Music, *Music Representation,*

AND

Music Retrieval

Donald Byrd

Center for Research on Concepts & Cognition

Indiana University Bloomington

20 April 2023

(rev. 10 May 2023)

Motivation: Audio-to-Audio Music Retrieval

- Shazam (Wang, 2003): “just hit 2580 on your mobile phone and identify music”
  - Query: 
  - Match: 
- Have they solved all the problems of music IR? No, (almost) none!
- Reason: desired signal & match are identical => no time warping, let alone higher-level problems (perception/cognition)

Overview

- I'll mostly avoid details of current technology & focus on underlying principles
- Organization of the talk
  - I. Music Is An Art. What Else Is It?
  - II. Music Representation: Audio, Events, Notation
  - III. Music Retrieval
  - IV. Conclusions
- Interrupt with burning questions; please save non-burning ones

You Are Here

I. Music Is An Art. What Else Is It?

II. Music Representation: Audio, Events, Notation

III. Music Retrieval

IV. Conclusions

Music Is An Art

- Music is an *art*
- => numerous works use the same elements, but very often in novel ways
- Example: string quartet in “Eleanor Rigby”
- Example: Ives’ *Concord Sonata* quotes Beethoven’s 5th 
- Example in visual art: Chagall story
- Text analogy: Music is more like poetry than prose

Music Is A Performing Art

- Music is a *performing art*
- => a performance is rarely the work itself
- It's one of numerous possible instantiations of the work
  - Audio: "Yesterday" has been covered over 3,000 times
  - Notation: at least eight editions of Mozart Piano Sonatas published by 1956
- Text analogy: play in verse
- Exception: Jimi Hendrix's "Star-Spangled Banner" improvisation at Woodstock. Two excerpts:  
- Exception: "electronic" music

Music May or May Not Be Representational 1

- Music can be totally *nonrepresentational*
- => everything can be abstract: need not focus on anything from the real world
- Instrumental music is rarely representational
- ...but it can refer to other music
  - Meaning unclear
    - Ex.: Ives' *Concord Sonata* quoting Beethoven's 5th
  - Meaning pretty clear!
    - Ex.: Berlioz: *Symphonie Fantastique*, V ("Witches' Sabbath") quotes the Gregorian chant "Dies Irae" ("Day of Wrath", i.e., Judgment Day)

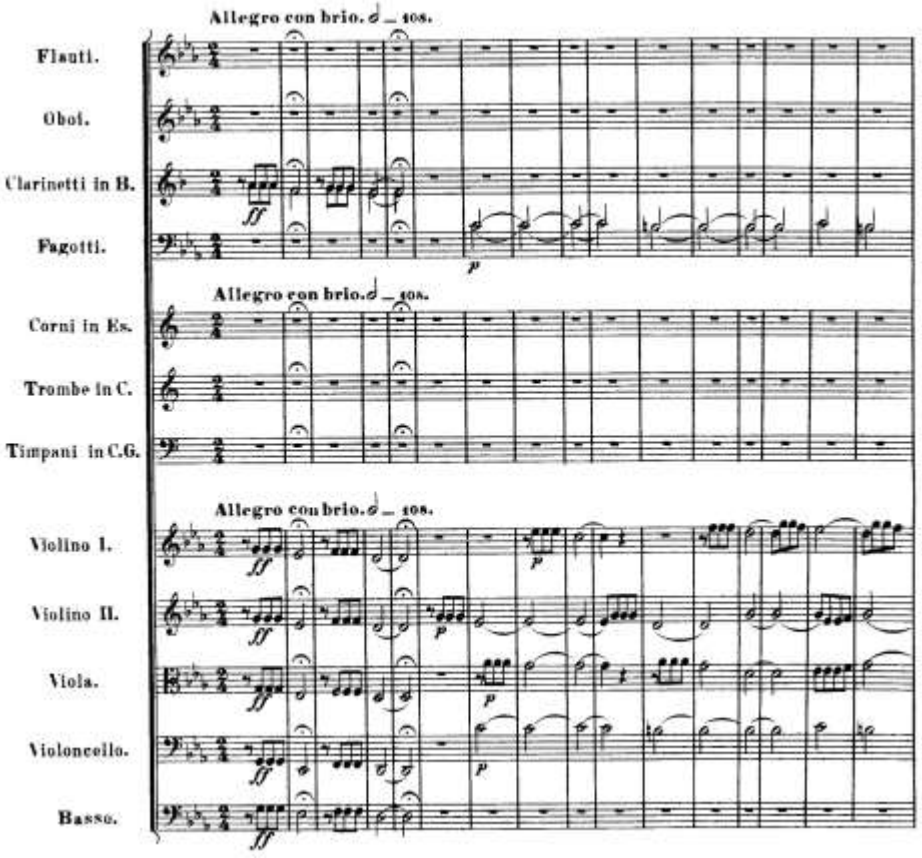
Music May or May Not Be Representational 2

