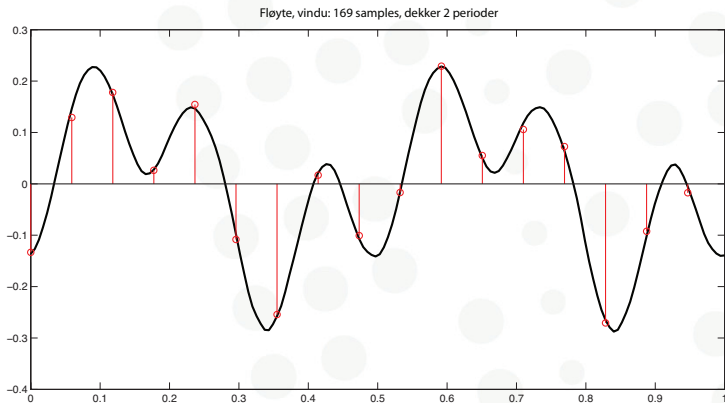
Det viktigste i dette kurset er ikke forskjellen på de ulike typene.

# Fra tidsdomenet til frekvensdomenet: Fourier Transform



Her er vårt fløytesignal

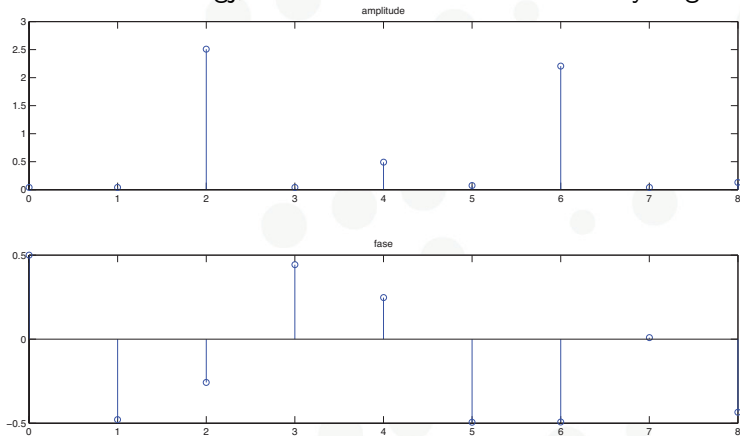
# Fra tidsdomenet til frekvensdomenet: Fourier Transform



For enkelhets skyld nedsampler vi til et signal med 17 sampler

# Fra tidsdomenet til frekvensdomenet: Fourier Transform

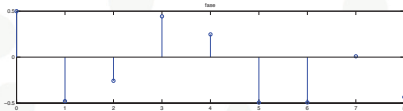
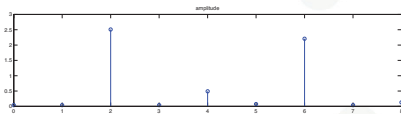
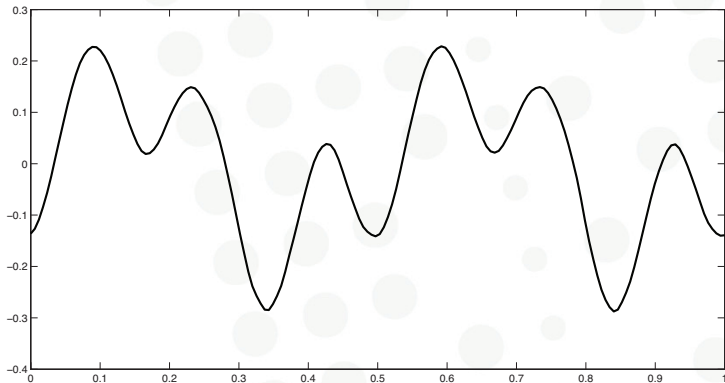
Resultatet av å gjøre en Fourier Transform er to nye signaler:



Her er ikke x-aksen tid, men 'bin'.

