

Innføring i MIR,
'Music Information Retrieval'

Music Information Retrieval:

- MIR = det å trekke ut informasjon av alle typer musikk-relaterte dokumenter, dvs. både noter (symbol-basert), tekst, lydifiler, og multimediafiler
- Kan dreie seg om forskjellig type informasjon, dvs. alt fra ulike typer såkalt 'metadata' (tekster som angir verknavn, komponist, musikere, produksjonsdetaljer, genre, osv. og/eller ulike opplevelsesattributter) eller mer notenære og/eller lydnære attributter
- Reiser derfor en rekke grunnleggende spørsmål om musikk
- Delvis nært forbundet med 'musikkanalyse' i mer tradisjonell forstand, men nå med ny teknologi

ISMIR2010-emner (International Society for Music Information Retrieval):

Music perception and cognition
Musical knowledge and meaning
Content-based querying and retrieval
Automatic classification
Music recommendation and playlist generation
Fingerprinting and digital rights management
Score following, audio alignment, and music synchronization
Transcription and annotation
Music summarisation
Music structure analysis
Optical music recognition
Music signal processing
Libraries, archives and digital collections
Database systems, indexing and query languages
Text and web mining
Compression and streaming
Modification and transformation of music data
Evaluation of mir systems
Knowledge representation, social tags, and metadata
Melody and motives
Harmony, chords and tonality
Rhythm, beat, tempo and form
Timbre, instrumentation and voice
Genre, style and mood
Performance analysis
Similarity metrics
Computational musicology
User interfaces and user models
Emotion and aesthetics
Applications of mir to the performing arts and multimedia

Ontologi

- Helt grunnleggende i MIR (og lydforskning generelt) er refleksjoner om ontologi, dvs. hvilken dimensjon av virkeligheten vi velger å forkusere på
- Generelt: alt kan sees på som statistiske fordeling, og mange likheter her med annen innholdsbasert datasøk, f.eks. bilder, tekster, bevegelser, osv.
- Orios differensieringer er veldig gode og velformulerte, og det vil være nyttig å lage sammendrag av dem (oppgave nr. 3 - oppgave nr. 4 vil være mer signal-basert og relatert til MIR)
- Dagens forelesning har fokus på Orios MIR og dernest en oversikt over div. PD-objekter/patcher

Bakgrunn for MIR

- Musikk tilgjengelig over alt
- Behov for å kunne finne fram i det store mangfoldet av musikk
- Teknologi muliggjør nå søk i musikk som tidligere ikke var mulig, selv om det gjenstår mange utfordringer her
- MIR må sees i sammenheng med det mer generelle feltet Information Retrieval (IR) og den enorme mengden forskning og utvikling som finnes der, dvs. en del av metodene i MIR er overført direkte fra IR i forskjellige sammenhenger

Bakgrunn for MIR

- Grunntanke: Innholdsbasert IR i MIR:
- "The main idea underlying content-based approaches is that a document can be described by a set of features that are directly computed from its content." (Orio 2006, s. 2)
- Og i forhold til IR generelt: "the underlying models are likely to describe fundamental characteristics being shared by different media, languages, and application domains" ... "For this reason, the research results achieved in the area of IR, in particular in the case of text documents, are a continuous reference for MIR approaches." (ibid)

Bakgrunn for MIR

- Og vi trenger innholdsbasert MIR: "The basic assumption behind content-based approaches is that metadata are either not suitable, or unreliable, or missing. In the case of MIR, all the three assumptions may hold." (s. 3)
- Men: Det finnes også visse muligheter for metadata, f.eks. på CDer og ved digital kringkasting
- Vår interesse her (MUS4831) er å se på utfordringer i, og metoder for, å knytte forbindelser mellom lyd som signal og ulike musikk-relaterte begreper, jfr. Schaeffers prosjekt om å kartlegge *korrelasjoner* signal-begrep

Musikkbegrep i forhold til MIR:

- Utgangspunkt i psykoakustiske begreper, dvs. *tonehøyde* (og ikke frekvens), *intensitet* (og ikke amplitude), og *timbre* (og ikke stasjonært spektrum), samt en redgjørelse om utfordringene ved slike komplekse sansekvaliteter
- Videre, begreper om tempo, tonalitet, metrum og faste fortegn, samt notesymbol-representasjon
- Men: Også mulig å ta utgangspunkt i det sub-symboliske, f.eks. ved å kalkulere liket (similarity) bare basert (helt blindt) på lavnivå statistiske egenskaper
- Derfor spørsmålet om hva innholdet (dvs. i forhold til innholdsbasert MIR) i musikken er for noe

Musikkbegrep i forhold til MIR:

