

MUS4831 - Lydanalyse

Oppsummering

14 november 2013

Fra første forelesning:

Hovedmålsetning:

Å kunne knytte flest mulig forbindelser mellom musikkopplevelsen og lyden, med andre ord: å bli en enda bedre lytter

Table 1.1
Three Levels of Musical Experience

	Events per second	Seconds per event
EVENT FUSION (early processing)	16,384	1/16,384
	8,192	1/8,192
	4,096	1/4,096
	2,048	1/2,048
	1,024	1/1,024
	512	1/512
	256	1/256
	128	1/128
Functional units = individual <i>events</i> and <i>boundaries</i> ; pitches, simultaneous intervals, loudness changes, etc.	64	1/64
	32	1/32
	16	1/16
	8	1/8
	4	1/4
MELODIC and RHYTHMIC GROUPING (short-term memory)	2	1/2
	1	1
	1/2	2
	1/4	4
	1/8	8
Functional units = <i>patterns</i> ; rhythmic and melodic groupings, phrases.	1/16	16
	1/32	32
	1/64	1 min 4 sec
	1/128	2 min 8 sec
	1/256	4 min 16 sec
	1/512	8 min 32 sec
	1/1,024	17 min 4 sec
	1/2,048	34 min 8 sec
FORM (long-term memory)	1/4,096	1 hr 8 min 16 sec.
Functional units = large scale <i>constancies</i> ; sections, movements, entire pieces.		

- Fra Snyder (2000). Tidsskalaer for lydegenskaper og musikkrelaterte bevegelser.
- Vårt fokus i dette kurset er innenfor kategoriene “Event Fusion” og “Melodic and Rhythmic Grouping”

Pierre Schaeffer: “Écoute réduite” / Focused Listening

- *Sonic object*: Chunk of musical sound (≈ 0.5 to 5 seconds)
- Taking subjective impressions seriously:
‘What do you hear now?’
- *Acousmatic*: Not seeing the sound-source
- Typo-morphology based on subjective features
 - *Typology*: Different types/classes of sonic objects
 - *Morphology*: Internal features of the sonic objects

Typology and Morphology

Different types — Similar morphology



Schaeffer's klassifisering av lydobjekter

<i>Criteria of facture</i>				
<i>Criteria of mass</i>	Ordinary note	N	N'	N''
	Complex note	X	X'	X''
	Varied note	Y	Y'	Y''

- 3 typer opphav (criterion de facture)
 - impulsiv
 - utholdt
 - iterativ
- 3 typer for tone (criterion de masse)
 - stabil
 - kompleks
 - variabel

Lydeksempler fra
Solfège de l'objet sonore:
cd3 40-42

Multimodalitet

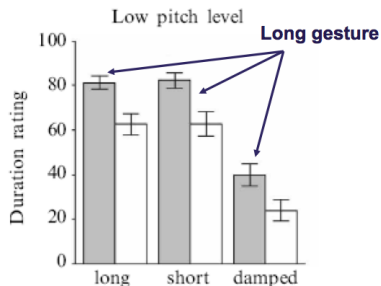
- Scaheffer's typologi peker klart mot lydperepsjon som *multimodal*
- Når vi hører lyder forstår vi lyden ved hjelp av vår erfaring med omverdenen (ecological listening):
 - Lydproduserende bevegelse
 - Lydproduserende objekt: materiale / form
- Visualisering av lyd: "Shapes" / "Sound-Tracing"

Multimodalitet

Perceived note length affected by visual movement information



“Long” and “short” strokes played on marimba shown to affect listeners rating of the duration of the tone.



(Schutz & Lipscomb, 2007)



AALBORG UNIVERSITY
DENMARK

fourMs
Music, Media, Motion, Modelling

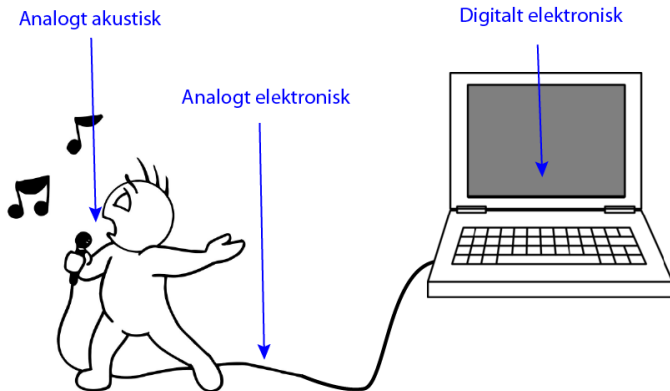
What can we learn from Schaeffer's theory:

