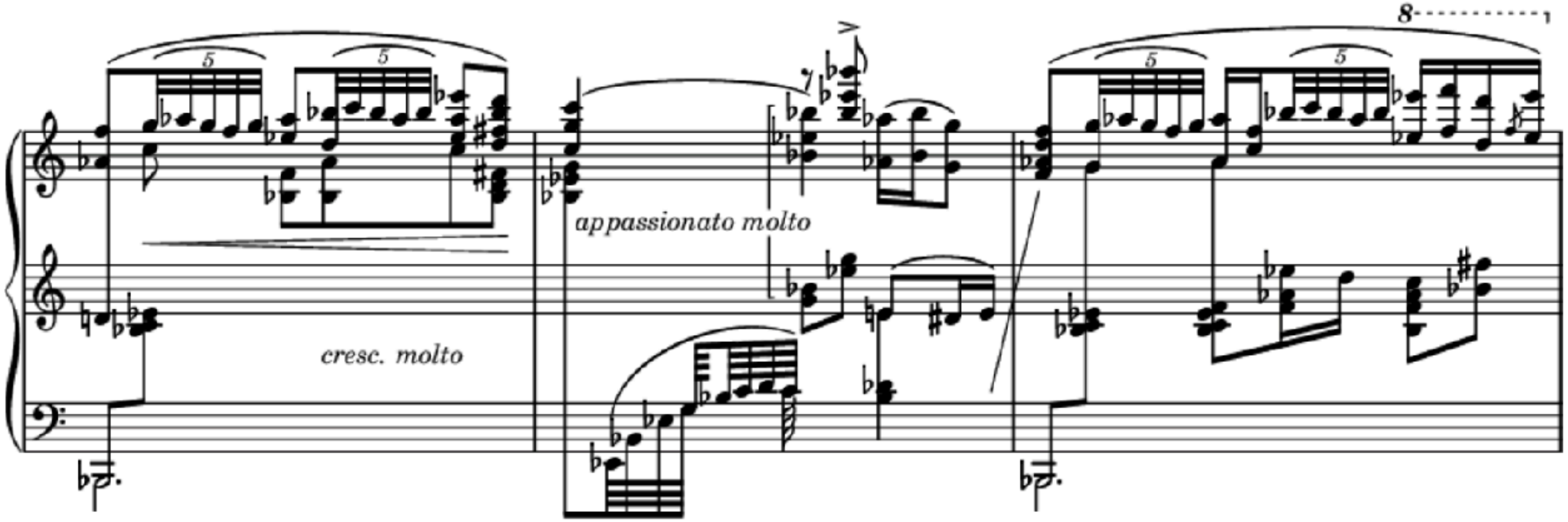
Exploratory Project Codex

Music Notation Processing



Philippe Rigaux
le cnam Paris

Florent Jacquemard
Inria

Raphaël Fournier-S'niehotta
le cnam

Lydia Rodriguez-de la Nava
PhD (Codex, Inria)

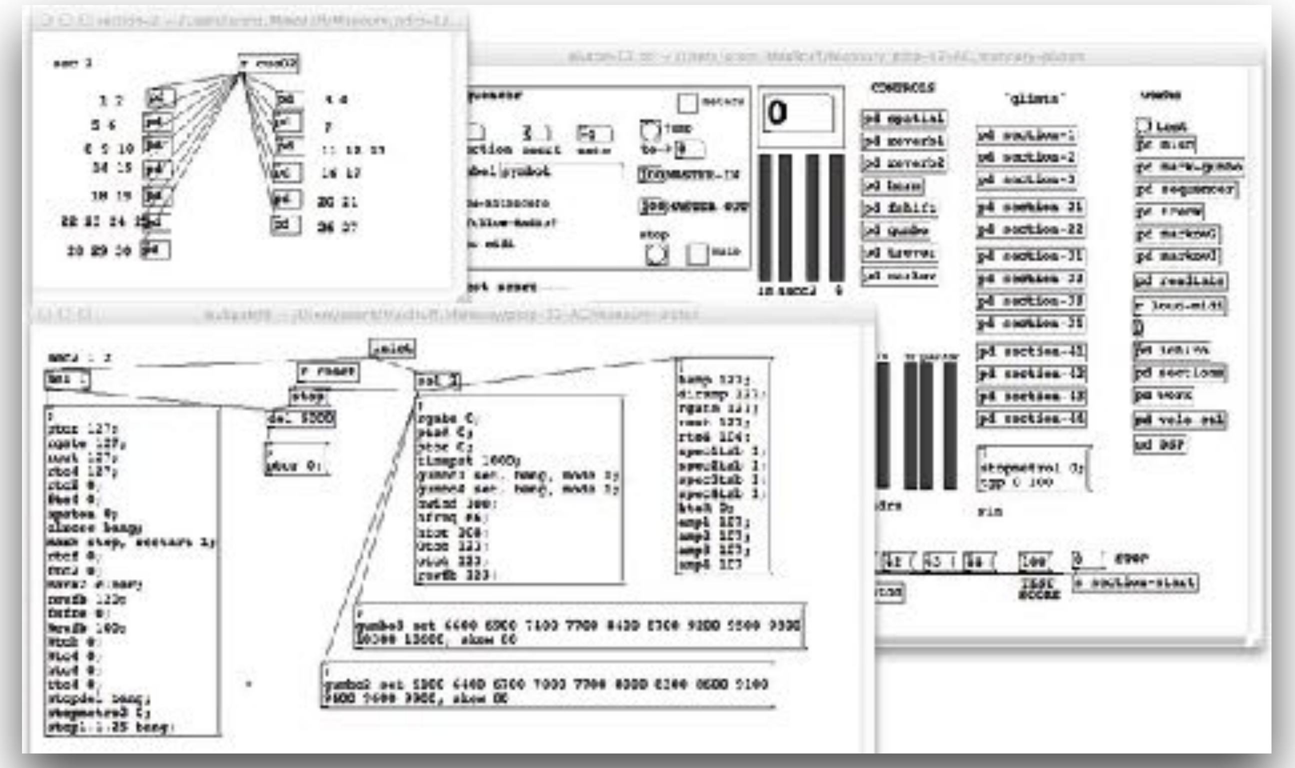
Tiange Zhu
PhD (Polifonia, H2020)

X
post-doc (Collabscore, ANR)

1/2 heure de Science - 06/10/2022

Why studying Music Notation Processing?

Western Music Notation = graphical format for music practice,
for exchange of musical information since ~1000 years (Guido d'Arezzo)



vs

Philippe Manoury
Tensio for string quartet and electronics

(digital) music scores, a language for

- players
  performance : real-time reading or memoization
- composers
  authoring, exchange
- teachers & students
  transmission
- editors
  access digital score libraries *e.g.* nkoda.com
- librarians
  cultural heritage preservation: *e.g.* Gallica
- scholars (historians, musicologists…)
  research, analysis

XML Music Notation Formats

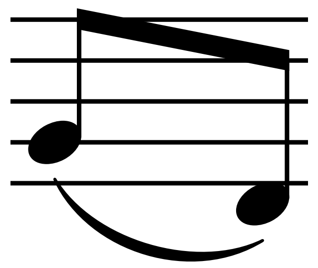
machine-readable encodings of musical documents

- MusicXML : industry standard (MakeMusic / Finale)
- MEI (Music Encoding Initiative) open source
  analog for music scholars of the TEI (text)
- MNX : next standard (W3C)

- multiple
- verbose
- redundant (mix logical and graphical information)
- ambiguous

```xml
<note default-x="80.72" default-y="-30.00">
  <pitch>
    <step>G</step>
    <octave>4</octave>
    </pitch>
  <duration>1</duration>
  <voice>1</voice>
  <type>eighth</type>
  <stem>up</stem>
  <beam number="1">begin</beam>
  <notations>
    <slur type="start" placement="below" number="1"/>
    </notations>
  </note>
<note default-x="132.57" default-y="-45.00">
  <pitch>
    <step>D</step>
    <octave>4</octave>
    </pitch>
  <duration>1</duration>
  <voice>1</voice>
  <type>eighth</type>
  <stem>up</stem>
  <beam number="1">end</beam>
  <notations>
    <slur type="stop" number="1"/>
    </notations>
  </note>
```