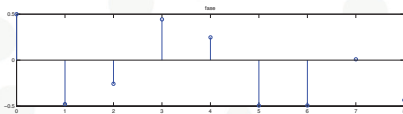
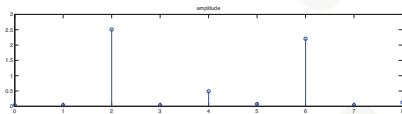
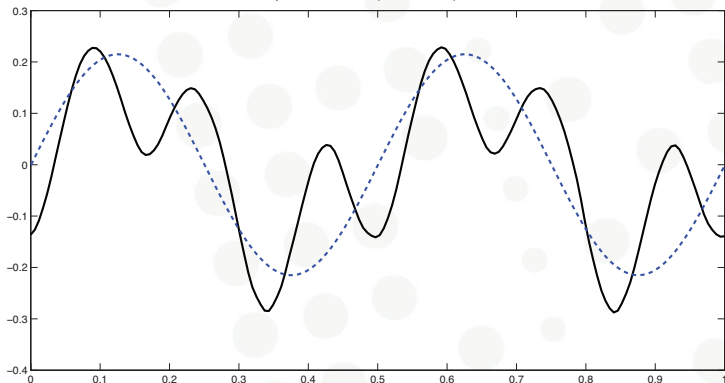
# Fra tidsdomenet til frekvensdomenet: Fourier Transform

Fløyte, vindu: 169 samples, dekker 2 perioder



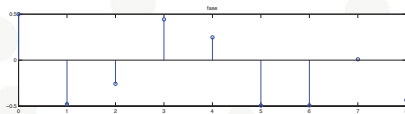
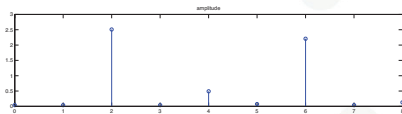
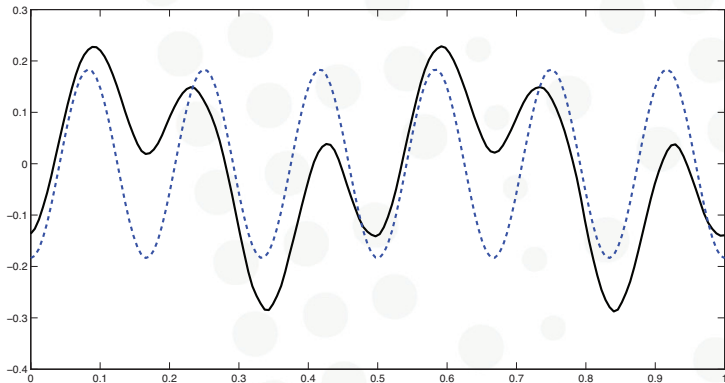
# Fra tidsdomenet til frekvensdomenet: Fourier Transform

Fløyte, vindu: 169 samples, dekker 2 perioder



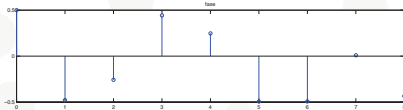
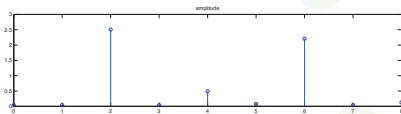
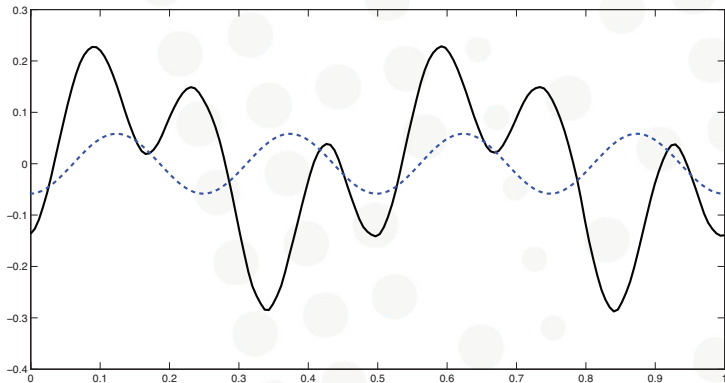
# Fra tidsdomenet til frekvensdomenet: Fourier Transform