- Focus on fragments of musical sound, i.e. on *sonic objects*; that's where we find very significant features
- Taking subjective impressions as point of departure
- Top-down subjective feature “questionnaire” approach
- Sonic objects as multidimensional entities
- Distinguishing as many features of musical sound as possible
- Western music theory limited in focus, very many features of musical sound not mentioned at all
- Western music theory mostly limited to notated music; very much music is not notated

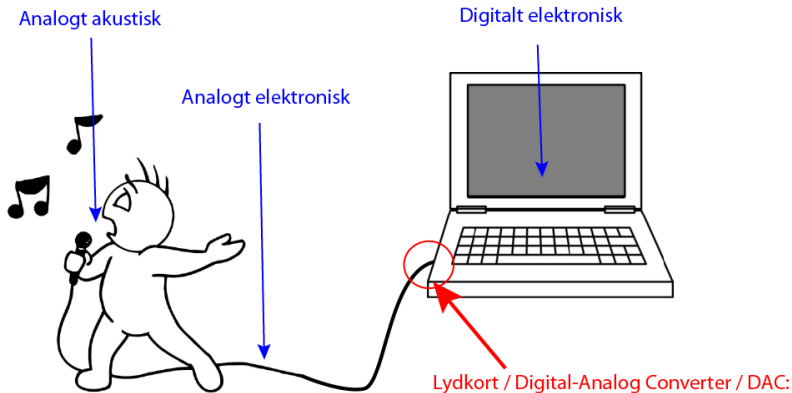
Hvordan studere lyd?

- Lytte
- Tid-/frekvensplot (Waveform/Spektrogram)
- Analyse gjennom syntese
 - Additiv syntese
 - Fysisk modellering
 - Kilde - Filter (impulsrespons)
- Feature extraction
- Mønstergjenkjenning/Klassifisering
- Kvalitative undersøkelser. Eksempler:
 - Hva gjør at akkurat denne lyden høres ut som en ?
 - Hvor mange overtoner trengs for å kjenne igjen lyden?
 - Hva skjer om man fjerner enkelte overtoner?
 - Hva skjer om man kombinerer anslaget fra en lyd med halen fra en annen?
 - Når går en lyd over fra å være impulsiv til å bli iterativ og videre til å bli utholdt