- Dimensjoner i musikken:
- Timbre, både instrumenter og spillemåter
- Instrumentasjon, ulike kombinasjoner
- Akustikk, romakustikk og div. prosesseringer
- Rytme, dvs. alle typer mønstre i tid
- Melodikk, dvs. gjenkjennbare konturer
- Harmonikk, både eksplisitt og implisitt
- Musikalsk form, dvs. strukturer over lengere tidsstrek
- Skille mellom korttidsegenskaper (det vi kaller *chunk level* og *sub-chunk-level*) og egenskaper ved lengere tidsstrek (det vi kaller *supra-chunk level*)
- Tanken om et *Snapshot* av lyd, dvs. et øyeblikksbilde (chunk level eller sub-chunk level) som særlig signifikant (Gjerdingen & Perrott 2008)

Musikkbegrep i forhold til MIR:

- Orios oppsummering av tidsskalaer (s. 14):

| Time scale | Dimension | Content |
|-------------|---------------|----------------------------------|
| Short term | Timbre | Quality of the produced sound |
| | Orchestration | Sources of sound production |
| | Acoustics | Quality of the recorded sound |
| Middle term | Rhythm | Patterns of sound onsets |
| | Melody | Sequences of notes |
| | Harmony | Sequences of chords |
| Long term | Structure | Organization of the musical work |

- Som en test: Lydeksempler til kap. 12 i Sethares bok

- Og en nyttig oppsummering fra Snyder (2000):

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 |
| | 64 | 1/64 |
| | 32 | 1/32 |
| MELODIC and RHYTHMIC GROUPING (short-term memory) | 16 | 1/16 |
| | 8 | 1/8 |
| | 4 | 1/4 |
| | 2 | 1/2 |
| | 1 | 1 |
| | 1/2 | 2 |
| | 1/4 | 4 |
| | 1/8 | 8 |
| FORM (long-term memory) | 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 |
| 1/4,096 | 1 hr 8 min 16 sec. | |

Musikkbegrep i forhold til MIR:

- Og:
- "A musical **score** is a structured organization of symbols, which correspond to acoustic events and describe the gestures needed for their production."
- Og:
- "A musical **performance** is made of a sequence of gestures performed by musicians on their musical instruments; the result is a continuous flow of acoustic waves, which correspond to the vibration induced on musical instruments or produced by the human voice."
- Derfor noen betraktninger om performance vs. score

Musikkdokumentformater:

- Informasjon i notene (score): Symboler for hva som skal gjøres og når
- Informasjon i framføringer (performance): Hva musikerne faktisk gjør og de lydlige resultatene av det de gjør
- Derfor noen betraktninger om formater for musikkdokumenter ("dokument" her = enhver samling av data)
- Symbolske formater:
- Noter
- Andre symbolformater (f.eks. MUSIKODE og XML)
- MIDI, og ulike representasjoner av disse (f.eks. pianola)

Musikkdokumentformater:

- Audioformater
- Basert på PCM, dvs. samplede formater
- AIFF, WAVE, AU, osv., samt MP3
- Generative representasjoner i MPEG-7 ved å kode syntesemodeller/-parametre (egentlig også MIDI som et generativt format, faktisk med 'perfekt' gjengivelse i noen tilfeller som f.eks. diskklavier og elektromekanisk pipeorgel)
- Standard Midi File og General MIDI

Brukernes rolle:

- Brukernes ekspertise avgjørende for IR-effektivitet
- Kategorier brukere: "(i) casual users want to enjoy music, listening and collecting the music they like and discovering new good music; (ii) professional users need music suitable for particular usages related to their activities, which may be in media production or for advertisements; (iii) music scholars, music theorists, musicologists, and musicians are interested in studying music." (s. 28)
- Casual users: "Find me a song that sounds like this", "Given that I like these songs, find me more songs that I may enjoy", og "I need to organize my personal collection of digital music...."

Brukernes rolle:

- Profesjonelle brukere: "I am looking for a suitable soundtrack for...", eller: "Retrieve musical works that have a rhythm (melody, harmony, orchestration) similar to this one", osv.
- Musikkteoretikere, musikkforskere, musikere: Alle slags søkekriterier
- QBE (query by example) QBH (query by humming) og lignende input er utviklet, dog ennå ikke gode gestikk-baserte løsninger, noe vi vil arbeide med i SMA-prosjektet

Prosessering av musikken:

- Målsetting: Trekke ut relevante innholdsdeskriptorer
- Åpenbart store utfordringer i det å skulle trekke ut nyttige innholdsdeskriptorer fra audiodokumenter
- Symbol-baserte dokumenter: Uttrekking av hovedmelodi (forgrunnsmelodi) i flerstemmige satser,
- Segmentering av melodier, f.eks. ved bruk av N-grams: "A simple segmentation approach consists on the extraction from a melody of all the subsequences of exactly N notes, called N-grams. N-grams may overlap, because no assumption is made on the possible starting point of a theme, neither on the possible repetitions of relevant music passages. The idea underlying this approach is that the effect of musically irrelevant N-grams will be compensated by the presence of all the relevant ones." (s. 42, see also <http://en.wikipedia.org/wiki/N-gram>)
- Harmonikk, oftest avhengig av segmentering

Prosessering av musikken:

