MUS4218 - Metodologisk emne, kognitiv musikkvitenskap

Lydanalyse

9 mars 2017

Beskjeder

- Tilbakemeldinger fra presentasjonene forrige uke ligger på fronter.
- Oblig 3: Lydanalyse

  http://www.uio.no/studier/emner/hf/imv/MUS4218/v17/obligatoriske-oppgaver.html
- NM i stillstand

I dag: Learning outcomes

- Kjenne til noen generelle problemstillinger og prinsipper innen forskning på lyd og Music Information Retrieval
- Ha en konseptuell forståelse av Fourier Transform
- Kunne utføre enkle lydanalyser i Matlab

Orio (2006): *Music Retrieval: A Tutorial and Review.*

- Litt gammel, men god intro til MIR-fagfeltet.
- Paralleller til tradisjonell musikalsk analyse
- Digital representasjon av musikk
  - Bølgeform
  - Spektrum
  - Spektrogram
  - MIDI
- Deskriptorer sortert etter 'Time scales'
  - short term: Timbre, Orchestration, Acoustics
  - middle term: Rhythm, Melody, Harmony
  - long term: Structure

Forskning på lyd

- Sammenheng mellom det *sub-symbolske*, *symbolske* og *supra-symbolske*.
- MIR sjanger-klassifikasjon: Hvilke subsymbolske descriptorer er felles for en sjanger?
- MIR instrument-klassifikasjon: Hvilke subsymbolske descriptorer er felles for et instrument?
- Hvis vi kan trene en maskin til å kjenne igjen musikkinstrumenter/sanger/artister, så kan det gi oss en infallsvinkel til å studere hvordan mennesker sanser musikk.

Grunnleggende lydlære: Periodisitet

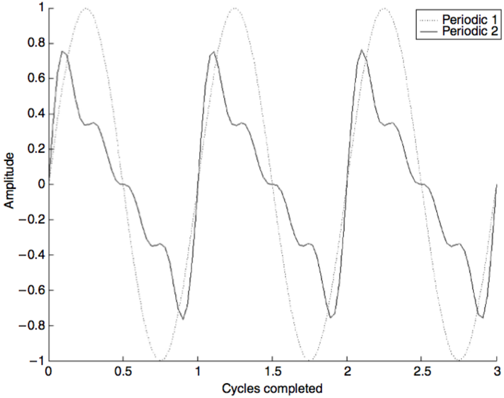
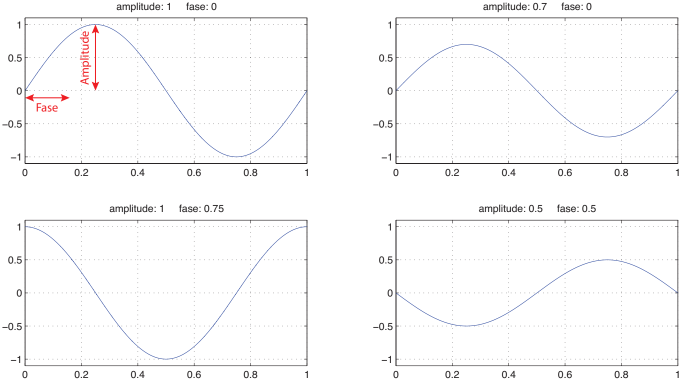


Figure 1.2 Periodic waveforms.

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

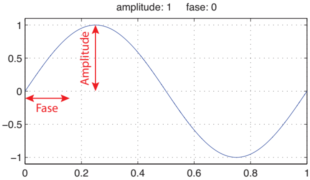
Grunnleggende lydlære: Amplitude og Fase

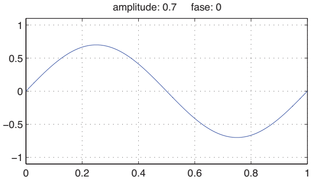
sinustoner gjengis med *amplitude* (også kalt magnitude) og *fase*:

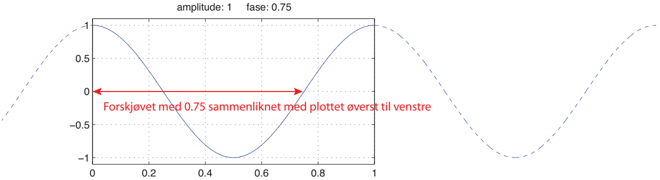


Grunnleggende lydlære: Amplitude og Fase

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







Grunnleggende lydlære: Grunnfrekvens og Overtoner

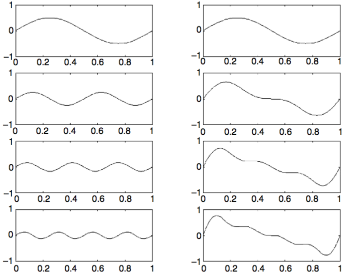


Figure 1.3 Sinusoidal components of a periodic sound. The left column shows each individual sine component (they vary in their amplitude). The right column gives the mix so far at each stage, as the sines are added together down the page.

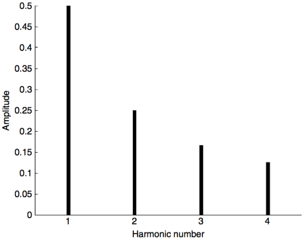
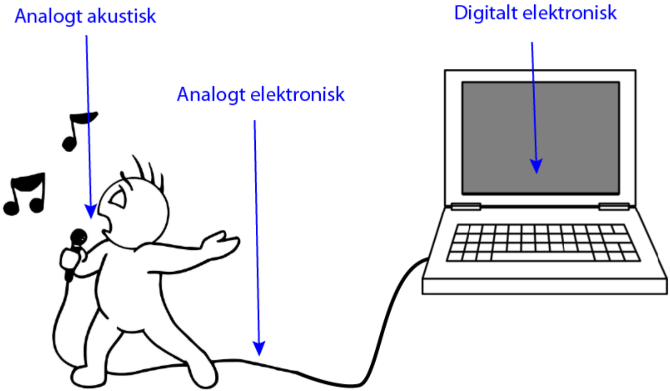