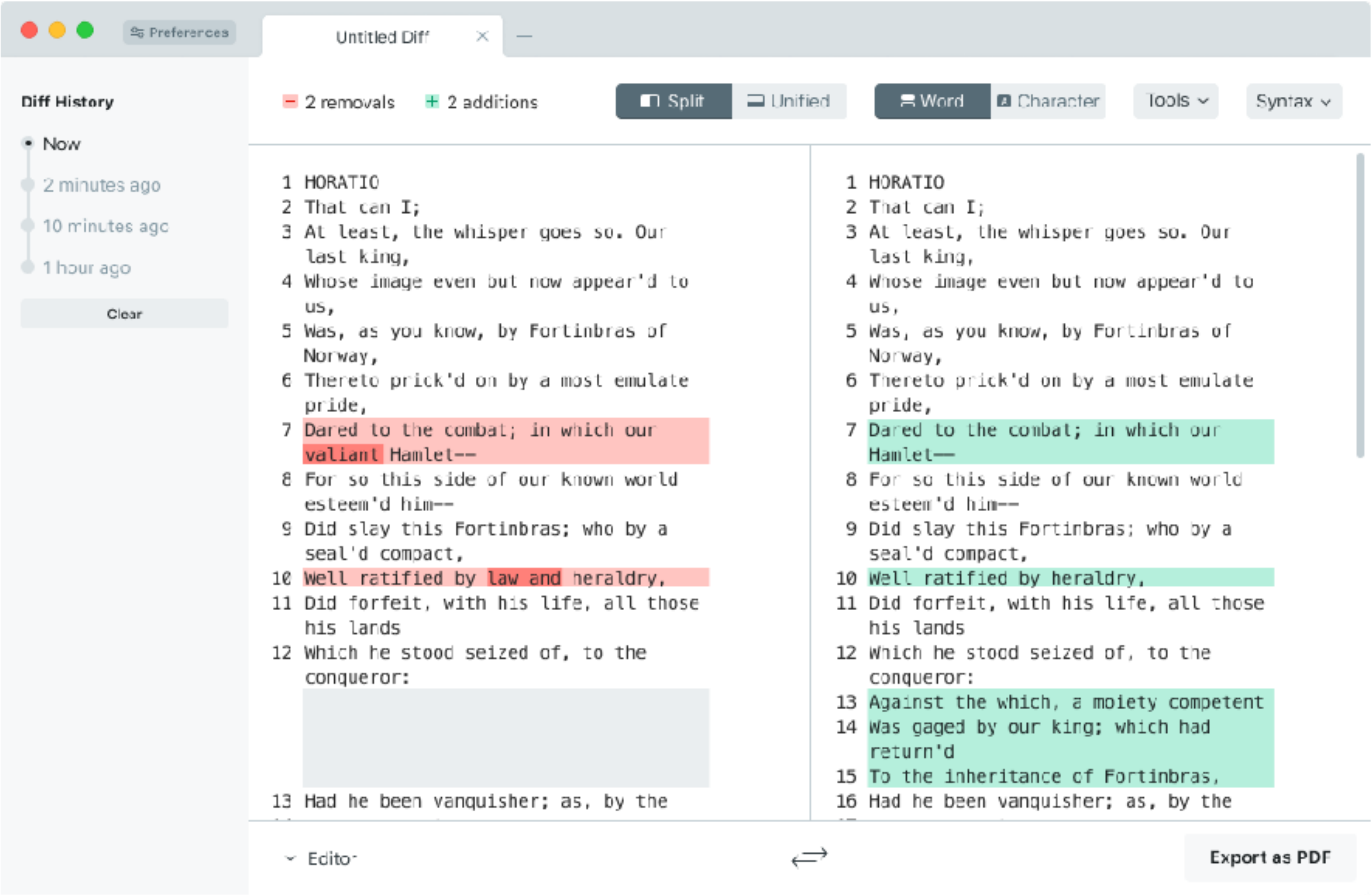
=



Text diff

tools to compare text files:

- Unix diff (1976)
- diffchecker.com
- git diff
- …

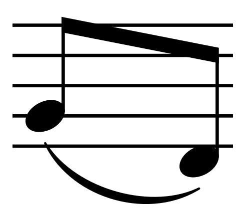


tools to compare music scores:

?

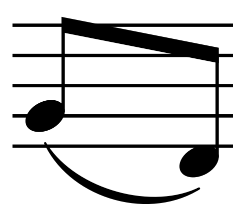
Music Score diff ?

```xml
<note default-x="80.72" default-y="-30.00">
  <pitch>
    <step>G</step>
    <octave>4</octave>
  </pitch>
  <duration>1</duration>
  <voice>1</voice>
  <type>eighth</type>
  <stem>up</stem>
  <beam number="1">begin</beam>
  <notations>
    <slur type="start" placement="below" number="1"/>
  </notations>
  </note>
<note default-x="132.57" default-y="-45.00">
  <pitch>
    <step>D</step>
    <octave>4</octave>
  </pitch>
  <duration>1</duration>
  <voice>1</voice>
  <type>eighth</type>
  <stem>up</stem>
  <beam number="1">end</beam>
  <notations>
    <slur type="stop" number="1"/>
  </notations>
</note>
```



?

=



```xml
<note default-x="80" default-y="-30.00">
  <pitch>
    <step>G</step>
    <octave>4</octave>
  </pitch>
  <duration>1</duration>
  <voice>1</voice>
  <type>eighth</type>
  <time-modification>
    <actual-notes>2</actual-notes>
    <normal-notes>2</normal-notes>
  </time-modification>
  <stem default-y="-30">up</stem>
  <beam number="1">begin</beam>
  <notations>
   <slur type="start" number="2" line-type="solid" />
  </notations>
</note>
<note default-x="132" default-y="-45.00">
  <pitch>
    <step>D</step>
    <octave>4</octave>
  </pitch>
  <duration>1</duration>
  <voice>1</voice>
  <type>eighth</type>
  <stem>up</stem>
  <beam number="1">end</beam>
  <notations>
    <slur number="2" placement="below" type="stop"/>
  </notations>
</note>
```

Intermediate Representation
abstract, unambiguous

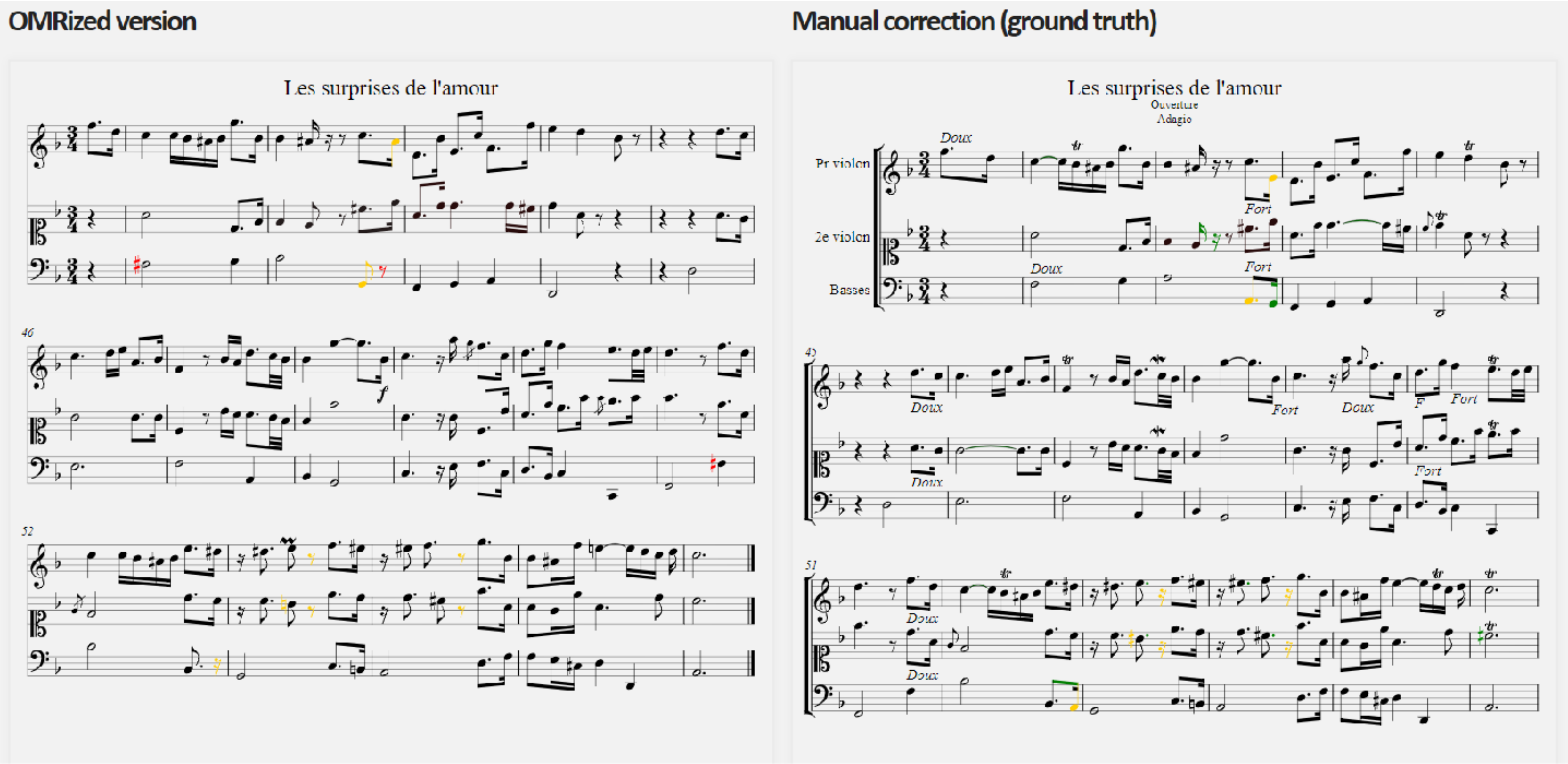
comparison

Intermediate Representation
abstract, unambiguous

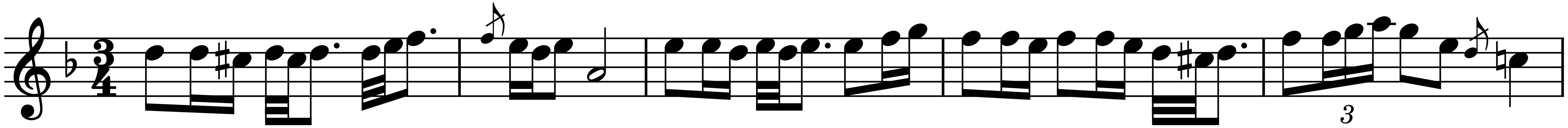
with Francesco Foscarin

- identify the differences between two XML music scores
- string/tree edit distance applied to a intermediate score representation