Analog → Digital

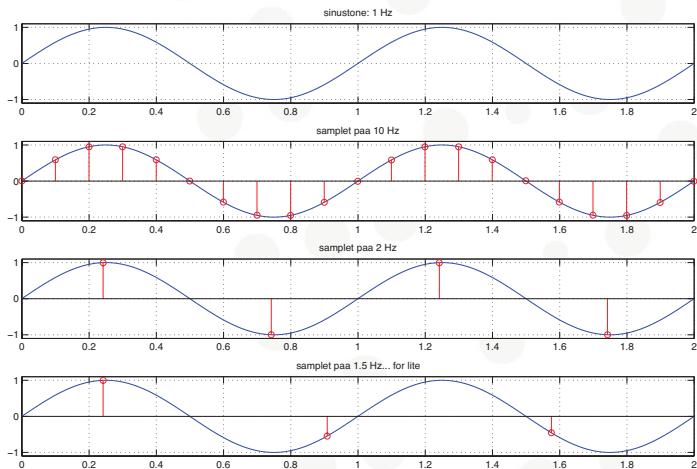


Analog → Digital

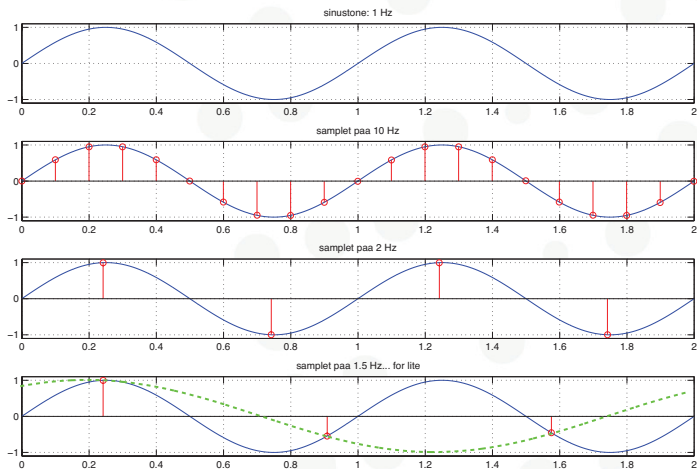


Sampling

Sampling



Sampling



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

Nyquist–Shannons samplingsteorem (litt omskrevet):

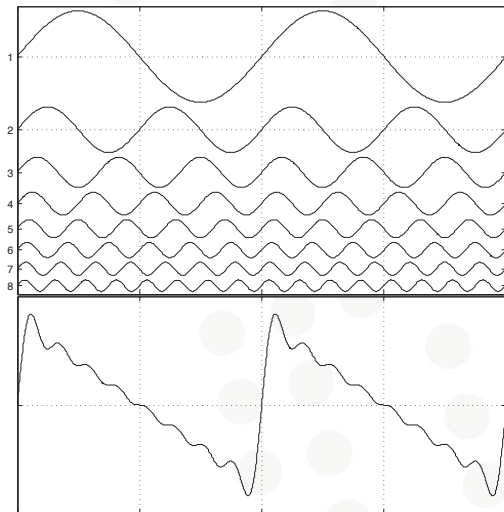
Hvis et signal inneholder frekvenskomponenter på opptil X Hz, må vi minimum bruke en samplingsrate på 2 ganger X Hz.

CD samplingsfrekvens: 44100 Hz

høyeste gjengivbare frekvens: 22050 Hz (Nyquistfrekvens)

Menneskelig hørsel: ca 20 – 20000 Hz

En bølgeform kan bygges opp av “partialer”



Analyse gjennom syntese

Idé:

- Noen fenomener kan være vanskelig å analysere ved hjelp av vanlige signalbehandlingsmetoder.
- For eksempel er *timbre*^a et flerdimensjonelt fenomen som er vanskelig å beskrive eksakt.
- Hvis vi ikke klarer å analysere oss frem til hva det er i signalet som gjør at en lyd høres ut slik den gjør kan vi ty til analyse gjennom syntese.
- Hvis vi klarer å bruke veldig enkle komponenter til å bygge opp en lyd som likner på lyden vi ønsker å analysere kan vi studere vår egen modell for å analysere lyden.

^a*Klangfarge* er et alternativt (men litt upresist) ord for *timbre* på norsk

Analyse gjennom syntese

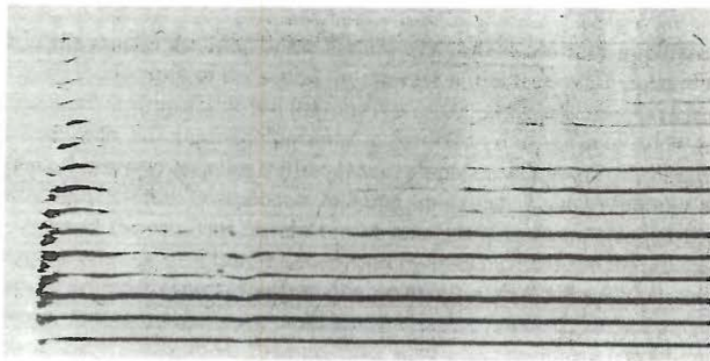


Figure 1.3

Sound spectrogram of an A4 trumpet tone with evolving dynamics, described by the markings *sf p cresc...f*. This represents frequency versus time—the portion represented lasts 1.7 s and goes up to 7.0 kHz in frequency. The amplitude of a component at a given time is indicated by its darkness. One can also see that the spectrum extends with loudness.