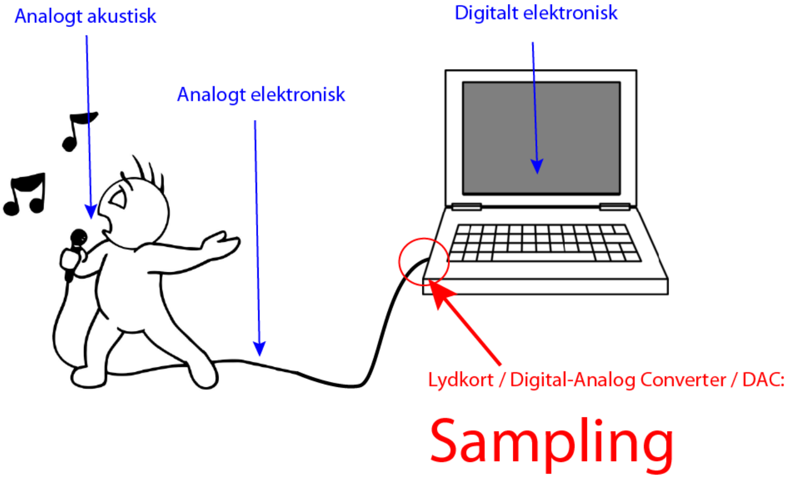
Figure 1.4 Line spectra. The frequency and amplitude of each component of the periodic sound are indicated (though phase information is dropped).

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

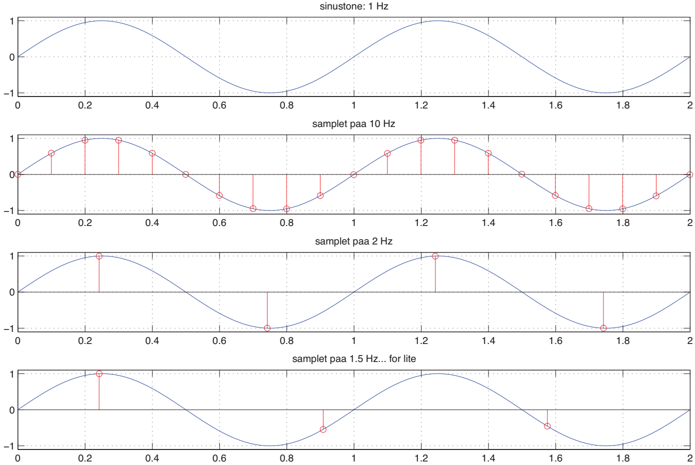
Grunnleggende lydlære: Analog $\rightarrow$ Digital



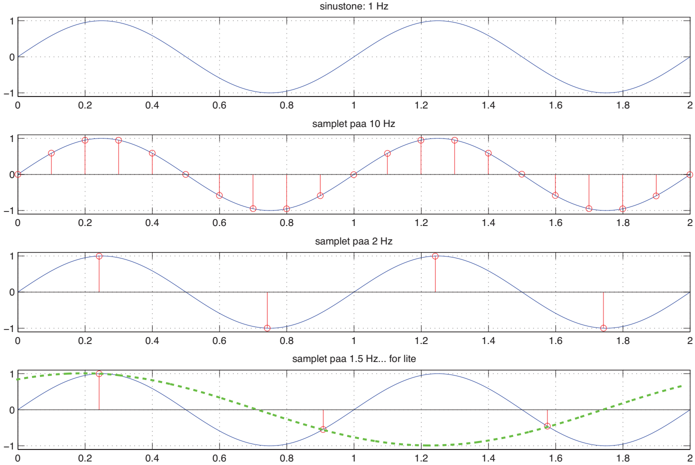
Grunnleggende lydlære: Analog → Digital



Grunnleggende lydlære: Sampling



Grunnleggende lydlære: Sampling



ALIASING: Signalet ser ikke lenger ut som en 1 Hz sinustone

Grunnleggende lydlære: Nyquist

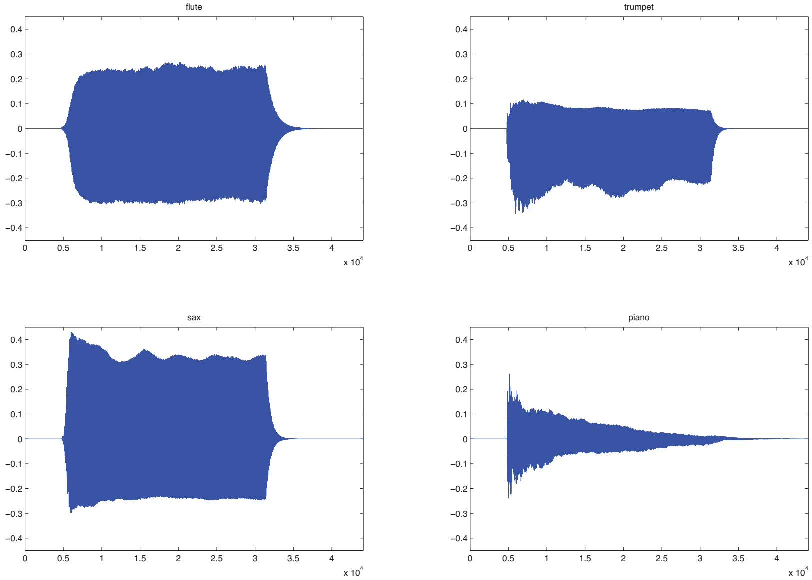
Nyquist–Shannons samplingsteorem (litt omskrevet):