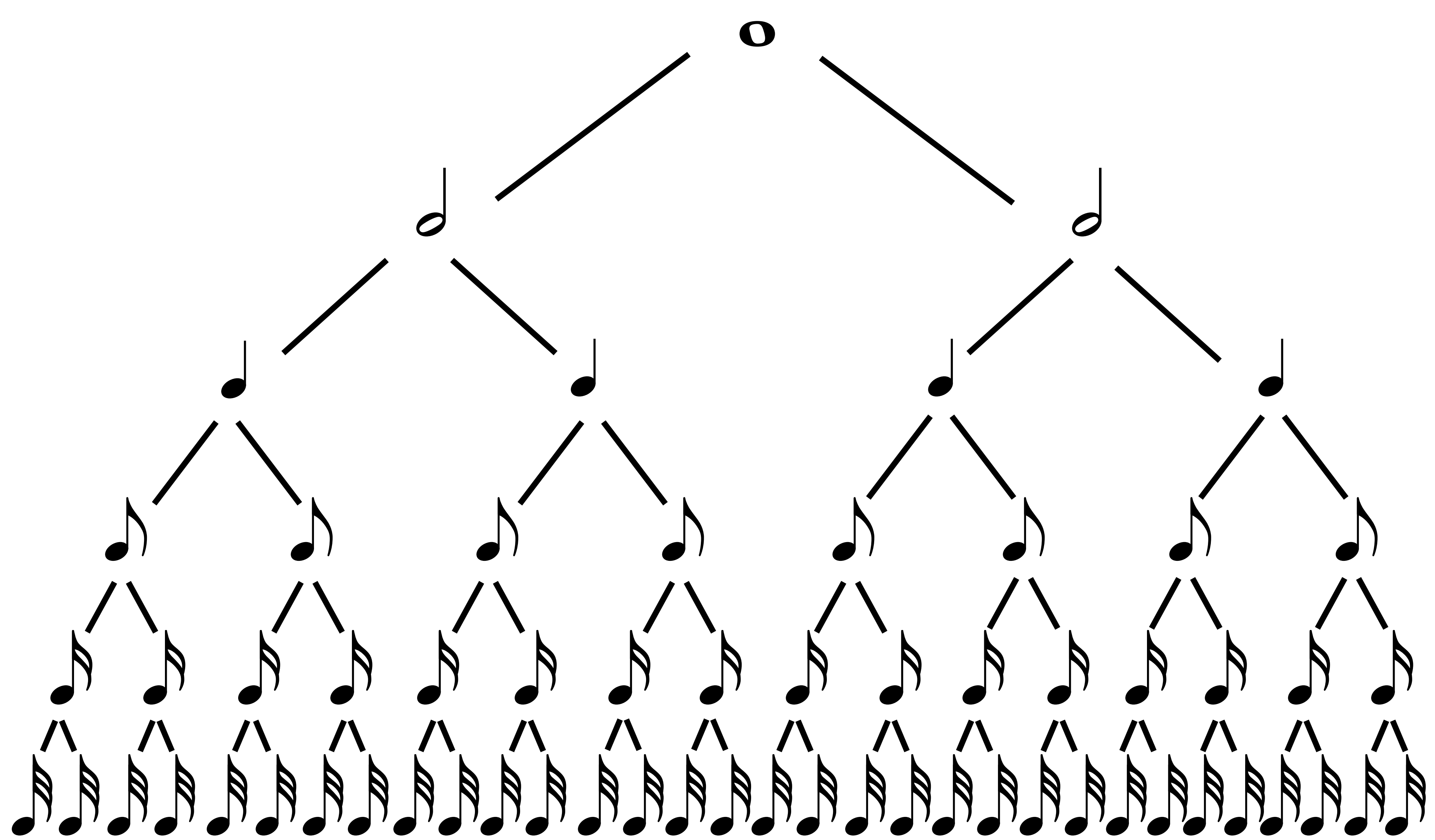
Case study on OMRization of a collection with IReMus lab



Polonaise in D minor from Notebook for Anna Magdalena Bach   BWV Anh II 128



hierarchical
note
durations



Common Western Music Notation

Polonaise in D minor from Notebook for Anna Magdalena Bach   BWV Anh II 128

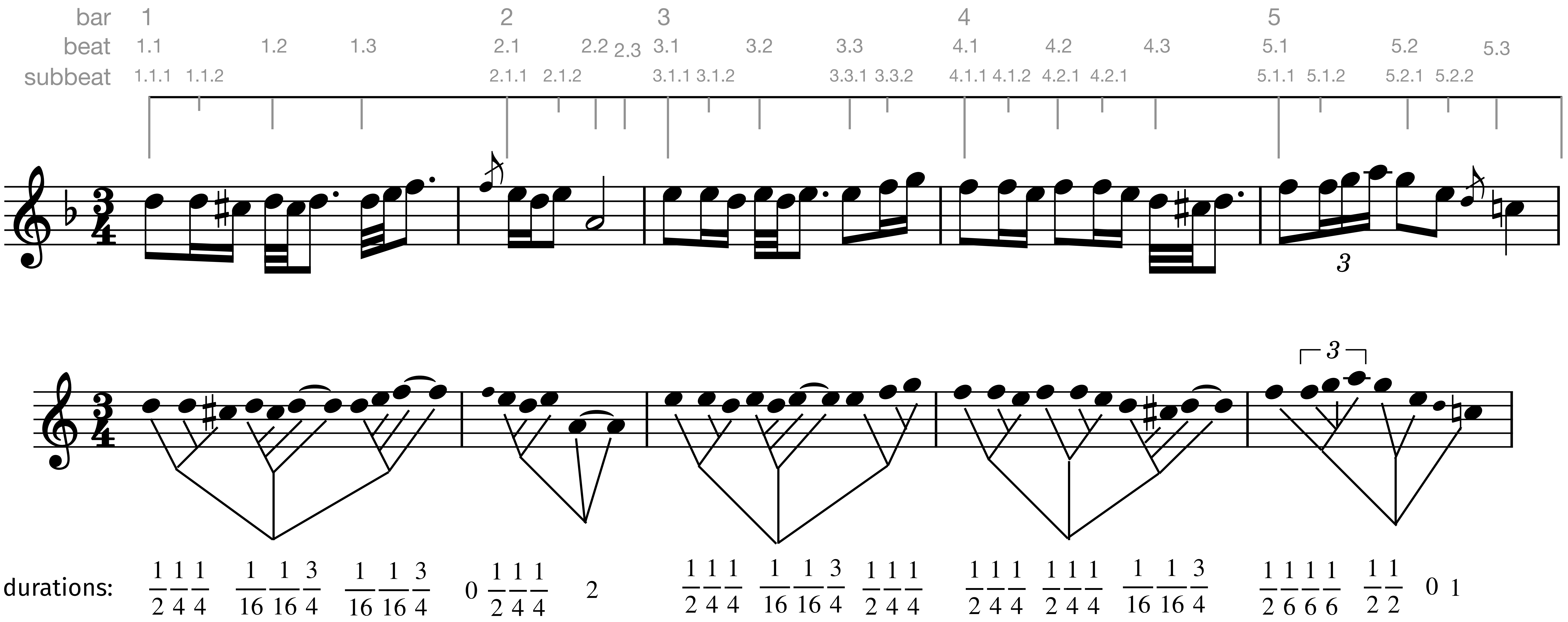
grouping notes with measure bars and beams

- eases readability (player reads in a real-time context)
- highlight the metric structure
  hierarchy of strong / weak beats