- Vocal music is different
- ...but not completely, since lyrics are often poetry!
- Movie music is different for obvious reason
- Cf. denotation vs. connotation in semantics (later in this talk)

Music Is An Art, and...

- Most music, especially Western art-tradition music, has independent "voices" (melodic lines)
  - Unlike text, it's fundamentally 2-dimensional!
- Music involves many different instruments
  - including human voices, of course
- => "numerous possible instantiations" can be for other ensembles or at other levels of difficulty
- ...especially for music in notation form
  - Example: Beethoven's Fifth Symphony published in arrangements for at least nine different ensembles
  - "Easy" versions of pieces are common

Independent Voices in Music



Beethoven: Symphony no.5, I, beginning

Independent Voices in Text

MARLENE. What I fancy is a rare steak. Gret?

ISABELLA. I am of course a member of the / Church of England.*

GRET. Potatoes.

MARLENE. *I haven't been to church for years. / I like Christmas carols.

ISABELLA. Good works matter more than church attendance.

--Caryl Churchill: "Top Girls" (1982), Act 1, Scene 1

Performance (time goes from left to right):

M: What I fancy is a rare steak. Gret? I haven't been...

I: I am of course a member of the Church of England.

G: Potatoes.

You Are Here

I. Music Is An Art. What Else Is It?

II. Music Representation: Audio, Events, Notation

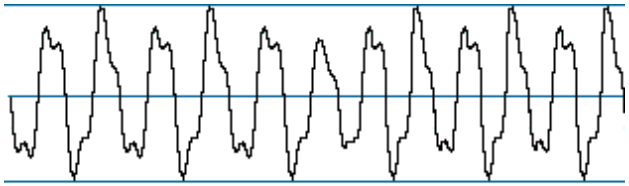
III. Music Retrieval

IV. Conclusions

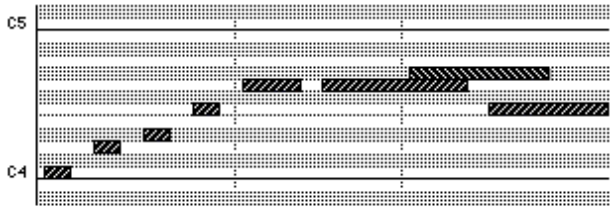
Basic Representations of Music

- Basic representations, with text analogs:

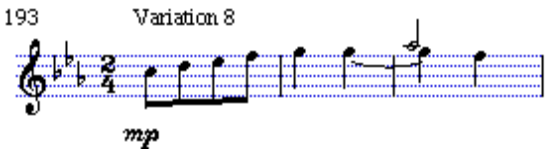
Audio (e.g., CD, MP3): like recording of speech



Time-stamped Events (e.g., MIDI file): like unformatted text



Music Notation (e.g., "Conventional"): like HTML or RTF text



- Essential difference: *explicit structure*

Basic Representations of Music

- MIDI = Musical Instrument Digital Interface: simple, standard low-bandwidth protocol from 1980's