Fløyte, vindu: 169 samples, dekker 2 perioder



# Fra tidsdomenet til frekvensdomenet: Fourier Transform

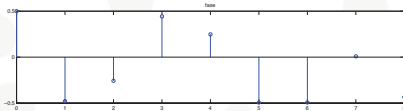
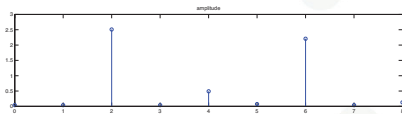
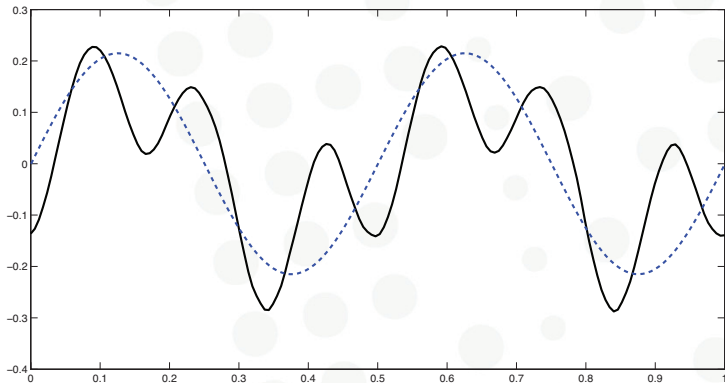
Fløyte, vindu: 169 samples, dekker 2 perioder





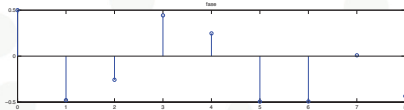
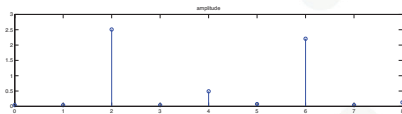
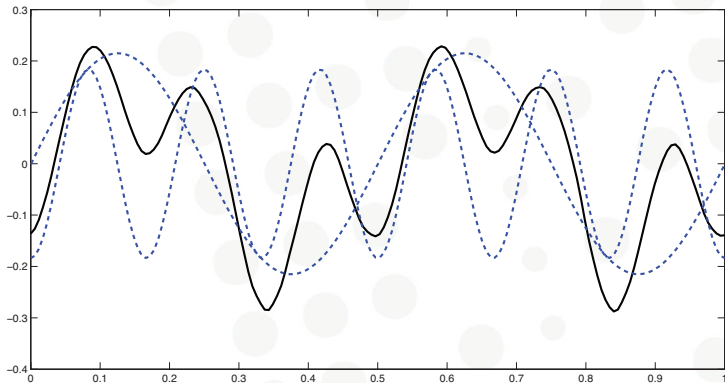
# Fra tidsdomenet til frekvensdomenet: Fourier Transform

Fløyte, vindu: 169 samples, dekker 2 perioder



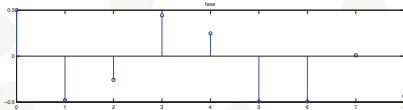
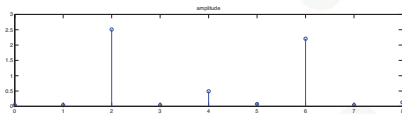
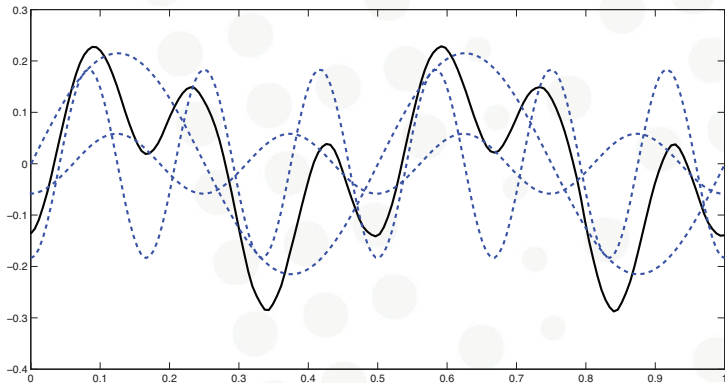
# Fra tidsdomenet til frekvensdomenet: Fourier Transform