Common Western Music Notation

Polonaise in D minor from Notebook for Anna Magdalena Bach BWV Anh II 128

metric structure

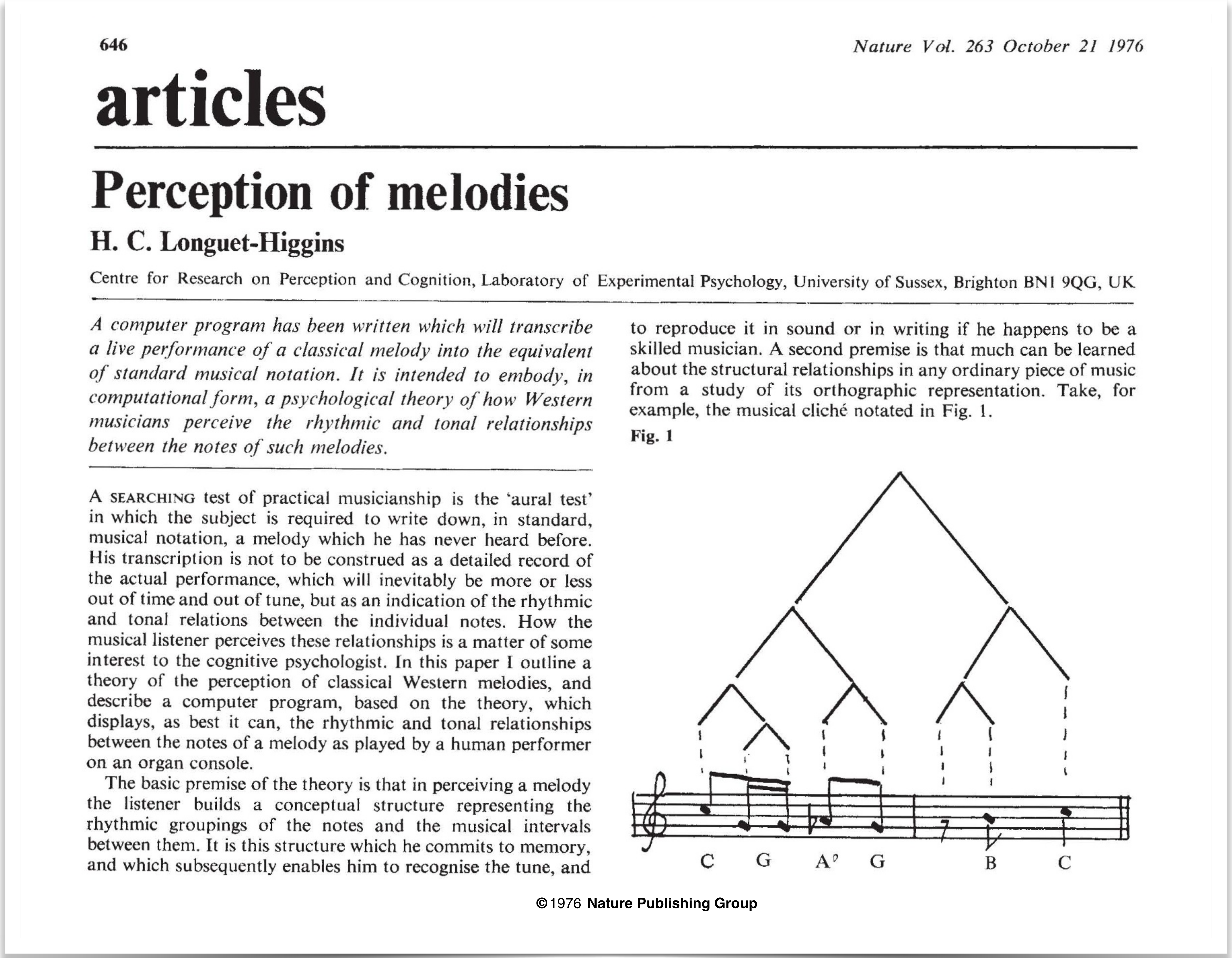


in our work, we consider a

tree-structured Intermediate Representation
(abstract & unambiguous)

Conversion of a recorded music performance into a music score ~ *speech-to-text* in NLP

one holy graal in Computer Music since 1970's…



646 *Nature Vol. 263 October 21 1976*

articles

Perception of melodies

H. C. Longuet-Higgins

Centre for Research on Perception and Cognition, Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, UK

A computer program has been written which will transcribe a live performance of a classical melody into the equivalent of standard musical notation. It is intended to embody, in computational form, a psychological theory of how Western musicians perceive the rhythmic and tonal relationships between the notes of such melodies.

A SEARCHING test of practical musicianship is the 'aural test' in which the subject is required to write down, in standard, musical notation, a melody which he has never heard before. His transcription is not to be construed as a detailed record of the actual performance, which will inevitably be more or less out of time and out of tune, but as an indication of the rhythmic and tonal relations between the individual notes. How the musical listener perceives these relationships is a matter of some interest to the cognitive psychologist. In this paper I outline a theory of the perception of classical Western melodies, and describe a computer program, based on the theory, which displays, as best it can, the rhythmic and tonal relationships between the notes of a melody as played by a human performer on an organ console.

The basic premise of the theory is that in perceiving a melody the listener builds a conceptual structure representing the rhythmic groupings of the notes and the musical intervals between them. It is this structure which he commits to memory, and which subsequently enables him to recognise the tune, and to reproduce it in sound or in writing if he happens to be a skilled musician. A second premise is that much can be learned about the structural relationships in any ordinary piece of music from a study of its orthographic representation. Take, for example, the musical cliché notated in Fig. 1.

Fig. 1

Automated Music Transcription today

source(s)
“recorded music performance”

Audio recording

MIDI device
(score edition)

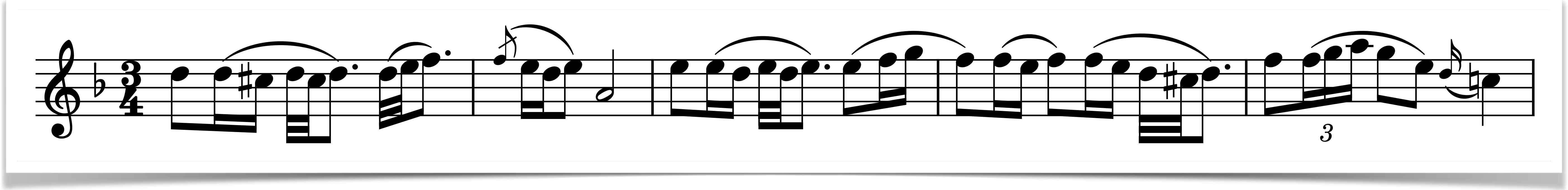
Algorithmic composition
DAW

audio Music Information Retrieval
- fundamental freq. estimation
- onset detection
- beat tracking …

symbolic representation of performance
piano roll (MIDI file)
- quantized pitches
- unquantized onsets & durations

symbolic Music Information Retrieval
- rhythm quantization
- tempo tracking
- score engraving…

target representation
music score (XML file)



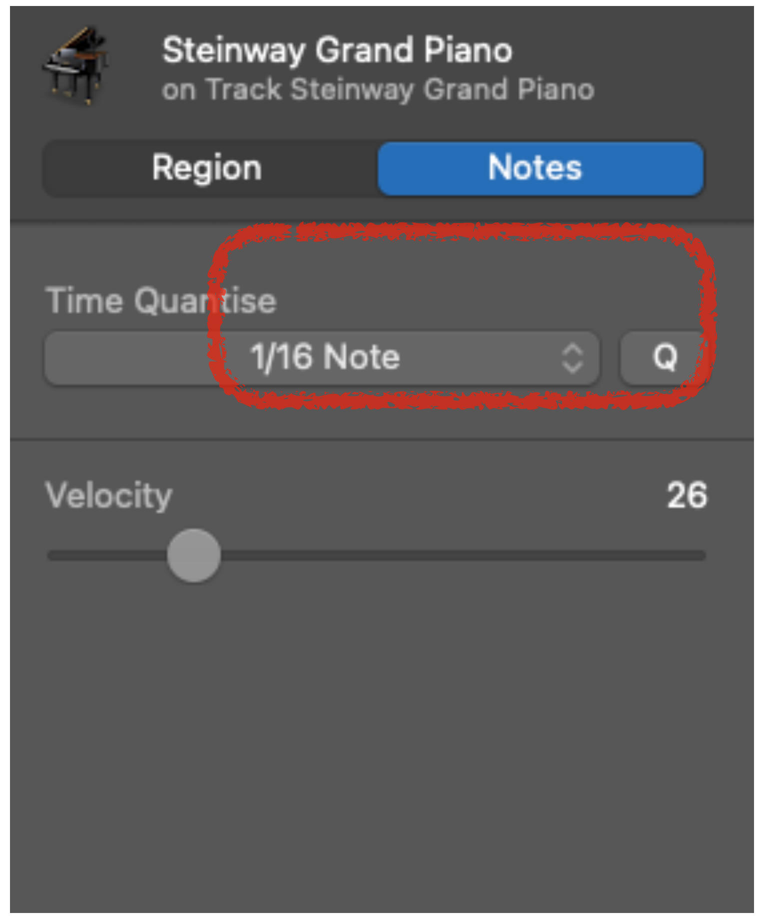
Rhythm Quantization “on Grid”