	Audio	*Time-stamped Events*	*Music Notation*
Common examples	CD, MP3 file	Standard MIDI File	CMN sheet music
Unit	Sample	Event	Note, clef, lyric, etc.
Explicit structure	none	some (partial voicing information)	much (complete voicing information)

- Convert to form with less explicit structure (to left): moderately hard to do well
- Convert to form with more explicit structure (to right): *very* hard to do well

From Representation to Notation

- A *notation* is a visualization of a representation
- Choosing a representation inevitably introduces bias
- Choosing notation for it inevitably introduces more bias
- For huge body of important music, we have no choice: notation is CMN (Conventional Music Notation)!
  - Really "CWMN" (W = Western)
  - Alternative for some music: tablature (guitar, lute, etc.)
  - CMN is among the most successful notations ever
  - …but also among the most complex and subtle

Music Notation vs. Math and Chinese

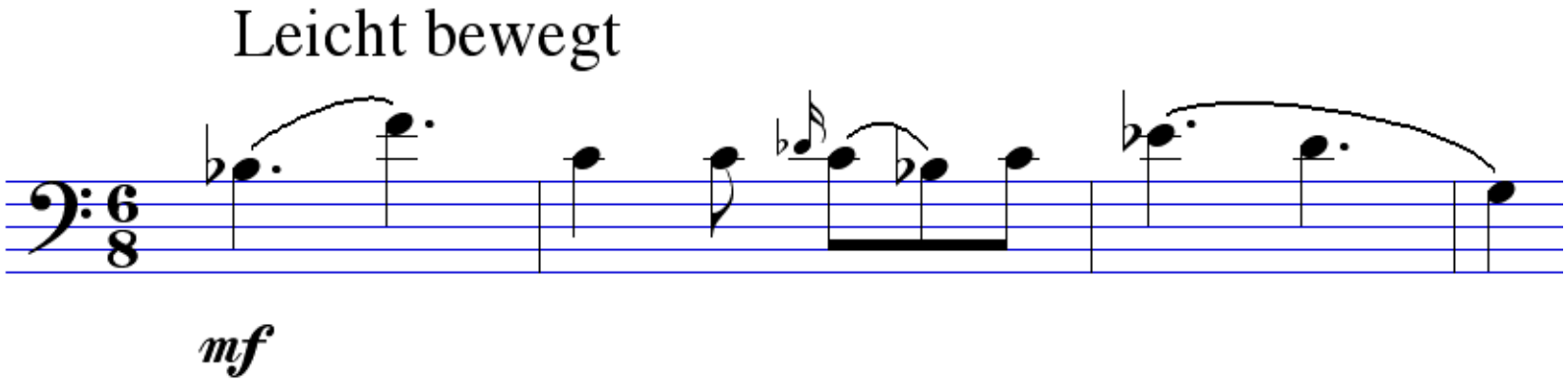
- Don's dissertation: CMN more complex/subtle than mathematical notation or Chinese
  - Typographic complexity, is relevant, not syntactic or semantic
- Chinese has a heck of a character set, but you just put the characters down in rows as with other languages
- Mathematical notation is a more serious opponent, but CMN is necessarily more complex because of constraints:
  - Music *must* be read in real time => as efficient as possible: "Omit Needless Symbols"
  - Ability to sight read music is important => no "macros"
  - Musician's hands are full => minimize page turns

How to Read Music Without Really Trying 1

- Basic parameters of a musical note
  1. *Pitch:* how high or low sound is
  2. *Duration:* how long the note lasts
  3. *Loudness:* perceptual analog of amplitude
  4. *Timbre* or tone quality
- Above in decreasing order of importance for most Western music