Fløyte, vindu: 169 samples, dekker 2 perioder



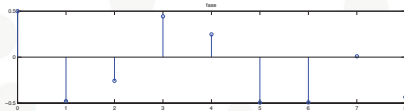
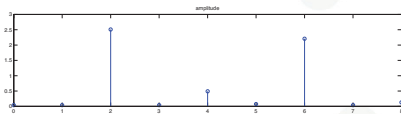
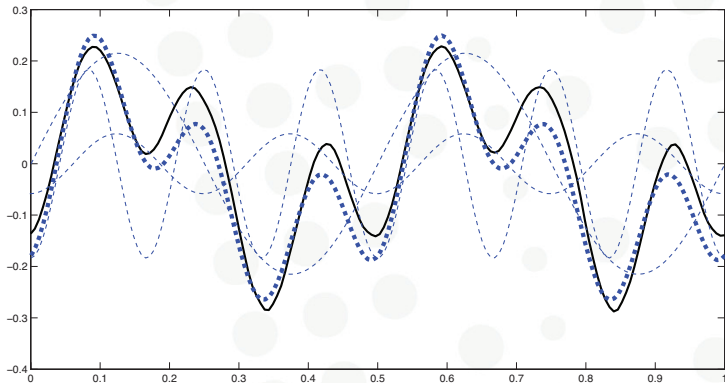
# Fra tidsdomenet til frekvensdomenet: Fourier Transform

Fløyte, vindu: 169 samples, dekker 2 perioder



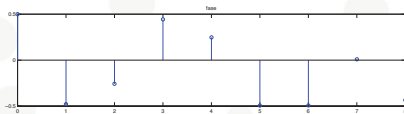
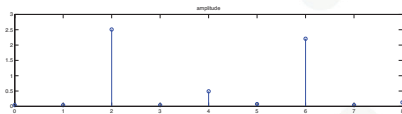
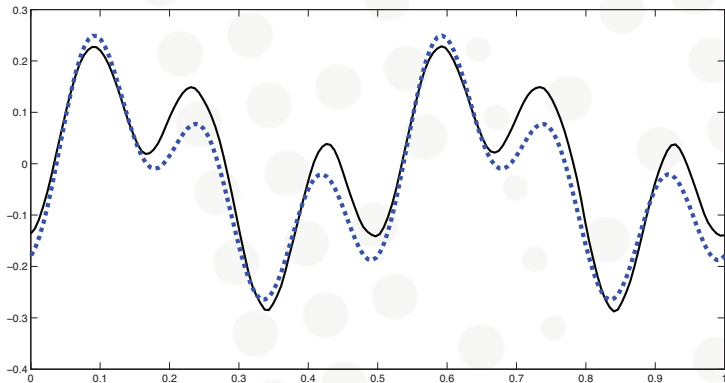
# Fra tidsdomenet til frekvensdomenet: Fourier Transform

Fløyte, vindu: 169 samples, dekker 2 perioder



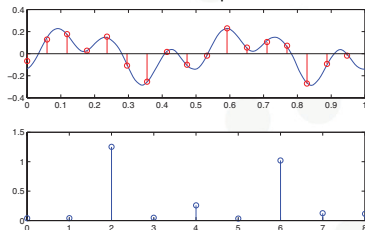
# Fra tidsdomenet til frekvensdomenet: Fourier Transform

