Music Informatics

Alan Smaill

Jan 18 2016

Today

- More on Midi vs wav representation.
- Rhythmic and metrical analysis.

Recall difference between procedural midi representation (key press and time based), and digital versions of sound wave representations. There is a trade-off between the expressiveness (the fine details of performance) and manipulability (allowing abstract analysis).

Comparing representations

There are obvious strengths and weaknesses between these different representations; consider some musical manipulations that musicians do fairly often on the basis of listening to another musician:

Task	midi	wav
copy	✘	✔
keep melody, change instrument	✔	✘
add echo	?	?
generate score	✔?	✘
transpose, same tempo	✔	✔?

midi to wav/mp3

This is one of the main uses of midi, and is supported by many tools.

midi files indicate orchestration for each channel as a “Program change” message; the resultant sounds depend on the quality of an associated synthesiser. The standard specifies timbres in terms of common musical terms (eg as a vibraphone note).
Compare sound output direct from such a midi synthesiser, and via digital or acoustic piano:

`http://www.piano-midi.de/`

It’s easy to get the notes played by (something sounding like) other instruments.

As we expect, this is much harder; midi files are much much smaller than audio, giving a view of the sounds that is biased by the discrete pitch set in the standard use of midi.
A small example that just about succeeds in doing this is at:

`http://www.pluto.dti.ne.jp/~araki/amazingmidi/`

This is polyphonic music (ie several notes are played at the same time), but it's on an instrument that only makes notes at semi-tone intervals. As the site says, this is harder with sung music. Even one solo voice is hard; in that case we can track pitch fairly well, but the mapping from pitch to midi note can get misled very easily.

Note that:

- The resultant midi file is much smaller (422 KB goes to 1 KB);
- Works well where pitches are stable (keyboard instruments);
- Works badly for vocal music;
- Rhythm copies over well (no quantisation, compared to pitch).

Metrical analysis

We'll now consider the temporal organisation of music.
In particular, we'll take a first look at metrical organisation as mostly used in western tonal music.
This is characterised by

- a regular underlying pulse
- a regular hierarchical grouping (and/or subdivision) of pulses
- in groups of 2, 3 or 4

Most people can pick up on dance rhythms, and recognise say the regular 3 pulses in a bar (measure) in a waltz.

Example

Music given a time signature of 6/8 has three levels of grouping, the underlying quaver (eighth not) being grouped in threes, in turn being grouped in twos. Depending on how fast the music is, the listener may tap along at any of these levels – the level at which the pulse is sensed is called the tactus (could be at any of these levels, depending on speed):

```
bar:      x                       | x                       |
mid:      x           x           | x           x           |
lowest:   x   x   x   x   x   x   | x   x   x   x   x   x   |
```

Here we distinguish between rhythm, as in a (short) sequence of organised duration, and metrical structure which involves longer scale setting up of expectations of hierarchical layers.

Metre and notes

Note that this underlying framework can be part of the organisation even when the notes played do not coincide with the beginning of the metrical units (as in syncopated music). This organisation is also heard to persist even during a slowing down or speeding up of the underlying pulse.

For a lengthier account of the issues, see for example in pp 22-26, Scruton, Aesthetics of Music, OUP, 1997. Scruton points out that the German philosopher Leibniz described music as "a kind of unconscious calculation": the beat is thus measured out, during the stretching and contraction of time found especially in romantic music of the 19th century (rubato).

All this suggests that it is a hard task to analyse metrical structure by computer, even under the simplifying assumptions made so far. Notice that the task involves a cognitive dimension: how is the metre experienced? And the answer is probably different for different people.

However, to create a good test situation we can follow Longuet-Higgins (Mental Processes, MIT Press, 1987).

Although this is old work, it is a good example of experiments with a hand-crafted rule set, designed to correspond to judgements of human musical listeners (familiar with music of a particular style). So, nu machine learning here ...

The problem set starts from music with a score, and given time signature, so that:

- we know the composer's own specification
- there is a single line of music
- the music involves little or no differentiation in volume (no strong accents)

Rhythmic analysis of Bach Fugues

Longuet-Higgins (and Steedman) looked at analysing the metre from the initial statement of the fugues from Bach's 48 Preludes and Fugues. At that point of each piece, there is only a single line being played; these were played on early keyboard instruments originally (clavichord, which does not have a big dynamic range, & harpsichord where the volume is fixed).

To further simplify, take as input a sequence of durations as multiples of an appropriate unit of time. This means that it is given in the input that the semiquaver is half the length of the quaver, for example. But no information about the time signature or bar-lines is given.

Bach Fugue C minor, book 1 of the 48

Subject of fugue (first two bars only 1 voice)

Basic Patterns

Some terminology, related to terms from syllable lengths (the names are not important here):

- dactyl: long, short, short
  (where the last short is recognised as short because another note starts quickly)
- spondee: long, short (not followed by another short)

This is still underdefined;
– for example what about lengths [4,2,1,1...]?
– Is the 4,2,1 a dactyl?