How to Read Music Without Really Trying 2

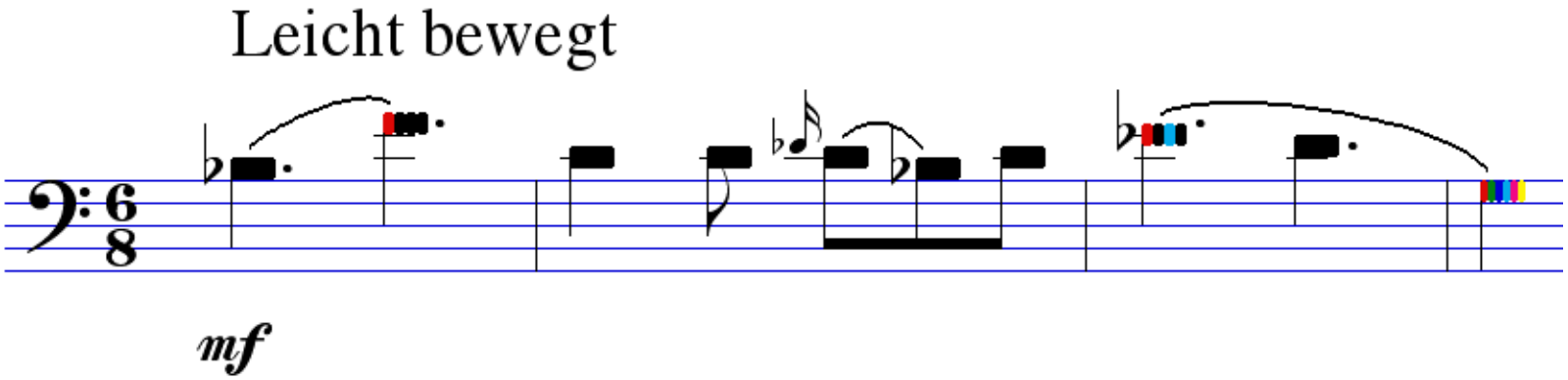
- Principles of CMN
  1. *Pitch* on Y axis: clef gives offset ("origin")
  2a. *Relative duration* indicated by note/rest shapes
  2b. Start times (sum of durations in the voice) on X axis
  3. *Loudness* indicated by signs like *p* , *mf* , etc.
  4. *Timbre* indicated with words like "violin", "horn", "pizzicato"



- But what if something changes *during* a note?

Adding to Notation Detail from Audio

- CMN bias: no way to show what happens during a note (with very limited exceptions)
- ...but it can be extended, e.g., with "notehead graphs"
- From an experimental version of Nightingale:



- Colors intended to show deviations from correct pitch
- ...but could show changes in loudness, tone quality, etc.

You Are Here

I. Music Is An Art. What Else Is It?

II. Music Representation: Audio, Events, Notation

III. Music Retrieval

IV. Conclusions

Audio-to-Audio Music Retrieval: No Explicit Structure Needed

- Shazam (Wang, 2003): "just hit 2580 on your mobile phone and identify music"
  - Query: 
  - Match: 
- Have they solved all the problems of music IR? No, (almost) none!
- Reason: desired signal & match are identical => no time warping, let alone higher-level problems (perception/cognition)

Similarity Scale for Content-Based Music IR

- Categories describe what's in common between query and items in corpus being searched (closest to most distant)

  NB: If in notation form, performance & recording don't apply.

1. Same music, arrangement, performance, & recording *(Shazam)*
2. Same music, arrangement, performance; different recording *(Dead)*
3. Same music, arrangement; different performance, recording *(OMRAS)*
4. Same music, different arrangement; or different but closely-related music, e.g., simpler variations (Mozart "Twinkle" Variations, etc.), many covers, minor revs. *(mainstream early music IR research)*
5. Different & less closely-related music: freer variations (Brahms, jazz, etc.) , extensive revisions, wilder covers (Hendrix in Part I) *(AI/Spotify)*
6. Music in same genre, etc. *(AI; Spotify?)*
7. Music influenced by other music *(AGI; GPT-5 or -6 or X?)*

OMRAS Polyphonic Audio Music IR: Requires Event Representation

- Ca. 2004; started with recordings of Bach preludes & fugues
- Did polyphonic (several notes at once) music recognition to convert to events
  - Converted results to MIDI & used as queries against database of ca. 3000 pieces in MIDI form
  - Example: one of worst-sounding cases: Prelude in G Major from Well-Tempered Clavier, Book I
  - Outcome: the actual piece was ranked 1st
- Models built from notation database, but note data only