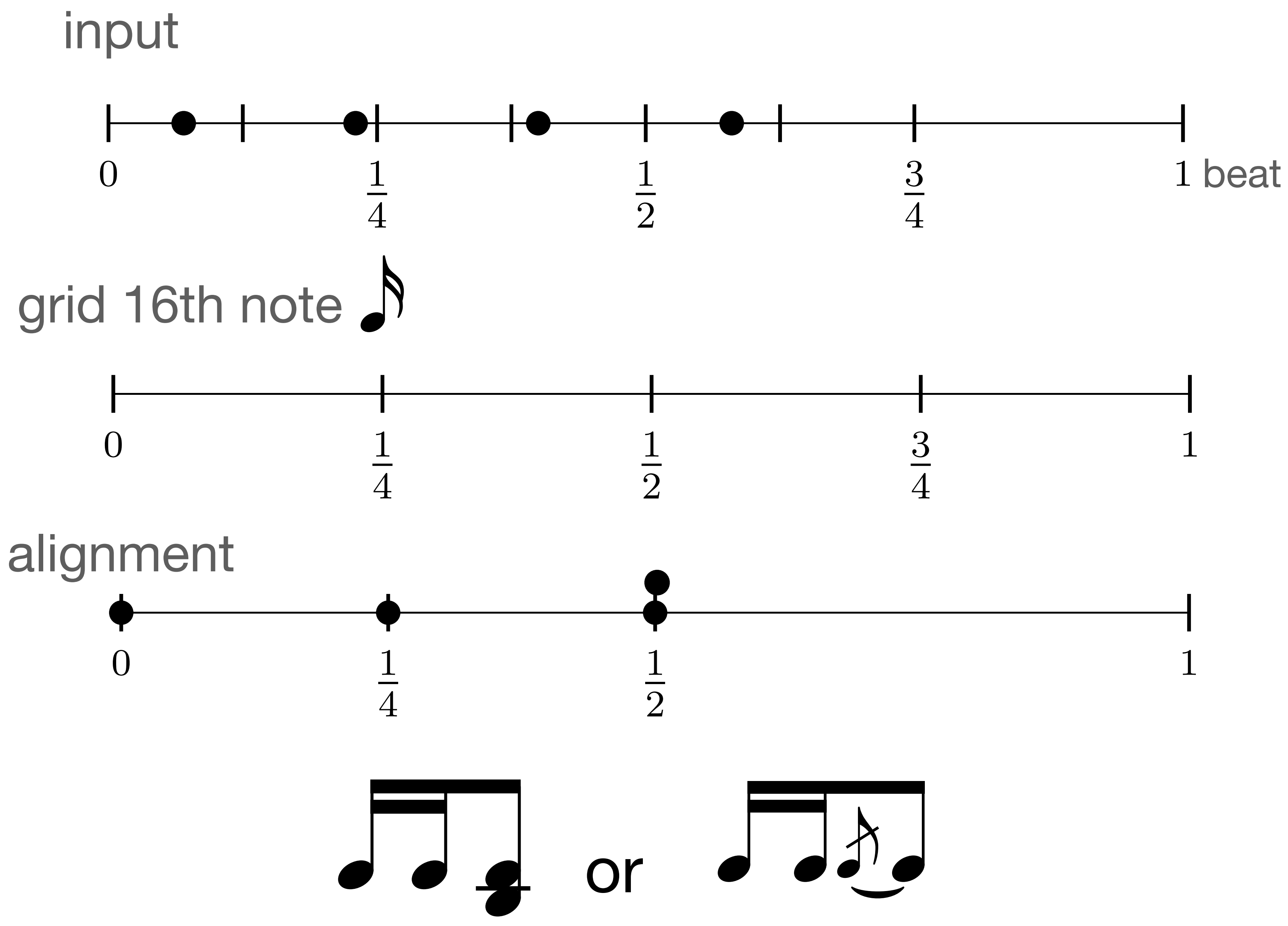
Rhythm quantization with grids, *e.g.* MIDI files import
• in score editors (Finale, Sibelius, Dorico, Musescore…),
• or in DAWs (Ableton Live, Logic…)

GarageBand, Apple

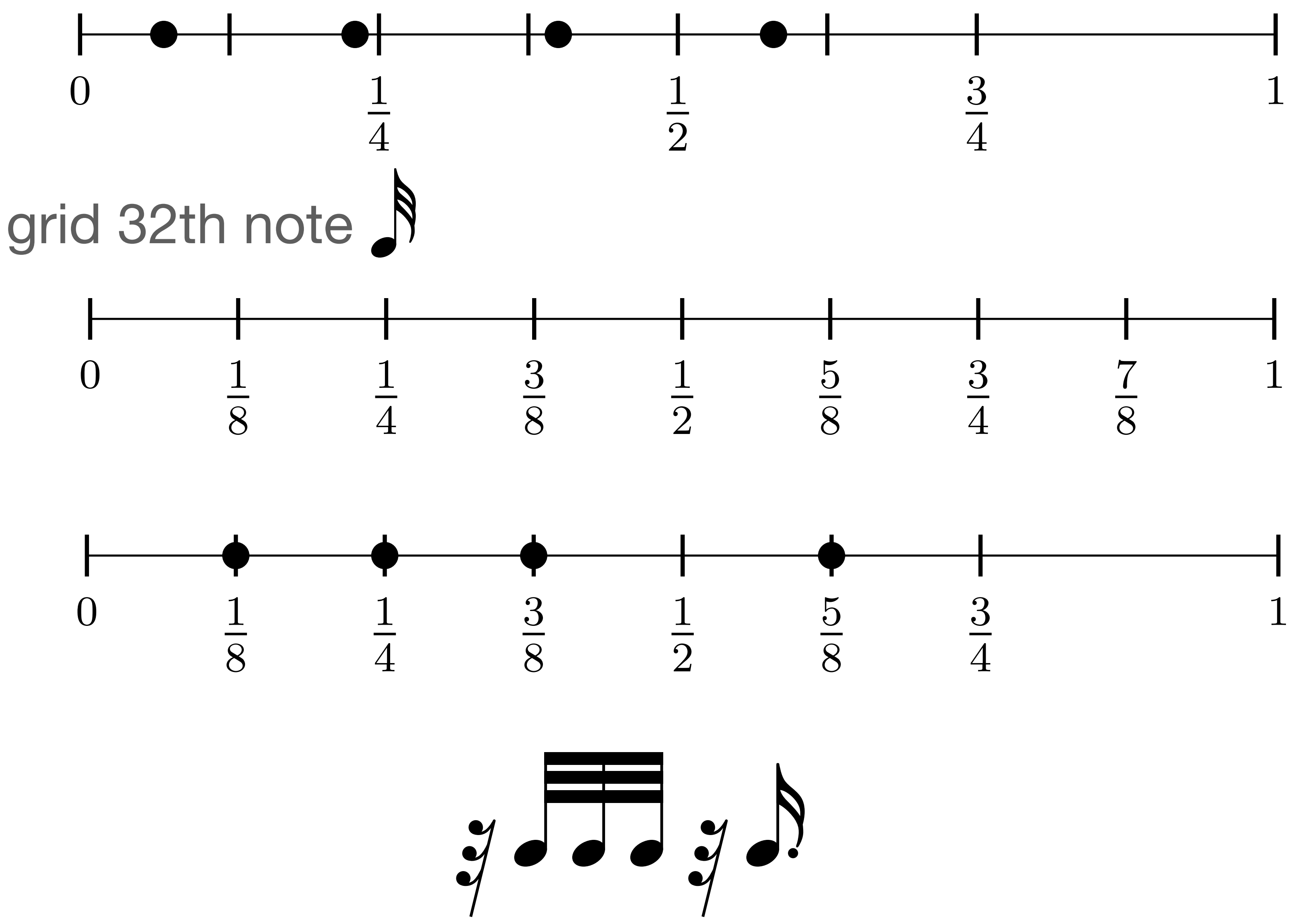
Grid based Approaches to Rhythm Quantization

Alignment of every input time point (onset) to the closest position in a *grid* = sequence of equidistant time position.