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

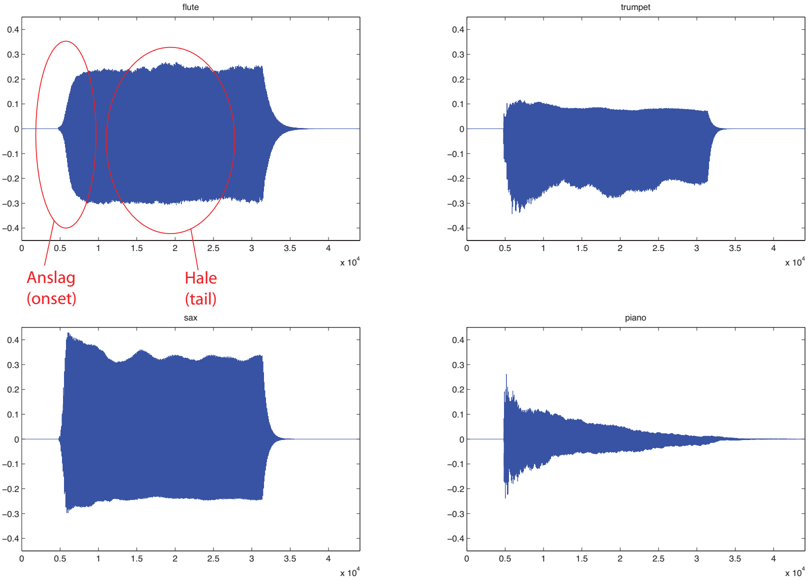
CD samplingsfrekvens: 44100 Hz
høyeste gjengivbare frekvens: 22050 Hz (Nyquistfrekvens)

Menneskelig hørsel: ca 20 – 20000 Hz

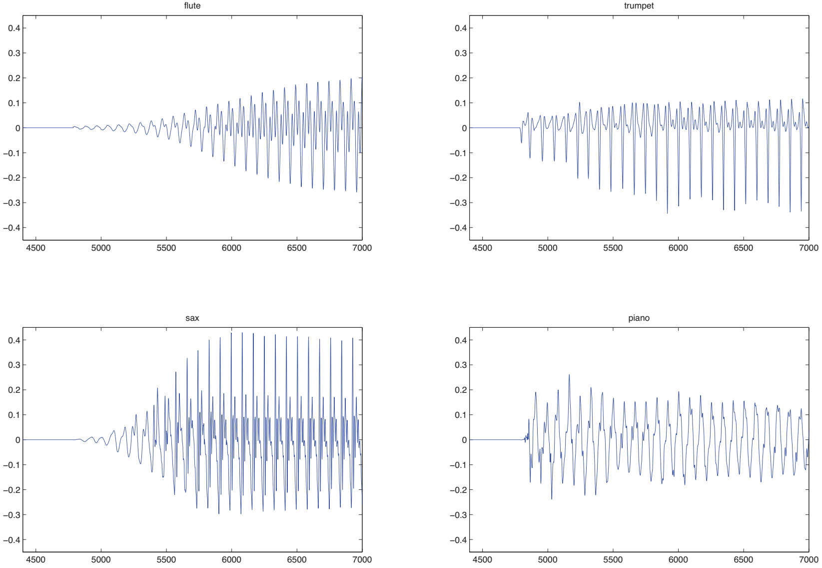
Noen “naturlige” (fra sampler) lyder i tidsdomenet



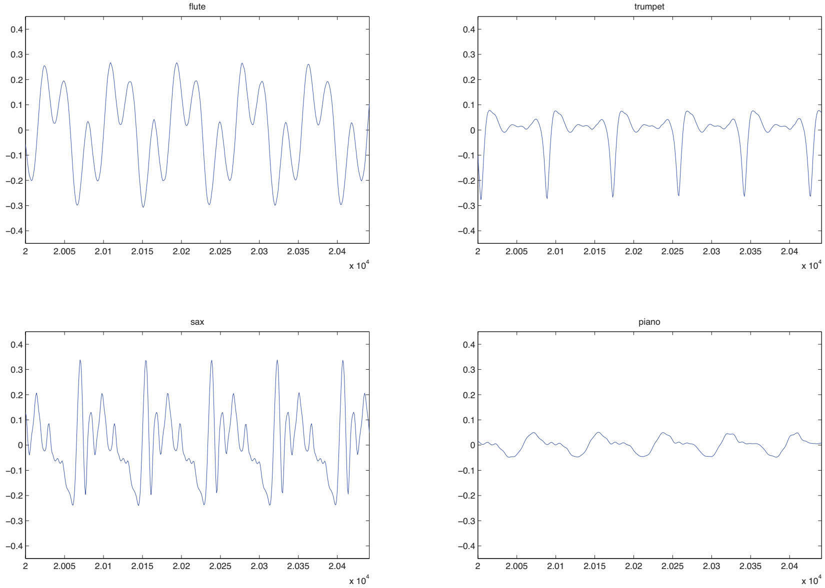
Noen "naturlige" (fra sampler) lyder i tidsdomenet



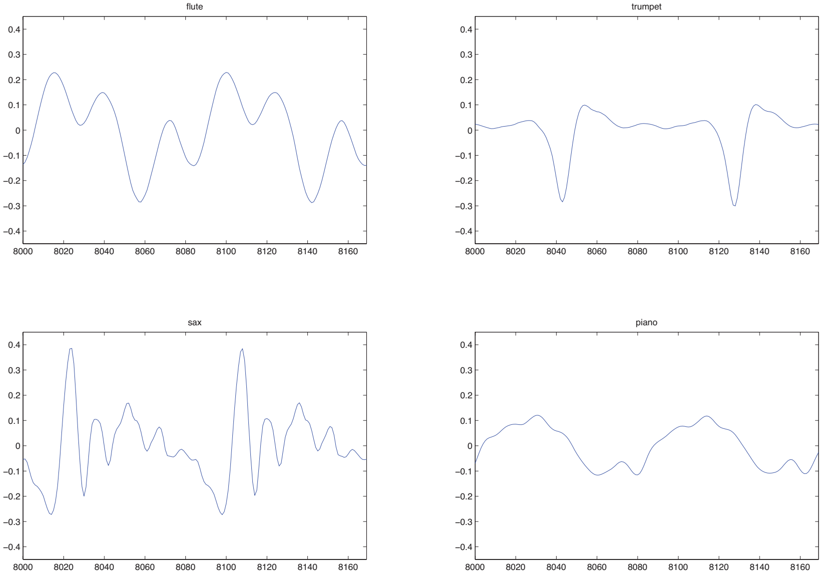
Vi zoomer inn på anslaget



Og på halen


flute
trumpet
sax
piano
0.4
0.3
0.2
0.1
0
−0.1
−0.2
−0.3
−0.4
2
2.005
2.01
2.015
2.02
2.025
2.03
2.035
2.04
x 10^4