Analyse gjennom syntese

Additiv syntese: bygge opp lyden fra overtoner

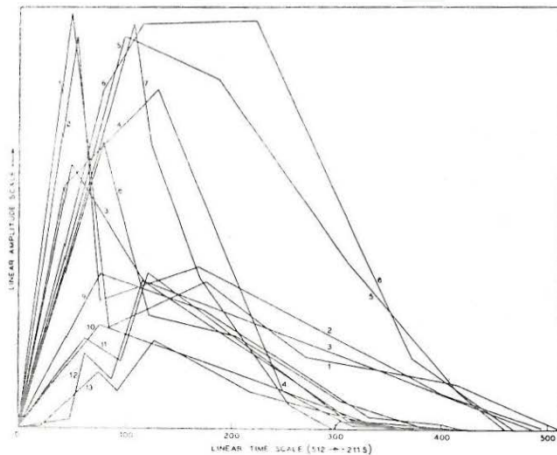
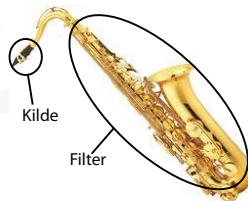
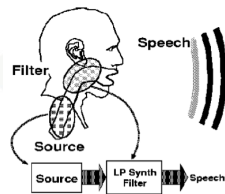


Figure 1.4
Plot of functions schematizing the envelopes of individual harmonics for a D4 trumpet tone of duration 0.21 s. These functions are drawn with a linear amplitude scale. They are made up of line segments. These functions have been used to control the harmonic amplitudes of synthetic tones.

Analyse gjennom syntese

- Lyd kan også analyseres ved å se på lydproduksjonen som en kilde og et filter.
- For eksempel stemmen.
 - Kilden (direkte fra stemmebåndene) kan litt forenklet representeres som en trekantbølge med varierende frekvens.
 - Filteret (munnhule) er et formantfilter som styrer hvilke vokaler vi tolker lydene som.
 - I sin enkleste form blir dette som regel en ganske middelmådig gjengivelse av lyden, fordi kilden i seg selv også kan fungere som et filter.
- Impulsrespons → Konvolusjonsfilter



Music Information Retrieval

- Trekke ut mening fra musikkrelaterte dokumenter (audio, midi, video, tekst, +++)
- Metadata (komponist, plateselskap, ...)
- Lavnivå (spektral centroide) / Høynivå (uttrykte emosjoner)
- Søkemotorer
 - Vanlig tekstbasert søk og automatisk tagging av lydklipp
 - Query by example
 - Query by gesture
 - Query by humming
 - Query by tapping
- Utfordringer:
 - Finne likheter mellom to veldig forskjellige lydklipp av samme type. (Hva gjør at vi hører at to ulike klappelyder er klappelyder)
 - Finne ulikheter mellom to veldig like lydklipp av forskjellig type. (For eksempel en kontrafagott og en kontrabass)
 - Kjenne igjen musikalske stiler/sjangere

Classification of emotions expressed by music (simplified)

- Researchers at University of Jyväskylä (Finland) asked people to evaluate the emotional content of a large number of music excerpts.
- Each sound file would get a score on two scales: *activity* and *valence*:



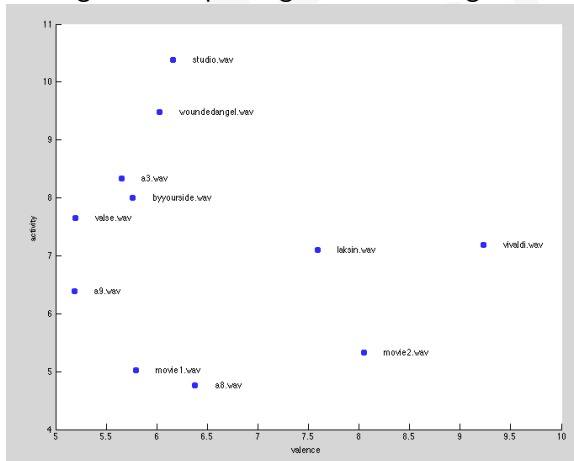
- By using the participants' answer as a ground truth, and extracting a large number of features, the researchers were able to train an algorithm to estimate the *expressed mood* of the music.