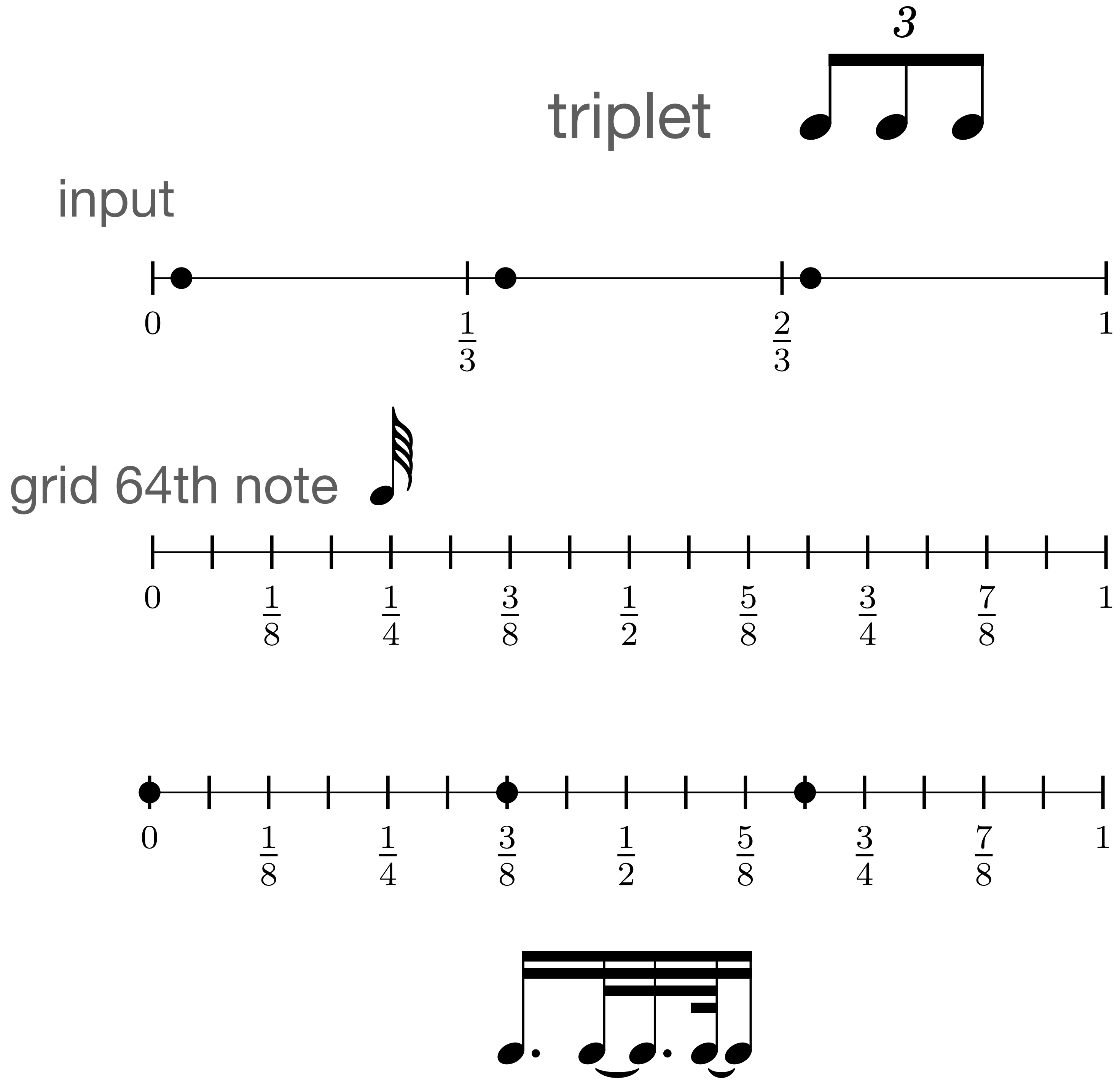




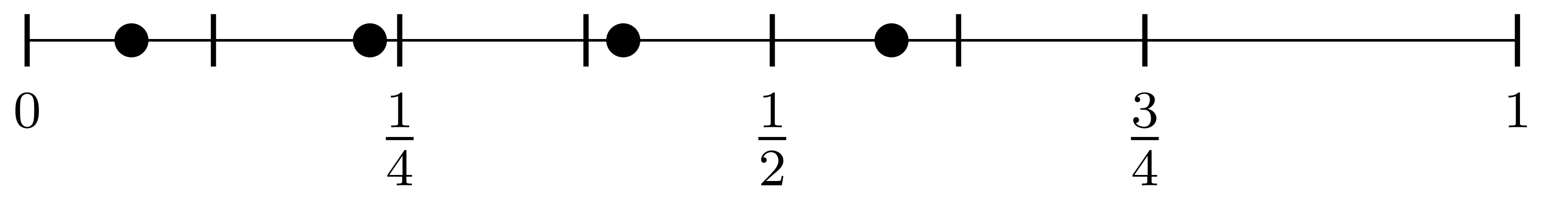
poor fit, good readability



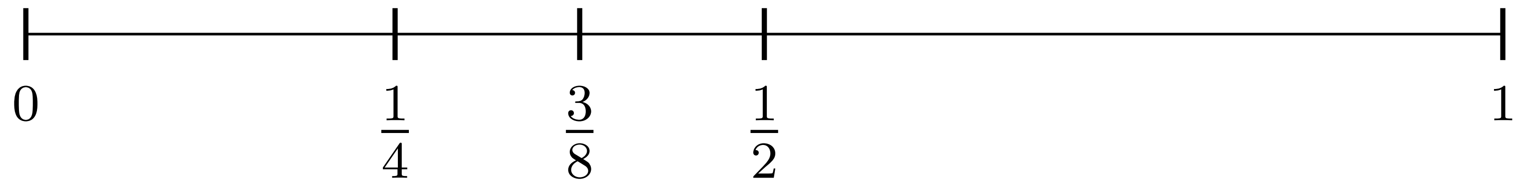
good fit, bad readability

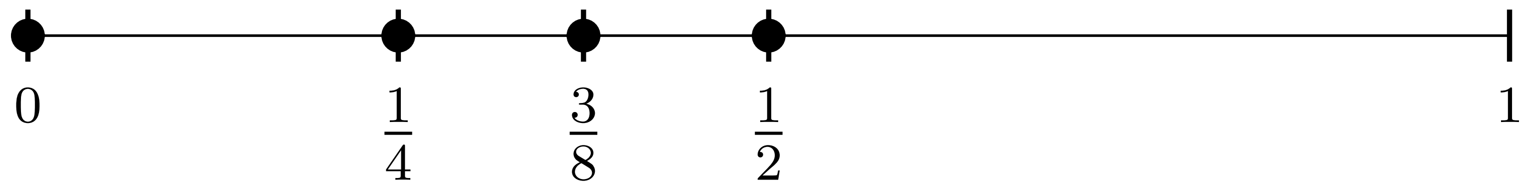


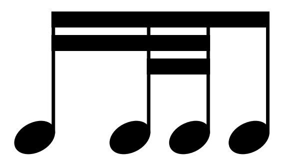
good fit, bad readability



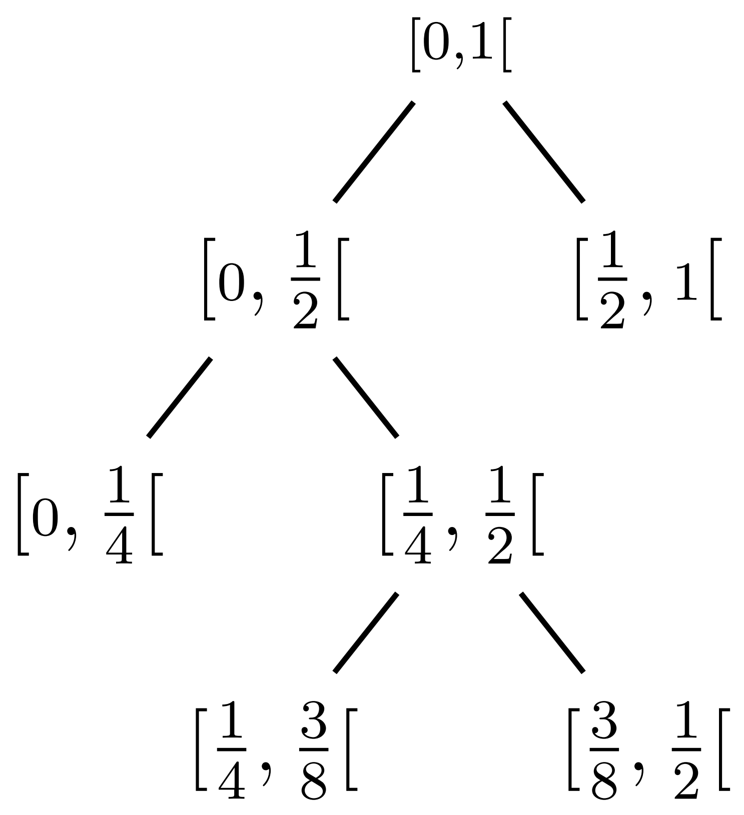
hierarchical grid







close to intuition



regular grids

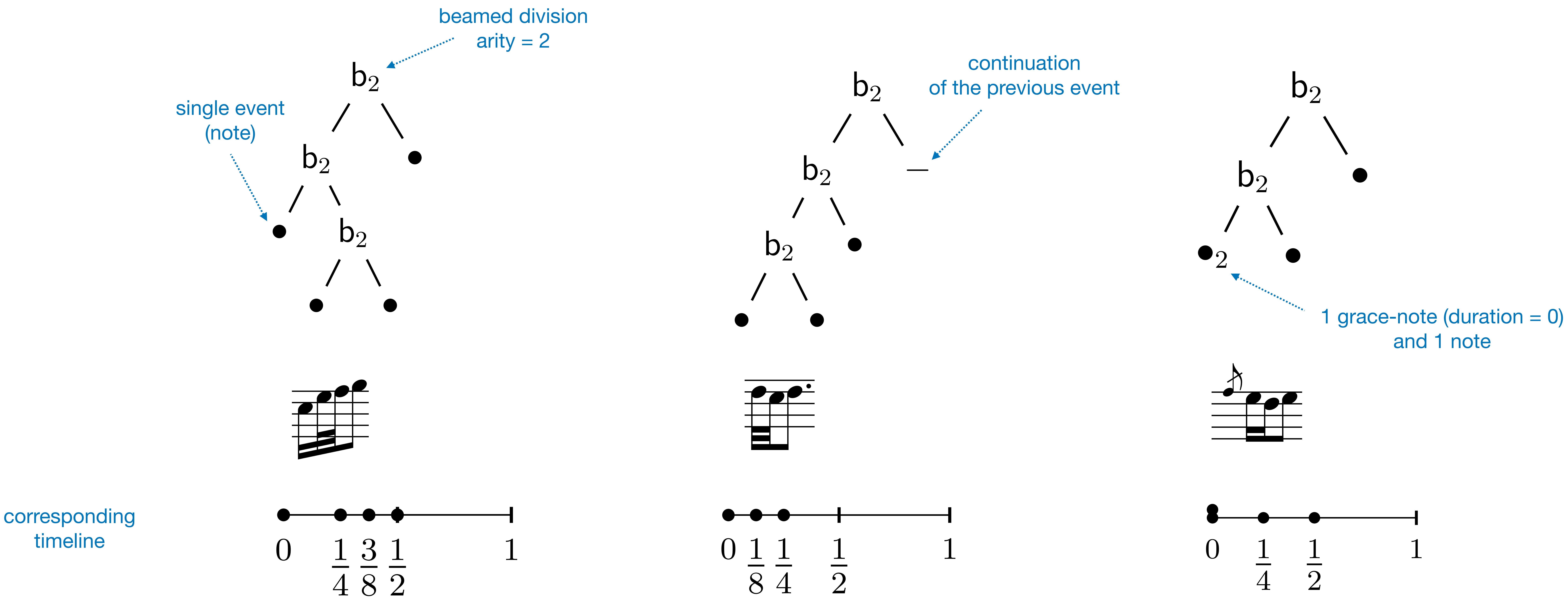
- search of a best quantization is possible by a brute-force enumeration: 8th note grid, 16th, 32th, 64th…
- result not always optimal
- problems with tuplets (so called “*irrationals*” 3, 5, 7…)

hierarchical grids

- more “natural” results
- brute force enumeration impossible
- how to specify the grids to try ?

Tree representation of the proportional rhythmic notation
with hierarchical encoding of durations: “*the* (duration) *data is in the structure*”

- the tree leaves contain the events
- the branching define durations, by partitioning of time intervals



Weighted Regular Tree Grammar $\mathcal{G}$:

- non-terminal symbols: $q, q_0, q_1, \ldots$
- terminal symbols (constants): • (1 note), $\bullet_2$ (1 grace-note + 1 note), — (continuation)
- every production rule is assigned a weight value (*e.g.* cost to read):

$$q \xrightarrow{0} \mathsf{m}_2(q_0, q) \qquad q \xrightarrow{0} \mathsf{m}_0$$

$$q_0 \xrightarrow{0.1} \mathsf{u}_3(q_1, q_1, q_1) \qquad q_0 \xrightarrow{1} \bullet$$

$$q_1 \xrightarrow{0.1} \mathsf{b}_2(q_2', q_2) \qquad q_1 \xrightarrow{1} \bullet \qquad q_1 \xrightarrow{1.9} \bullet_2 \qquad q_1 \xrightarrow{1} -$$