Fløyte, vindu: 169 samples, dekker 2 perioder

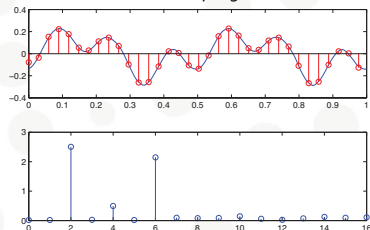


# Effekten av å doble vinduslengden

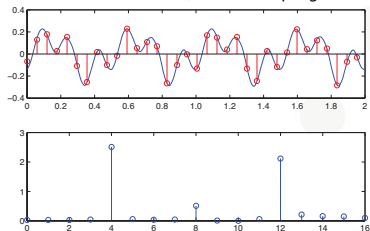
Vårt eksempel



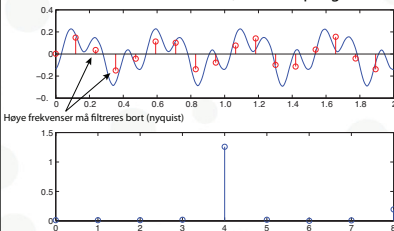
Dobbel samplingsrate



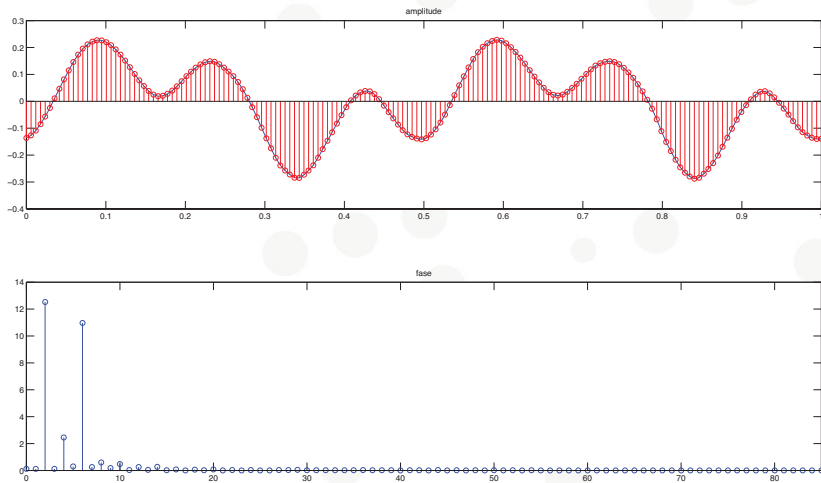
Dobbel vindusstørrelse, samme samplingsrate



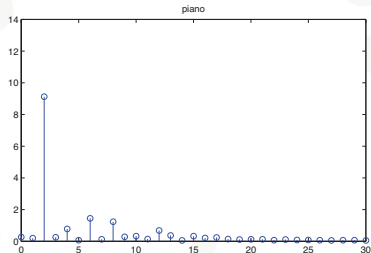
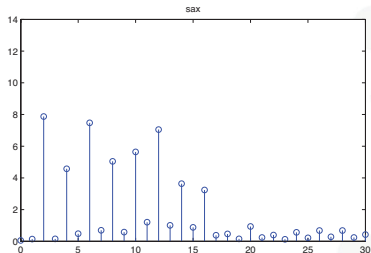
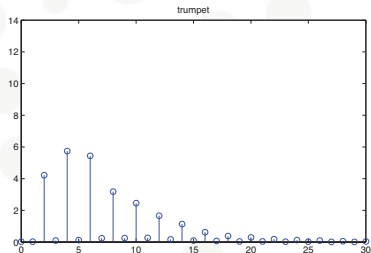
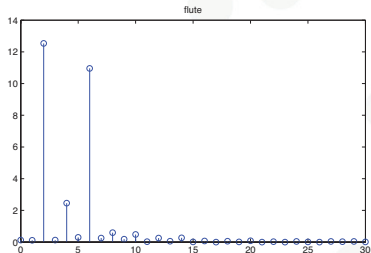
Dobbel vindusstørrelse, halv samplingsrate