Vi zoomer helt inn på kun to perioder



Fra tidsdomenet til frekvensdomenet: Fourier Transform

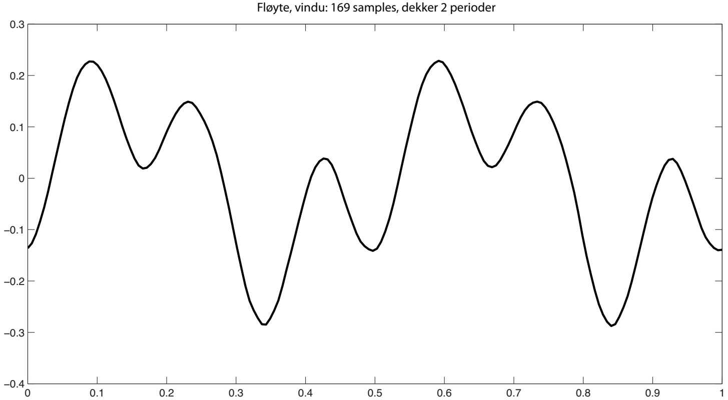
- Ethvert signal kan representeres som en sum av sinustoner med angitt amplitude og fase. (Spektrum)
- Vi ser vanligvis på et relativt kort signal for å finne frekvensinnholdet i dette. Dersom vi skal analysere et lengre signal deler vi det opp i overlappende “vinduer”.

Flere varianter.

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

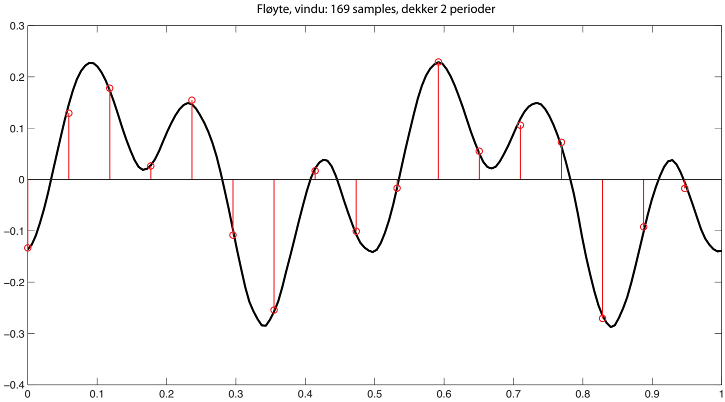
Det viktigste i dette kurset er ikke forskjellen på de ulike typene.

Fra tidsdomenet til frekvensdomenet: Fourier Transform



Her er vårt fløytesignal

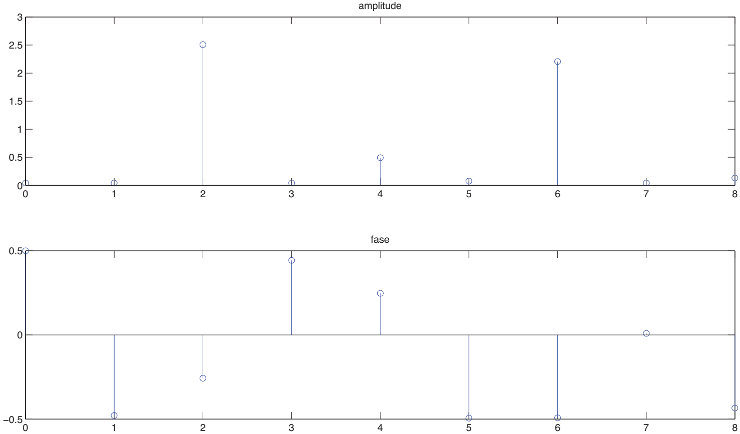
Fra tidsdomenet til frekvensdomenet: Fourier Transform



For enkelhets skyld nedsampler vi til et signal med 17 sampler

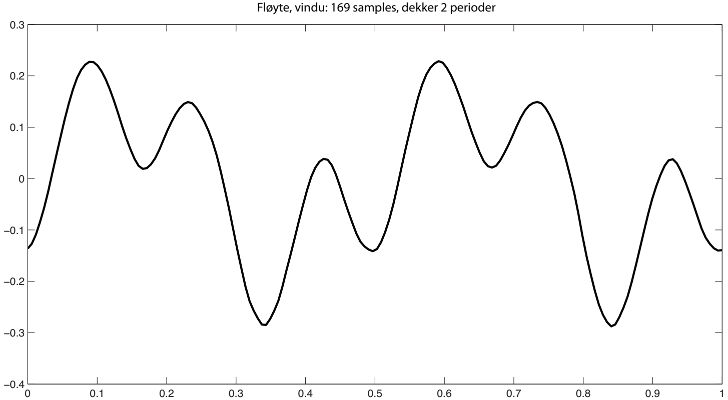
Fra tidsdomenet til frekvensdomenet: Fourier Transform

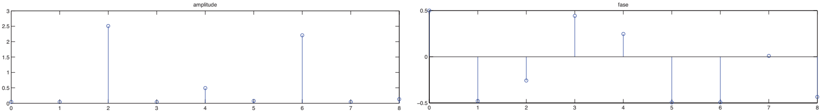
Resulatatet av å gjøre en Fourier Transform er to nye signaler:



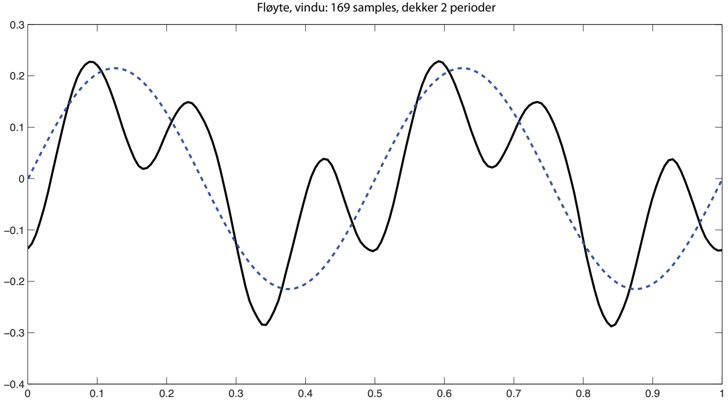
Her er ikke x-aksen tid, men 'bin'.

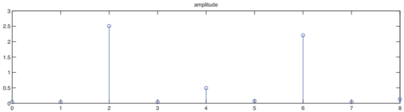
Fra tidsdomenet til frekvensdomenet: Fourier Transform

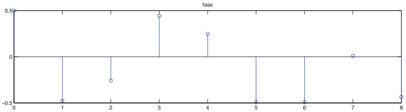




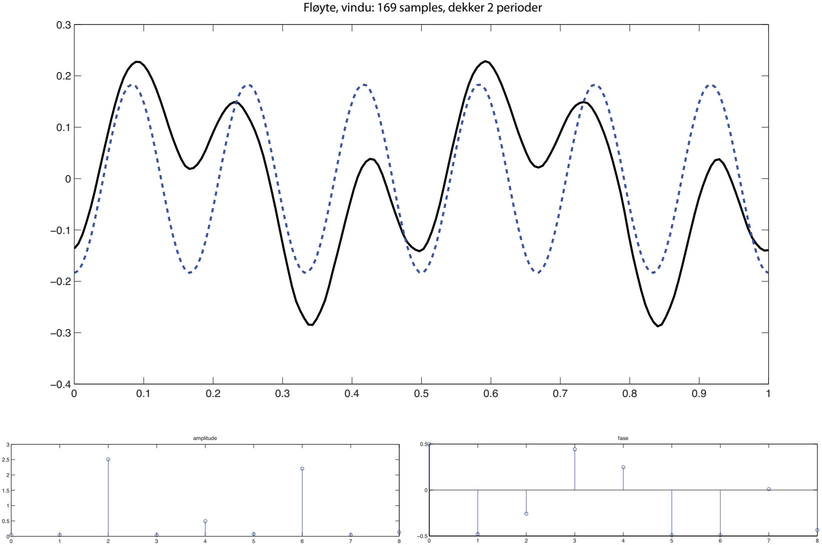
Fra tidsdomenet til frekvensdomenet: Fourier Transform



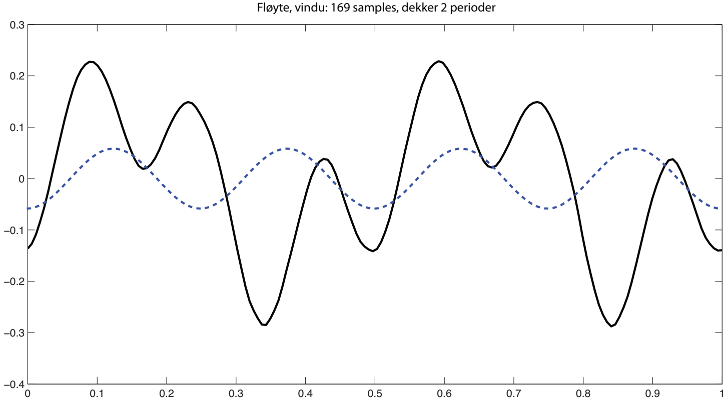


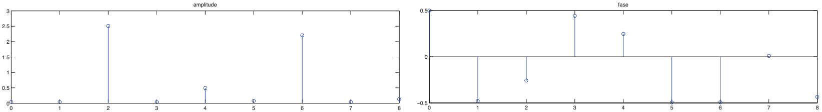


Fra tidsdomenet til frekvensdomenet: Fourier Transform

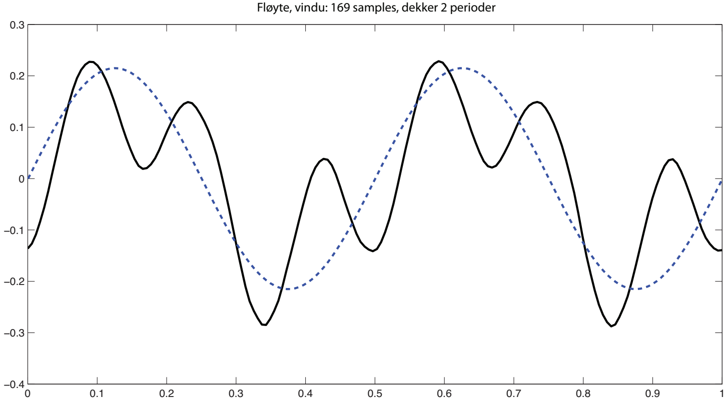


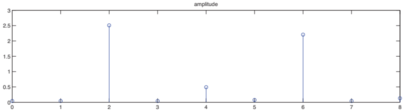
Fra tidsdomenet til frekvensdomenet: Fourier Transform