More about this:

Eerola, T. (2012). Computational modelling of emotions conveyed by music. *Topics in Cognitive Science*, 4(4), 607-624

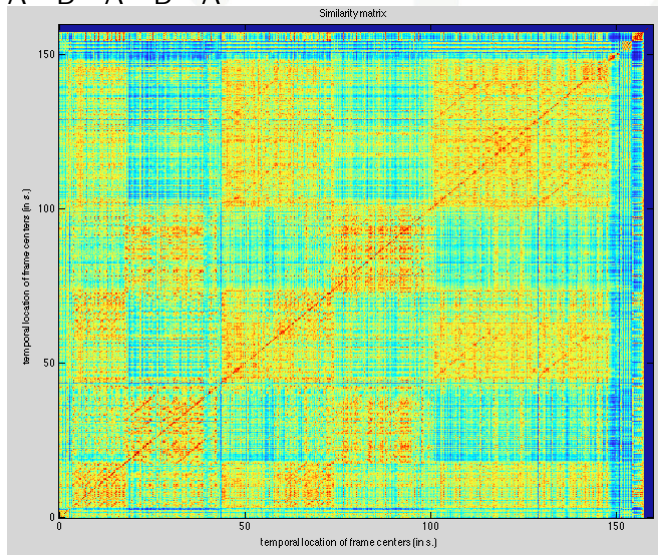
Classification of emotions expressed by music

Min egen lille utprøving av Eerola's algoritme:



Kjenne igjen formen i en låt: Self-similarity matrix

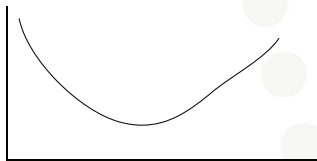
A - B - A - B - A



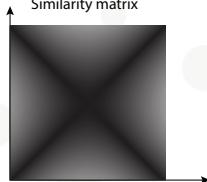
Kjenne igjen formen i en låt: Self-similarity matrix

- Selv om self-similarity matrix ofte gjøres på spektrumet i hver frame, kan det også gjøres på andre features.

Spektral centroide



Similarity matrix



Semesteroppgaven

Litt veiledning for semesteroppgaver i MUS4831

1. Dokumentasjon.

- Det du har gjort bør være så godt beskrevet at noen andre kan gjenta eksperimentet.
- Utvetydig
- Gjerne med bilder / video dersom dere gjør egne opptak.
- Hva slags utstyr er brukt?
- Hva slags programvare er brukt?
- Har du brukt tillegg til standarddistribusjonen av programmene? (f.eks. MIR toolbox til matlab)
- Hvilke innstillinger bruker du? (f.eks. vinduslengde)
- Hvorfor bruker du disse programmene/tilleggene/innstillingene?
- Forklar forkortelser første gang de brukes (RMS,ASDR,...)

Litt veiledning for semesteroppgaver i MUS4831

2. Referanser.

Tenk som i masteroppgaven. Vær nøye med referanser. La det være helt tydelig hva som er ditt eget arbeid, og hvilken informasjon du bygger på fra andre steder. Sammenlikn helst ditt eget arbeid (både resultater og metode) med det andre har gjort. Legg ved referanser til programvare.

I semesteroppgaven kreves det ikke referanser til programvare som er en del av pensum, men det oppfordres likevel til å inkludere det. (Og uansett om dere har med referanser eller ikke til programvaren må det selvsagt nevnes hvilken programvare som er brukt).

- Boersma, Paul and Weenink, David (2013). *Praat: doing phonetics by computer* [Computer program]. Version 5.3.56, retrieved 17 September 2013 from <http://www.praat.org/>
- Klingbeil, Michael (2005). "Software for Spectral Analysis, Editing, and Synthesis" in *Proceedings of the International Computer Music Conference*, pp. 1–4.

Litt veiledning for semesteroppgaver i MUS4831

3. Skrivefeil

En og annen sporadisk skrivefeil trekker ikke ned, men mange gjennomgående skrivefeil kan trekke en oppgave ned med en hel karakter. Bruk stavekontroll, eller eventuelt få noen andre til å lese over oppgaven din og se etter skrivefeil.

4. Tolking av plot

Følg med på verdien på aksene på plottene når du sammenlikner to eller flere plot. Ikke stol blindt på automatiske analyser. (MIR toolbox gjør "feil". For eksempel vil vindusstørrelsen som brukes i de underliggende funksjonene påvirke resultatet fra tempoanalyser)