# Vanlivis nedsmpler vi ikke:



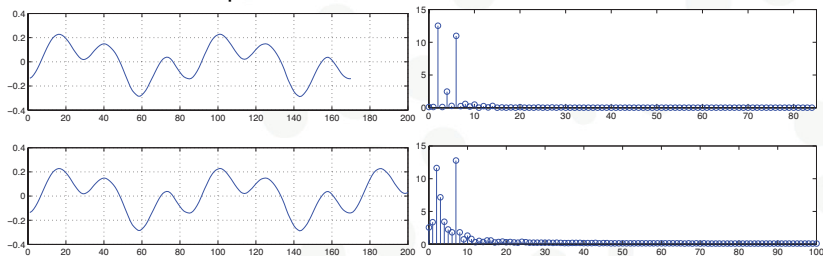
# Fire instrumenter, full samplingsrate, vindu 169 samples





# Mer om vindusstørrelse

Vinduet matcher som regel ikke grunnfrekvensen i signalet. Da faller frekvenskomponentene mellom flere 'bins'.



# Oppsummeringspunkter for Fourier Transform

- Ethvert signal kan representeres som en sum av sinustoner med angitt amplitude/magnitude og fase.
- Man kan gå frem og tilbake mellom frekvensdomenet og tidsdomenet ved hjelp av FFT og IFFT
- Innenfor et vindu bestående av  $N$  samples har vi  $N/2$  frekvenskomponenter. Det vil si at for å få god frekvensoppløsning må vi ha et langt nok vindu.
- Dersom vinduet er langt betyr det dårligere oppløsning i tid.
- Spektrogrammer lages derfor vanligvis med *overlappende* vinduer (hop size)

# FFT i Matlab (fløyteeksemplet):

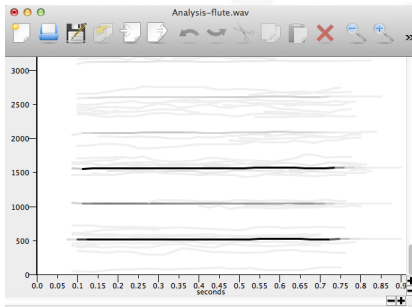
## Les inn, hør og plot:

les in lyden	<code>flute = wavread('flute.wav');</code>
hør på lyden	<code>sound(flute,44100);</code>
hør på litt av lyden	<code>sound(flute(5000:7000),44100);</code>
se på waveform	<code>plot(flute)</code>
zoom inn på waveform	<code>xlim([5000 7000])</code>

## FFT

Velg ut en del av lyden	<code>flute169 = flute(8000:8169);</code>
plot vinduet du har valgt	<code>plot(flute169);</code>
Gjør en FFT	<code>Y=fft(flute169);</code> <i>Variablen Y inneholder nå komplekse tall</i>
trekk ut amplitude og fase	<code>a = abs(Y);</code> <code>f = angle(Y);</code>
plot amplitude	<code>stem(a(1:85))</code>

FFT brukes blant annet i fasevocoder (eng. *Phase Vocoder*), som Michael Klingbeil's *SPEAR* illustrerer godt.



- Fasevocoderen analyserer signalet, og finner ut hvilke nabo-*bins* som er aktive.
- Ved å gjøre invers fourier transform kan man spille av lyden igjen.
- Ved å manipulere vindusstørrelsen kan man strekke lyden i tid uten at tonehøyden forvrenges.