- Audio-baserte trekk:
- Timbre: Spektrogrammer, men også andre representasjoner som attack-tid, spektral centroide, roughness, og mel-frequency cepstral coefficients (MFCC), dvs. spektrum av et spektrum fordelt langs en mel-skala, men debatt om dette er en god representasjon for musikalsk lyd eller ei
- Instrumentasjon: Fryktelig vanskelig, faktisk et avansert Auditory Scene Analysis-problem, jfr. f.eks. Klappuris arbeid.
- Rytme: Tempo tracking, beat tracking, osv., basert på enveloper (i tidsdomene), selvlikhetsmatriser, autokorrelasjon, men også spektrale egenskaper. Såkalt "p-center" problematisk, cf. Wright 2008

Prosessering av musikken:

- Audio-baserte trekk:
- Melodikk, basert på pitch-tracking, f_0 , segmentering, osv.
- Harmonikk, basert på spektralanalyse med tillegg av skjema for den kromatiske skala, dvs. korrelere spektralkomponenter med antatte tonehøydemønstre for akkorder
- Andre dimensjoner, f.eks. romakustikk og div. effektsprossesering som inngår i audiodokumentet (ser ut til å være et lite utviklet felt)
- Formtrekk (storform i musikken), utviklet i senere tid, blant annet i Meinard Müllers bok (2007)
- Synkronisering av symboldokumenter (noter) og audiodokumenter: Score following

MIR-systemer:

- Musikkøk:
- Melodisøk basert på indekseringer, dvs. på preprossering/omkoding til mer kompakte representasjoner (som i annen IR)
- Melodisøk basert på sekvenslikhet, dvs. man sammenligner fragmenter inntil man finner treff
- Melodisøk basert på geometriske metoder, dvs. man søker etter mer abstrakte (og dermed også fleksible) geometriske bilder av fragmenter, jfr. tidligere tanker om "shapes" i lydanalyse
- Mer "blinde" statistiske sammenligninger
- Musikkfiltering: komme med forslag til musikk

MIR-systemer:

- Musikkbrowsing, klassifikasjon, og visualisering:
- Audioklassifisering, f.eks. automatisk skille mellom tale, musikk og dagliglivslyder
- Sjangerklassifisering er i utgangspunktet også problematisk fordi folk (leg og lærd) kan være uenige i hva som er hva, men her en relativt grovmasket fordeling:

| Genre | Subgenre |
|-----------|----------------|
| Classical | Choir |
| | Orchestra |
| | Piano |
| | String quartet |
| Country | |
| Disco | |
| HipHop | |
| Jazz | BigBand |
| | Cool |
| | Fusion |
| | Piano |
| | Quartet |
| Swing | |
| Rock | |
| Blues | |
| Reggae | |
| Pop | |
| Metal | |

MIR-systemer:

- Klassifikasjon:
- Flere viktige sjangre er til nå ikke utforsket m.h.t. slik klassifikasjon, f.eks. ambient, elektroakustisk, etnisk, osv.
- Utfordringer mer visualisering, f.eks. spatial organsiering av forskjellige musikktyper, men også mer om innhold i ved hjelp av ulike lydavbildninger (spetrogram, pianola, osv.), egenskapsforhold (blant annet gjennom multidimensjonal skalering)
- Og hva med 'thumb nails'?
- Utprøving av MIR teknologier gjennom de årlige MIREX-konkurransene, se oversiktene s. 71 og 74
- Til slutt: Behov for også en *Multimedial Information Retrieval*

MIR, en foreløpig oppsummering:

- Et stort (kanskje det største) musikkforskningsfelt(et) i vår tid
- Symbolbasert MIR: Flere muligheter her, blant annet MIDIToolbox for Matlab
- Audiobasert MIR: Enda flere muligheter, først og fremst i PD-objeker/patcher og MIRToolboxen for Matlab som vi skal arbeide med, men også en rekke andre toolboxer og annen programvare rundt omkring
- I alt dette: Ha klart for seg hva man søker etter, dvs. hvilke trekk i musikken man vil undersøke, og derfor:
- Tenke trekk i musikken som *Shapes!*

MIR-kilder:

- <http://www.ismir.net/>
- <http://marsyas.info/>
- Og ganske enkelt, søk på 'music information retrieveval' + ulike andre søkeord
- Men særlig relevant for oss:
- <https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox> og div. publikasjoner nevnt der (særlig Lartillot & Toiviainen 2007 anbefales)

MIRToolbox, en oversikt:

- Utviklet av våre kolleger i Jyväskylä, men inneholder også komponenter utviklet av andre
- Består av en rekke scripts for Matlab
- En grei manual som forklarer hva de forskjellige komponentene gjør og hvordan man bruker dem
- Strekker seg fra helt grunnleggende og signalnære trekk til mer høynivå trekk, dvs. til semantiske og emotive elementer i musikken
- Muliggjør undersøkelser av alt fra enkeltlyder til store samlinger av musikkverk