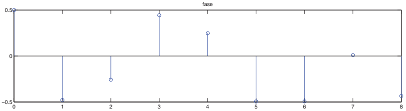




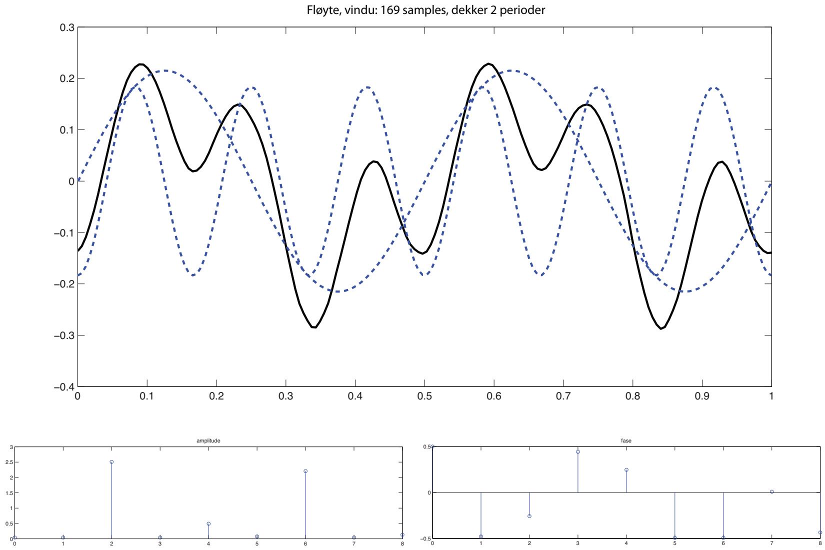
Fra tidsdomenet til frekvensdomenet: Fourier Transform



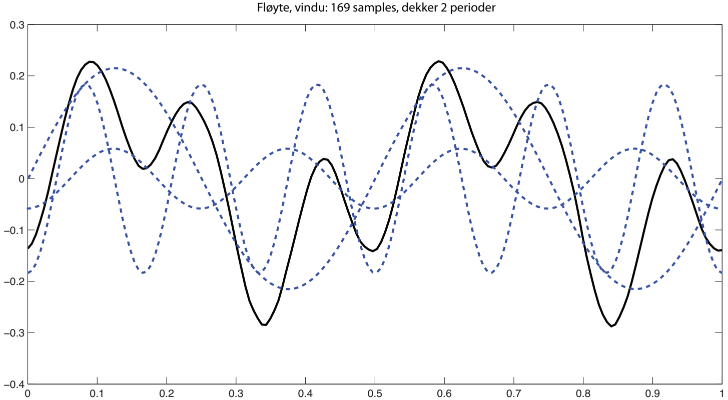


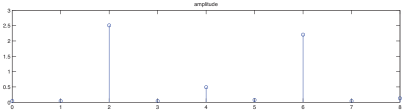


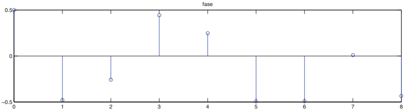
Fra tidsdomenet til frekvensdomenet: Fourier Transform



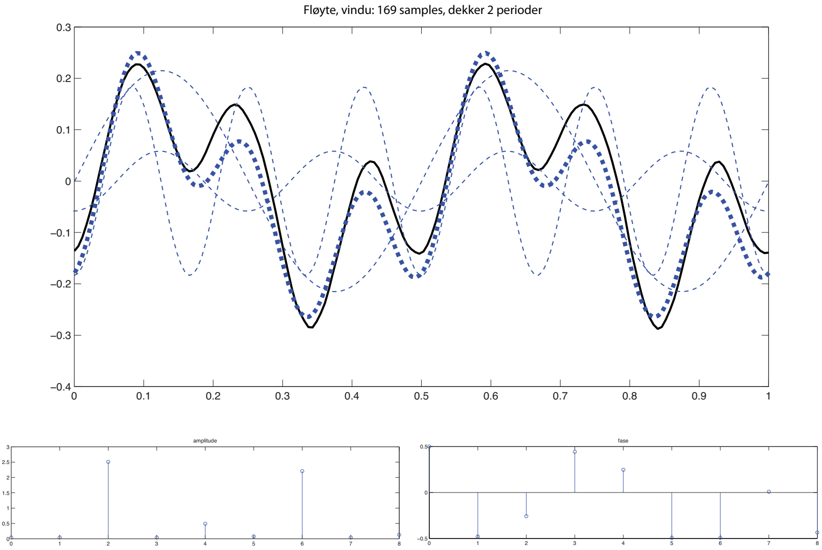
Fra tidsdomenet til frekvensdomenet: Fourier Transform