# Litt om filtre (mer om dette 17. oktober)

- Et filter endrer amplitude og/eller fase
- Deler av frekvensspekteret kan forsterkes eller dempes.
- Selv om filtre ofte beskrives i frekvensdomenet, er digitale filtre som regel implementert i tidsdomenet ved å forsinke signalet og legge det sammen med det originale signalet.
- Filtre er overalt:
  - Ulike rom former lyden forskjellig
  - Øret / hodet / kroppen former den lyden vi hører
  - Eksempel: lyden fra en elgitarstreng går gjennom flere filtre:
    - Pickup
    - Tonekontroll
    - Gitarforsterker
    - Rom

# Litt om filtre (mer om dette 17. oktober)

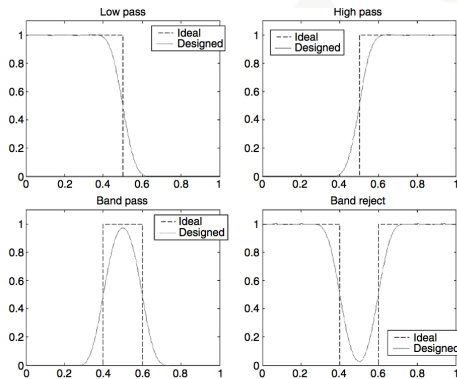


figure 1.7 Four basic filter types. Magnitude spectra illustrate the frequency responses for four standard filter types. The band-reject filter is also often called a band-stop or notch filter. Each plot shows the frequency response for an ideal filter, and the actual frequency response for some real finite impulse response filters approximating the ideal specification. In this diagram, the x-axis is marked in terms of 'normalized frequency', with a reference of the Nyquist rate, so 0=0 Hz and 1=half the sampling rate. This makes the diagram independent of the sampling rate.

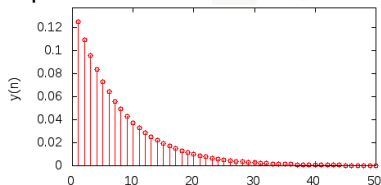
- I frekvensdomenet beskrives et filter med en *frekvensrespons*
- $Y = X \cdot F$

Y: Signal ut  
X: Signal inn  
F: frekvensrespons

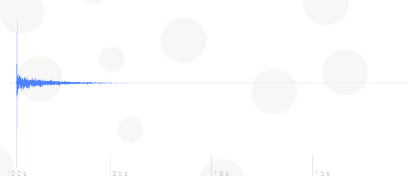
# Litt om filtre (mer om dette 17. oktober)

I tidsdomenet beskrives et filter med en *impulsrespons*:

Lavpassfilter:



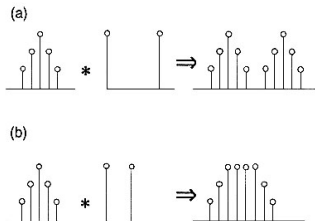
Katedral:



Multiplikasjon i  
frekvensdomenet tilsvarer  
*konvolusjon* i tidsdomenet.

$$Y = X * I$$

Y: Signal ut  
X: Signal inn  
\*: Konvolusjon  
I: impulsrespons



# Fysiologisk inspirerte modeller for å beskrive lyd

Siden vi ikke er i stand til å høre alle detaljer i lyden som når ørene våre er det interessant å modellere hørselssystemet vårt.

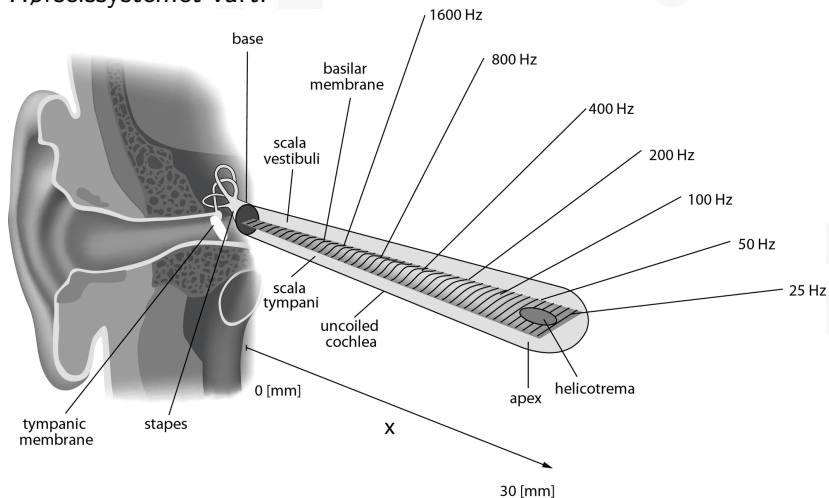
- 1 Tidlig prosessering: Modellere filtrering som skjer i det ytre øret, og mellomøret
- 2 Filterbankmodellering av det indre øret. Denne modellen er vanligvis en kraftig forenkling av virkeligheten. Øret vårt har 3500 hårceller, mens 64 eller 128 kanaler brukes i modellene.
- 3 Modellering av hårcellenes mekaniske bevegelse
- 4 Modellering av nervebanene fra hårcellene til hjernen
- 5 Modellering av prosesseringen som skjer i hjernen