$$q_2' \xrightarrow{0.1} \mathsf{b}_2(q_3', q_3) \qquad q_2' \xrightarrow{1} \bullet \qquad q_2' \xrightarrow{2.25} \bullet_2 \qquad q_2' \xrightarrow{1} -$$

$$q_2 \xrightarrow{0.1} \mathsf{b}_2(q_3, q_3) \qquad q_2 \xrightarrow{1} \bullet \qquad q_2 \xrightarrow{1} -$$

$$q_3' \xrightarrow{1} \bullet \qquad q_3' \xrightarrow{3.25} \bullet_2 \qquad q_3' \xrightarrow{1} - \qquad q_3 \xrightarrow{1} \bullet \qquad q_3 \xrightarrow{1} -$$

measure

beat = 𝅗𝅥

sub-beat = 8th-note = ♪

sub-sub-beat = 16th note = 𝅘𝅥𝅯

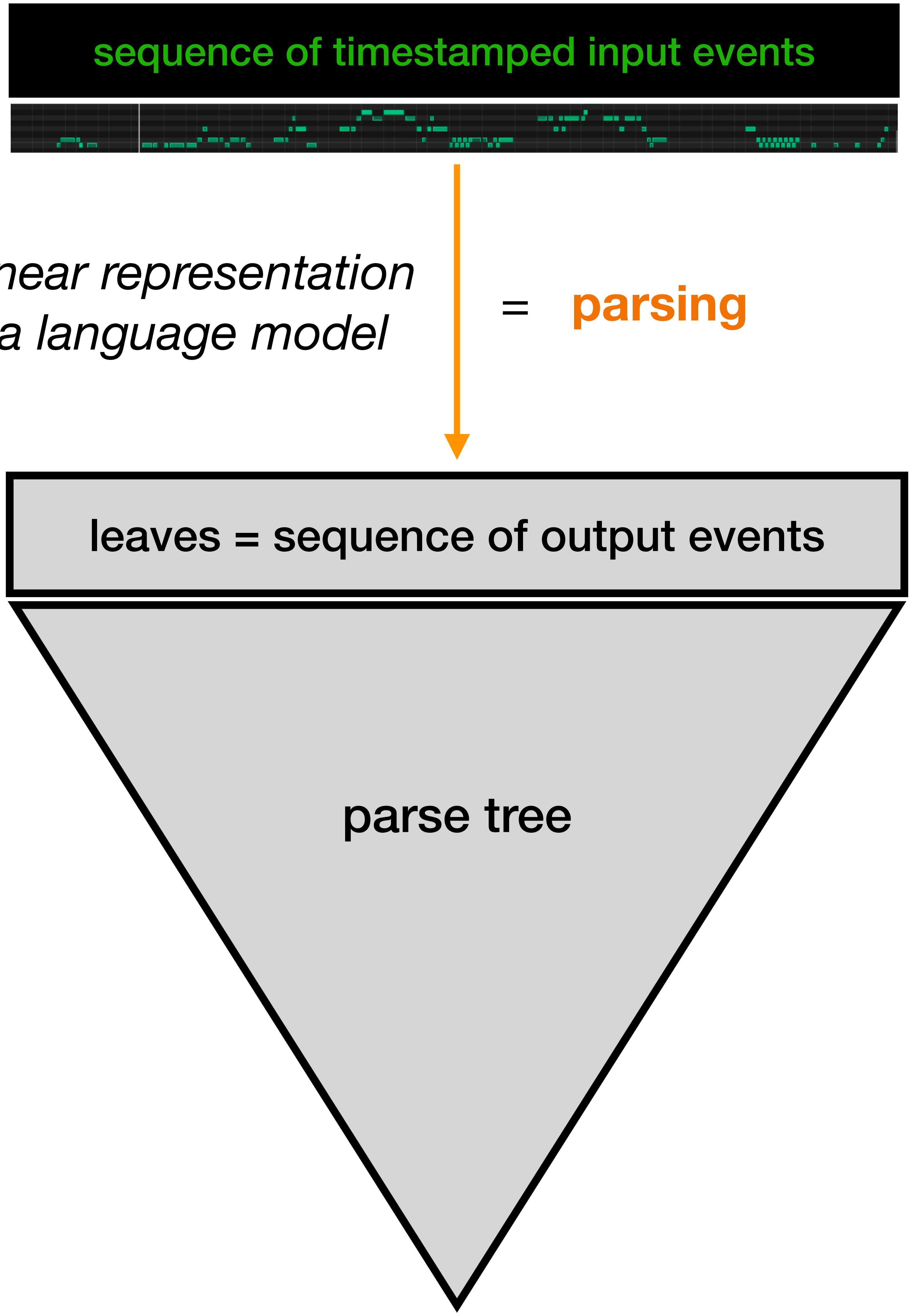
derivation (lefmost): $d : q_1 \xrightarrow{0.1} \mathsf{b}_2(q_2', q_2) \xrightarrow{0.1} \mathsf{b}_2(\mathsf{b}_2(q_3', q_3), q_2) \xrightarrow{3.25} \mathsf{b}_2(\mathsf{b}_2(\bullet_2, q_3), q_2) \xrightarrow{1} \mathsf{b}_2(\mathsf{b}_2(\bullet_2, \bullet), q_2) \xrightarrow{1} \mathsf{b}_2(\mathsf{b}_2(\bullet_2, \bullet), \bullet)$

cost of derivation: $\mathsf{weight}(d) = 0.1 + 0.1 + 3.25 + 1 + 1$

learning weight values from corpus statistics
Francesco Foscarin

Music Transcription as Parsing

piano roll



structuring a linear representation according to a language model

= **parsing**

tree-structured representation of an output music score

conform to a prior language (expected notation) defined by a Regular Tree Grammar

quantitative extensions of the problem are needed for music transcription:

(timestamped) input events are in an infinite alphabet

input and output events of different nature
the goal is not equality between sequences but a minimal distance.

with an ambiguous grammar, there can be several parse trees
→ select the most readable one.