- Query (audio -> MIDI) 
- Match (original audio recording) 

Music-IR Problems that Need More Structure

- Joan Public's problem: find a song, given some of the melody and some lyrics
  - Needs notes and text (lyrics)
  - Common question for music librarians (formerly?)
- Jean Public's problem: find music they're likely to like, given examples of music they do like
- Music scholar's problem: authorship/origin of works in manuscripts
  - Full symbolic data is important, even "insignificant" details of notation (John Howard)

Semantics in Music

- Meaning = *denotation* (explicit, well-defined)
- ...plus *connotation* (implicit, ill-defined)
- In text:
  - Two "definitions" of pig:
    - 1. Mammal with short legs, cloven hoofs, bristly hair, and a cartilaginous snout used for digging (dictionary)
    - 2. Ugh! Dirty, evil-smelling creatures, wallowing in filthy sties! (Hayakawa & Hayakawa example)
  - Prose is mostly denotation
  - Poetry is art => connotation much more important
- Music is always art, & almost all connotation!

Why is Musical Information Hard to Handle?

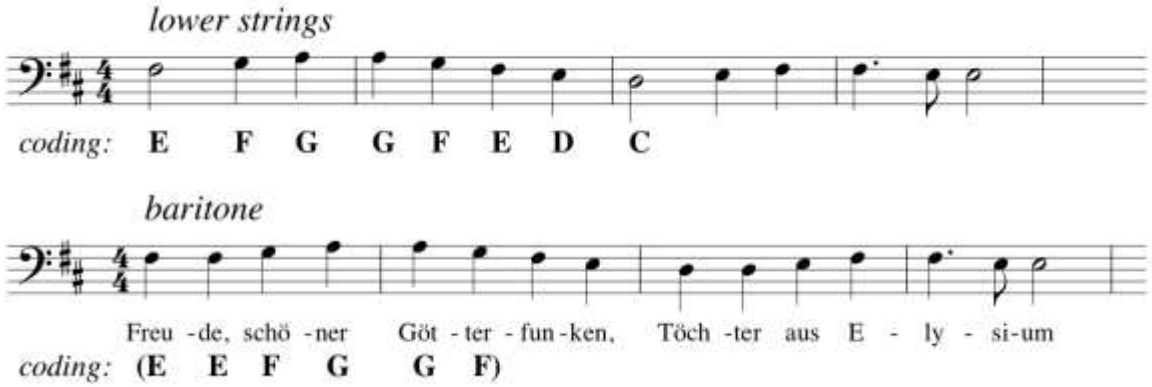
1. Polyphony: "parallel" independent voices, something like characters in a play *(all representations)*
   - "Music is fundamentally 2-dimensional"
2. Connotations are always important *(all representations)*
3. Recognizing notes *(audio only)*
4. Units of meaning: not clear anything is analogous to words *(all representations—NEW TO US)*
5. Melodic similarity (*all representations—NEW)*

Problem 4: Units of Meaning

- Not clear anything in music is analogous to words
  - No explicit delimiters (like Chinese)
  - Experts don't agree on "word" boundaries (unlike Chinese)
- Are *notes* like words?
- No. Relative, not absolute, pitch is important
- Are *pitch intervals* like words?
- No. They're too low level: more like characters
- Are *pitch-interval sequences* like words?
- In some ways, but
  - Ignores note durations
  - Ignores relationships between voices (harmony)