(Se Collins, Tabell 3.1, side 94)



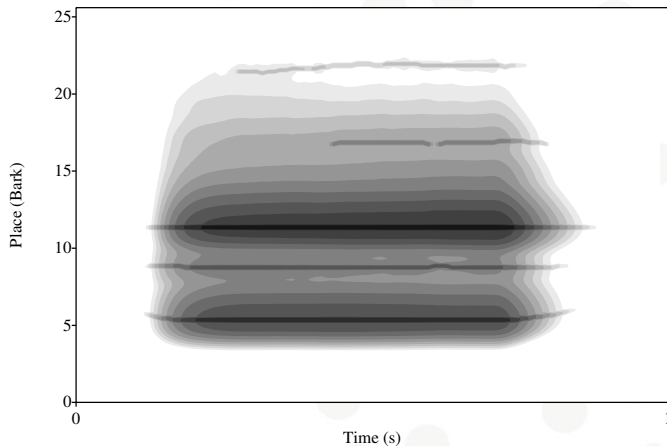
# Fysiologisk inspirerte modeller for å beskrive lyd

Hørselssystemet vårt:



# Fysiologisk inspirerte modeller for å beskrive lyd

Cochleagram av fløytelyden vår:



# Lyddeskriptorer

Table 3.3 Examples of low-level features.

Feature	Description	Calculation
ZCR	Count (positive) zero crossings within $N$ samples	$\sum_{k=0}^{N-2} x(k+1) \geq 0 \wedge x(k) < 0$
RMS	Root mean square amplitude calculated over $N$ samples	$\sqrt{\frac{\sum_{k=0}^{N-1} x(k)^2}{N}}$
Max power	Maximum power in a block of $N$ samples; often used in sample editor waveform displays when zoomed out	$\max_{k=0}^{N-1} x(k)^2$
Spectral centroid	Statistical measure over the spectrum	$\frac{\sum_{k=0}^{N/2-1} k  X_m(k) ^2}{\max(\sum_{k=0}^{N/2-1}  X_m(k) ^2, 1)}$
Spectral flux	Change of spectrum between frames	$\sum_{k=0}^{N/2-1}   X_{m+1}(k) ^2 -  X_m(k) ^2 $
Spectral fall-off	The spectral envelope can be modeled by fitting a curve to the magnitude spectrum. Spectral fall-off fits a single line to model the typical drop in energy at higher frequencies in sound, as one helpful timbral indicator, but more complex models are available	Rodet and Schwarz [2007]
LPC coefficients	Linear predictive coding models the spectrum of the input with a source-filter model; it is a useful compression technique	Gold and Morgan [2000]; Rabiner and Juang [1993]; Makhoul [1975]
MFCCs	Mel-frequency cepstral coefficients; given a spectrum, the <b>cepstrum</b> approximates the principal components, and is a useful timbre descriptor; it also deconvolves (separates) an excitation and body response and gives some idea of pitch	Gold and Morgan [2000]; Logan [2000]; Roads [1996, pp. 514–8]