Automated Music Transcription: qparse framework

https://gitlab.inria.fr/qparse/qparselib 75 Kloc C++

https://qparse.gitlabpages.inria.fr

- Command lines tools
- Python binding - Lydia Rodrigez-de la Nava
- Online port, real-time - Leyla Villaroel
- Other subtasks: pitch-spelling, key estimation, beat tracking…

monophonic : one note at a time
Good results for complex cases (ornaments, mixed tuplets, mixed note durations, silences…)
~ 100ms for the transcription of 1 score

Polonaise in D minor from Notebook for Anna Magdalena Bach   BWV Anh II 128

original score

transcription of MIDI recording by qparse

Monophonic transcription

Polonaise in D minor from Notebook for Anna Magdalena Bach   BWV Anh II 128

original score

transcription of MIDI recording by Finale (MakeMusic)

Dataset for Monophonic transcription

Lamarque-Goudard dataset (w. Francesco Foscarin, Teysir Baoueb)

- 283 monophonic extracts
  of classical repertoire
  inspired by a rhythm learning method
- ~ 20 measures per extract
- progressive difficulty
  cover a very large spectrum of rhythmic features
- score files (XML) and MIDI performances
  for evaluation and calibration of transcription tools

Generation of artificial performances

- fuzz testing ("*humanizing*" music score)
- more musical approach
  Madoka Goto, Masahiko Sakai (Nagoya U.), Satoshi Tojo (JAIST)
  - segmentation into phrases according to musicology analysis
  - performance following the phrase structure (*Director Musices*)

Monophonic transcription: datasets and case studies

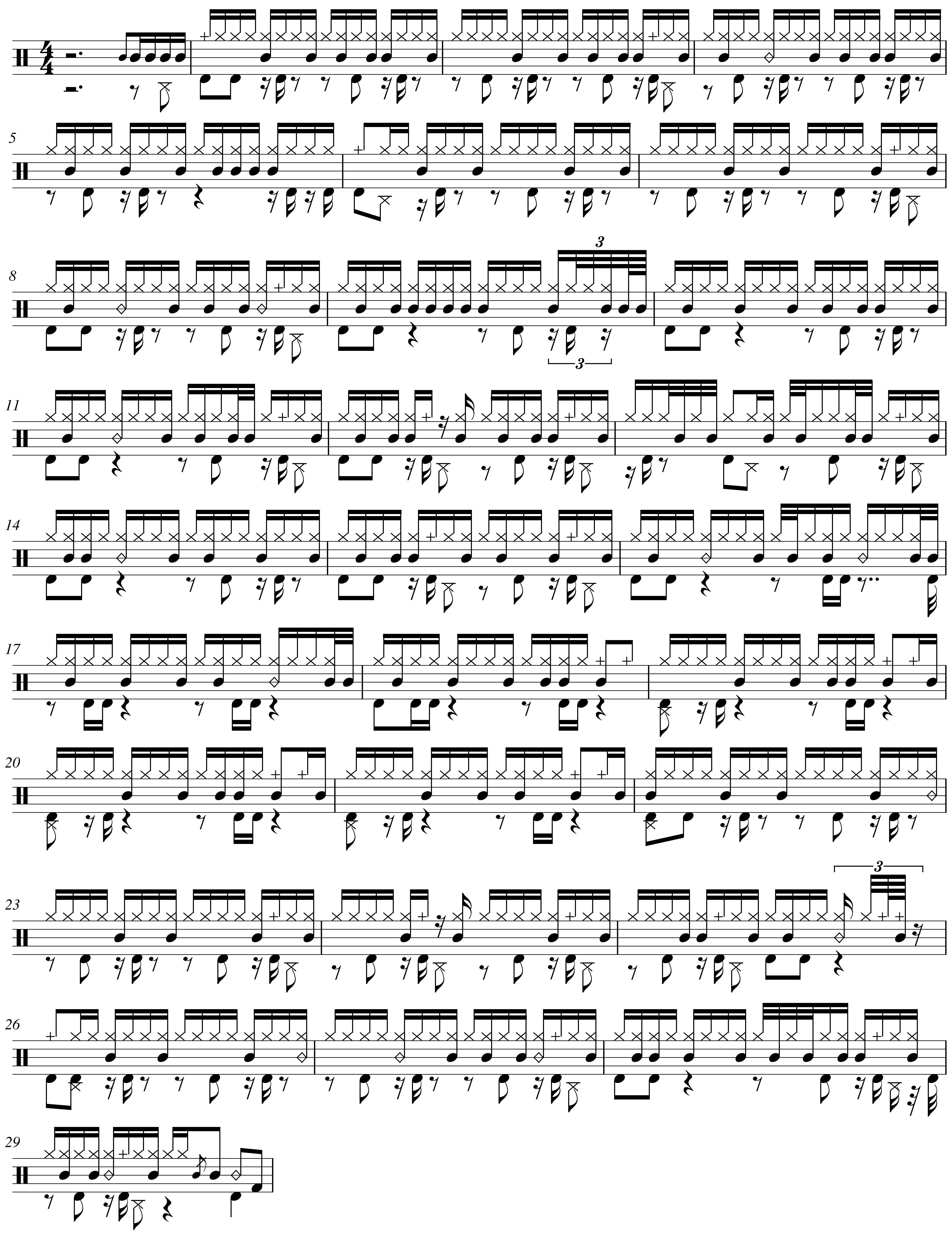
FiloBass by John-Xavier Riley (C4DM, QMUL)
EU project “*Dig That Lick*”

- jazz bass lines,
  companion of **FiloSax** collection of saxophone soli
- 24 recorded hours of melodies and improvisations
- automated extraction from youtube jazz videos
- qparse as backend of an audio-to-MIDI
  transcription procedure
- prior beat (measure) tracking

Drum transcription

Groove MIDI Dataset

- by Google Magenta
- 13.6 hours, 1150 MIDI files ~ 22000 measures recorded by professional drummers on a electronic drum kit
- audio (wav) files synthesized from (and aligned to) MIDI files for evaluation of audio-to-MIDI drum transcription
- no score files!



Scoring the GMD with qparse

Martin Digard (INALCO), Lydia Rodrigez-de la Nava

- all score files (XML) produced from the MIDI files with the same generic tree grammar (4/4 measure)
- polyphonic case-study, simpler than piano
- specific drumming constraints (hands ≤ 2, feet ≤ 2)
- processing errors from MIDI sensors

- **Voice separation** - Lydia Rodrigez-de la Nava, evaluation Augustin Bouquillard
  integration for piano guitar transcription:
  - before parsing, or
  - after parsing (on intermediate model), or
  - joint with parsing.

- **Dataset ASAP** - Francesco Foscarin, Andrew Mc Leod
  MIDI and audio recording from Yamaha piano competition
  + XML scores
  + alignments
  + beat tracking annotations

Structured Music Score models
hierarchical representation of music scores
finite representations of languages (*style*)

Search and Retrieval
indexing
exact and approximate
search and query

Similarity metrics
string and tree
edit-distances

H2020 Polifonia
U. Bologna
Open University
King's College London
Vrije U. Amsterdam

IReMus (Paris)

DCML (EPFL)

EDA (Paris Cité)

ANR Collabscore
IRISA (Renne)
BnF
fondation Royaumont

JSPS 採譜
Nagoya U. (Sakai lab)
JAIST (Tojo lab)
grant Yamaha M.F.

AlgoMus (Lille)

Digital Music Score Collections
Cultural heritage preservation and study
neuma.huma-num.fr

Digital Humanities
Computational Musicology

Optical Music Recognition
Crowdsourced correction

Automated Music Transcription
qparse

Structured Music Score models
hierarchical representation of music scores
finite representations of languages (*style*)

Search and Retrieval
indexing
exact and approximate
search and query

Similarity metrics
string and tree
edit-distances

H2020 Polifonia
U. Bologna
Open University
King's College London
Vrije U. Amsterdam

IReMus (Paris)

DCML (EPFL)

EDA (Paris Cité)

ANR Collabscore
IRISA (Renne)
BnF
fondation Royaumont

JSPS 採譜
Nagoya U. (Sakai lab)
JAIST (Tojo lab)
grant Yamaha M.F.

AlgoMus (Lille)

Digital Music Score Collections
Cultural heritage preservation and study
neuma.huma-num.fr

Digital Humanities
Computational Musicology

MERCI!

Optical Mu
Crowdso

Automated M

"I'll be Bach"