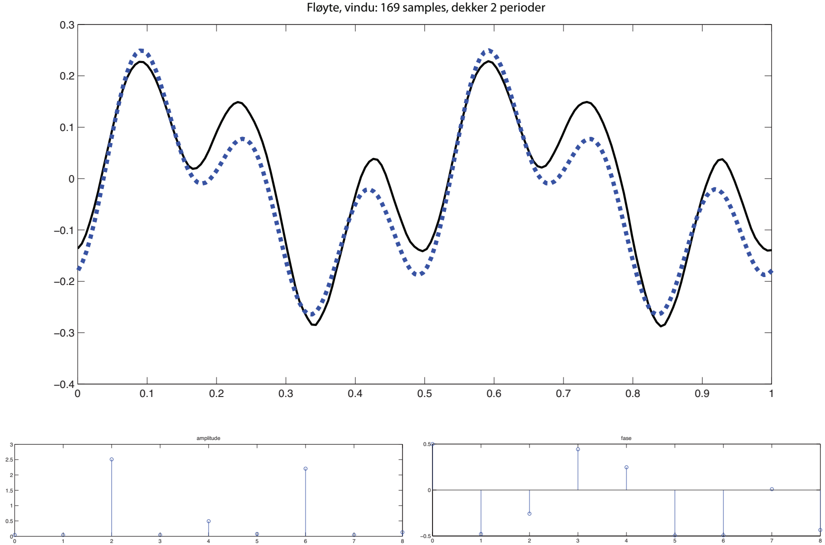




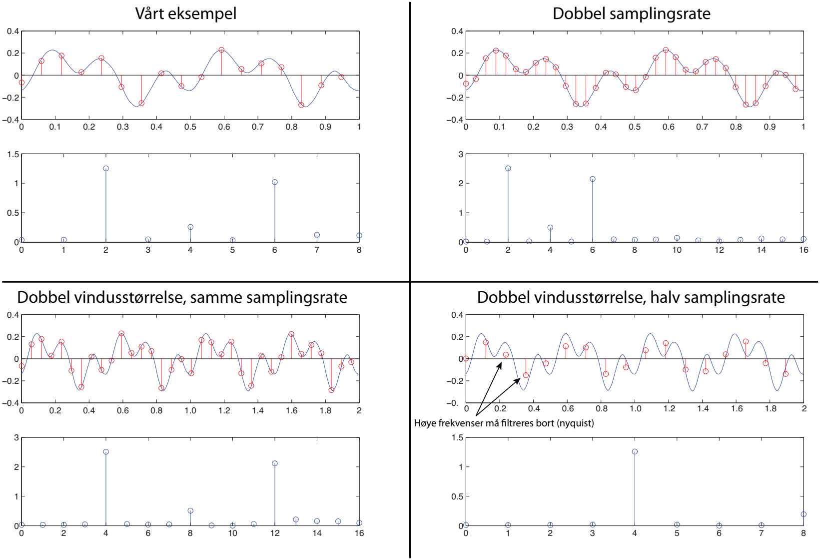
Fra tidsdomenet til frekvensdomenet: Fourier Transform



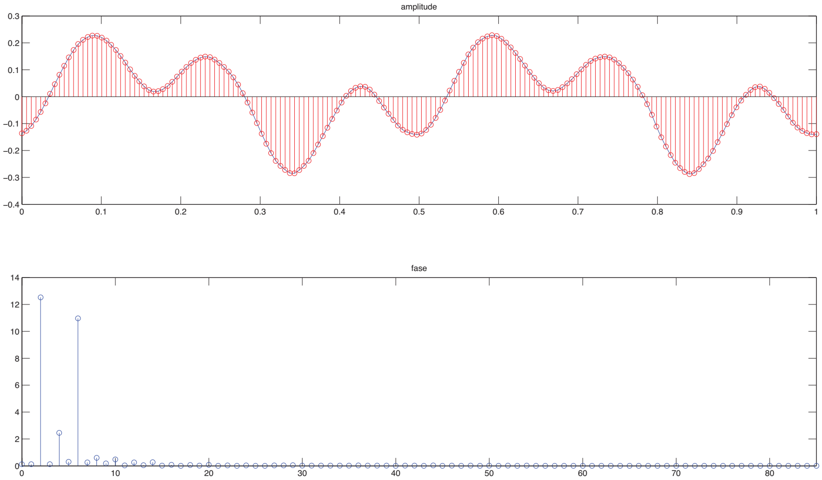
Fra tidsdomenet til frekvensdomenet: Fourier Transform



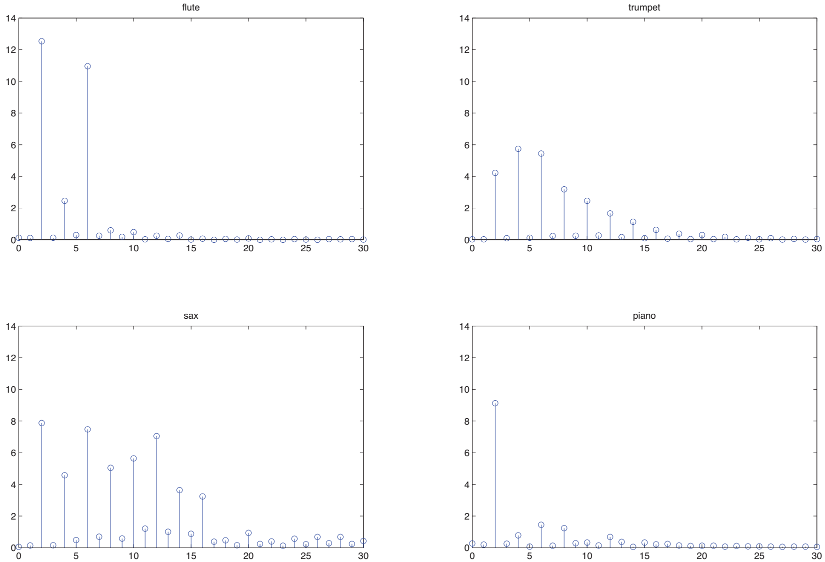
Effekten av å doble vinduslengden



Vanlivis nedsampler vi ikke:

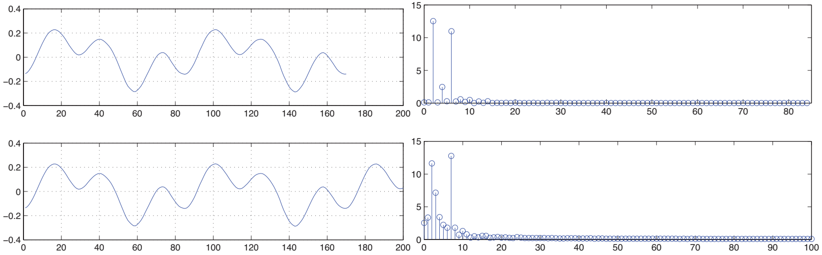


Fire instrumenter, full samplingsrate, vindu 169 samples

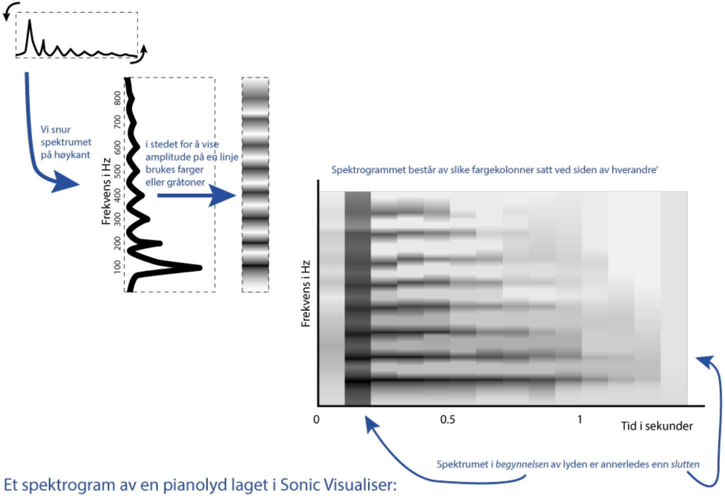


Mer om vindusstørrelse

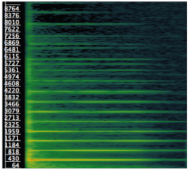
Vinduet matcher som regel ikke grunnfrekvensen i signalet. Da faller frekvenskomponentene mellom flere 'bins'.



Et **spektrogram** viser oss hvordan spektrumet endres over tid



Et spektrogram av en pianolyd laget i Sonic Visualiser:



sonicvisualiser.org

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)
- to fine videoer som forklarer FFT i programmet Max:
  - `https://www.youtube.com/watch?v=9gQAHf0Sf9I`
  - `https://www.youtube.com/watch?v=y0nlaHWEeaM`

FFT i Matlab (fløyteeksempelet):

Les inn, hør og plot:

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

FFT

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

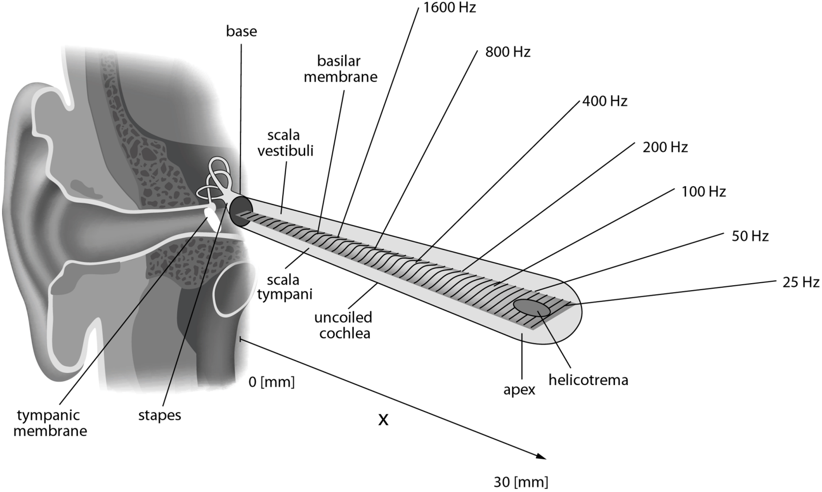
Fysiologisk inspirerte modeller for å beskrive lyd

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

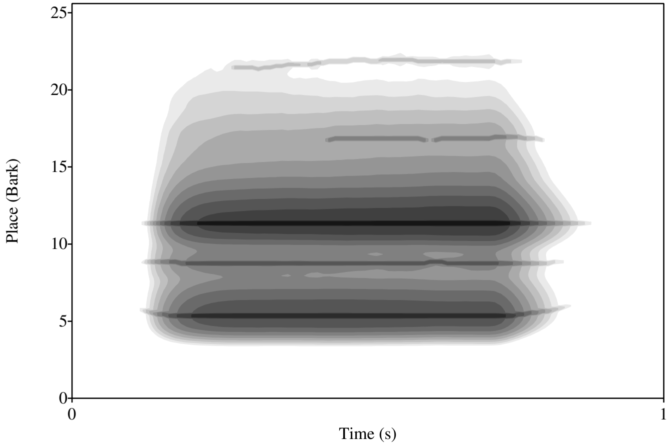
Fysiologisk inspirerte modeller for å beskrive lyd

Hørselssystemet vårt:



Fysiologisk inspirerte modeller for å beskrive lyd

Cochleagram av fløytelyden vår:



MIRtoolbox

- Toolbox (samling av funksjoner) for Matlab rettet mot lydanalyse og Music Information Retrieval
- Se MIRtoolbox primer (på pensumlisten) og manualen som følger med nedlasting av toolboxen for informasjon om installasjon osv.
- Nedlasting:

  `https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox/Download`

- På midirommet ligger MIRtoolbox ferdig nedlastet i mappen /Applications/matlab toolbox
- Noen grunnfunksjoner:
  - miraudio - les inn lydfil
  - mirplay - spill av lydfil
  - mirplayer - avspiller for lydfiler
  - mirframe - del opp en lydfil i 'frames'
  - mirfilterbank - del opp en lydfil i 'filterbanks'
  - mirgetdata - hent ut data fra et mirobject

Husk at du kan bruke hjelpefunksjonen i Matlab (F1)

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	$\frac{\sum_{k=0}^{N/2-1} k\lvert X_m(k)\rvert^2}{\max(\sum_{k=0}^{N/2-1} \lvert X_m(k)\rvert^2, 1)}$
Spectral flux	Change of spectrum between frames	$\sum_{k=0}^{N/2-1} \lvert\lvert X_{m+1}(k)\rvert^2 - \lvert X_m(k)\rvert^2\rvert$
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 **cepstrum** 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]

Noen analysemetoder og lyddeskriptorer i MIRtoolbox

- mirrms
- mirenvelope
- mirspectrum
- mirflux
- mirbrightness
- mircentroid
- mirroughness
- mironsets
- mirkey
- mirfeatures
- mirsimatrix
- mirzerocross

Repetisjon av beskjeder

- Tilbakemeldinger fra presentasjonene forrige uke ligger på fronter.
- Oblig 3: Lydanalyse

  http://www.uio.no/studier/emner/hf/imv/MUS4218/v17/obligatoriske-oppgaver.html
- NM i stillstand