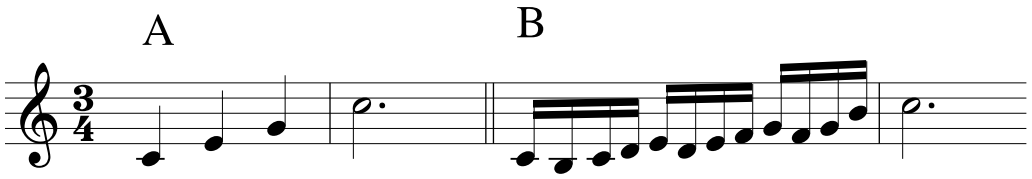
Problem 5: Melodic Similarity 1

- Evaluating similarity of versions of a melody isn't easy
- Minor (to our ears) differences may not be to a computer
- Easy example: repeated notes (Beethoven, Symphony no. 9, IV, "Ode to Joy" theme)
  - Version on top is how it first appears
  - Version on bottom is the familiar melody



Problem 5: Melodic Similarity 2

- Minor (to our ears) differences may not be to a computer
  - Approximate string matching, e.g., Levenshtein distance?
  - Fine for the repeated-note example, but isn't ideal
- Harder example: B is an "ornamented" version of A



  - In general, "Earth Mover's Distance" is much better
  - Why? "Music is fundamentally 2-dimensional"!

You Are Here

I. Music Is An Art. What Else Is It?

II. Music Representation: Audio, Events, Notation

III. Music Retrieval

IV. Conclusions & "By the Way"

Conclusions: Review

I. Music Is An Art. What Else Is It?

It's a performing art

It's fundamentally nonrepresentational

It involves many different instruments

It's polyphonic

II. Music Representation: Audio, Events, Notation

Essential difference: *explicit structure*

Audio: none; time-stamped events: some; notation (CMN): lots

Principles of CMN

III. Music Retrieval

- Representations express Semantics
- Semantics of Music; Denotation & Connotation
- Why Musical Information Is Hard to Handle: 4 reasons

Conclusions: Why Is This Important?

- Some problems directly related to other areas of informatics
  - Example: Approximate string matching in bioinformatics
- Handling music is *really* hard
  - GPT-4's "compositions" are mediocre; its attempts to improve them based on human feedback are usually unsuccessful (Eric Nichols)
- Encourages progress on real semantics
  - Connotation is an important part of meaning in everything
  - Can often be ignored, but not in arts
  - Music is probably more quantifiable than other arts, so likely more tractable
    - Evidence: compare music theory to theory of other arts

By the way, why does music in minor sound more serious than in major?

- Bob Krovetz's question (rephrased): Music in minor sounds more serious, even less "happy", than in major. Why?
- I don't know, but here's a plausible *speculation*:
  - The tonic triad in C minor consists of C, E-flat, and G
  - ...but the 5th harmonic of C is E-natural
  - Sounds of almost all acoustic instruments contain harmonics to the 5th and beyond
  - Interval from E-flat to E is a sharp dissonance
    - Technically, an augmented unison, but = minor 2nd
  - Q.E.D.? What do music psychologists say about it?

You Are Here

The End

References

- Avian Music Notation Foundation. Nightingale. https://github.com/AMNS/Nightingale
- Byrd, D. (1984). *Music Notation by Computer* (doctoral dissertation, Computer Science Dept., Indiana University). https://homes.luddy.indiana.edu/donbyrd/Papers/DonDissScanned.pdf
- Byrd, D., & Crawford, T. (2002). Problems of Music Information Retrieval in the Real World. *Information Processing and Management* 38, pp. 249–272.
- Hayakawa, S.I., & Hayakawa, A.R. (1991). *Language in Thought and Action,* 5th ed. Harvest.
- Nichols, E. (2023). Personal communication.
- Pickens, J., Bello, J.P., Crawford, T., Dovey, M., Monti, G., Sandler, M.B., & Byrd, D. (2002). Polyphonic Score Retrieval Using Polyphonic Audio Queries: A Harmonic Modeling Approach. In *Proceedings of the 3rd Int'l Conference on Music IR* (ISMIR 2002), Paris, France, pp. 140–149.
- Wang, A. (2003). An Industrial Strength Audio Search Algorithm. In *Proceedings of the 4th Int'l Conference on Music IR* (ISMIR 2003), Baltimore, MD