Main idea

Because in the Bach example these are the first notes the listener hears, and the aim is to model perception, it is expected that:

- the listener builds up the analysis incrementally
- the lower levels (shorter scale) are perceived first
- higher levels are built on lower levels at acceptable multiples of the current level's pulse

Justification

Why does this make sense as a model of understanding metre?

The claim is that

> *The progressive nature of the listener's comprehension is made explicit in an assumption about the permitted order of musical events in an acceptable melody. This assumption we call the "rules of congruence", and it is fundamental to the operation of both our harmonic and our metrical rules.*
>
> *Longuet-Higgins and Steedman, p 84 of L-H above*

Thus a note that fits the metre locally is congruent;
a note which is stressed rhythmically by the context, but not by the local metre (syncopated) is metrically non-congruent.

What is a good analysis?

The outcome could be any of these:

- time signature and bar-lines as in metrical analysis
- time signature and bar-lines as in metrical analysis, with grouping of bars
- time signature as in metrical analysis, but out of phase (eg 4/4 with bar-line displaced by half bar)
- metrical analysis correct but stopping beneath level of bar
- metrical analysis wrong at some level of the hierarchy

(second better than first, if right?, others not so good, last worst)

Algorithm: outline

Suppose we are at start of subject, or at start of current metrical unit, and first 3 notes are $n1, n2, n3$:

```
if at start of dactyl
   if duration of dactyl good multiple of current unit
      adopt duration as higher metrical unit
   else  if len n1 − (len n2 + len n3) good multiple
         adopt this length as higher metrical unit
if at start of spondee
   if len n1 − len n2 is good multiple
      adopt this as higher metrical level
if neither of above,
  & first note lasts n current metrical units
      if n is good multiple
        adopt this as higher metrical level
otherwise  keep current metrical analysis,
           and move to next pulse at this level.
```

Look at fugue 2, book 1, (C minor). The input is:

$$[1, 1, 2, 2, 2, 1, 1, 2, 2, 2, 1, 1, 2, 2, 1, 1, 4, 1, 1, 1, 1, 1, 1, 1, 1]$$

```
* First note: establish 1 as initial length
* Note 2 conforms
* Note 3 doubles length, and starts new metrical level
* Note 5 is start of dactyl; double length, start new level
```

Note that the crotchet (4) does not disturb this analysis, because it is not at the start of a pulse at the established level.

So analysis here gives:

```
nts:   x x x   x   x   x x x   x   x   x x x   x   x x x       x x x
l1:    x x x x x x x x x x x x x x x x x x x x x x x x x x x x x x x
l2:        x   x   x   x   x   x   x   x   x   x   x   x   x   x   x
l3:                x       x       x       x       x       x       x
```

Going a bit further

The algorithm so far does not go higher in the hierarchy. But listeners do hear at least another level.
A further stage involves marking metrical units themselves, and not just individual notes. Say a unit is marked for accent if a note or dactyl starts at the beginning, and lasts throughout the unit.
Now use the "isolated accent rule":

```
if a unit is marked for accent
   and is followed by 1 or more unmarked units
   which are followed by a marked unit
   which is followed by an unmarked unit
then the interval between start of marked units
   forms a new level in the metrical hierarchy
```

In the fugue at hand, the isolated accent rule applies between notes number 5 (start of [2,1,1]) and and 10 (another [2,1,1]) and this establishes metrical level of minim (the half bar level in Bach's notation).
Units marked for accent marked by `A`:

```
nts:    x x x   x   x   x x x   x   x   x x x   x   x x x         x x x
l1:     x x x x x x x x x x x x x x x x x x x x x x x x x x x x x x x
l2:         x   x   x   x   x   x   x   x   x   x   x   x   x   x   x
l3:                 x       x       x       x       x       x       x
                    A               A
```

The new analysis:

```
nts:    x x x   x   x   x x x   x   x   x x x   x   x x x       x x x
l1:     x x x x x x x x x x x x x x x x x x x x x x x x x x x x x x x
l2:         x   x   x   x   x   x   x   x   x   x   x   x   x   x   x
l3:             x       x       x       x       x       x       x
l4:             x               x               x               x
```

This is a reasonable final analysis, though it is at the half-bar level.

Starting elsewhere

Notice that if we take the input starting later, we can get a different analysis. Suppose we start as follows:

Now the “off-beat” dactyl is heard before the higher level is established, so we get a different analysis at level 3 (`l3`).

```
nts:   x x x   x   x x x       x x x x x x x x x x x
l1:    x x x x x x x x x x x x x x x
l2:        x   x   x   x   x   x   x
l3:            x       x       x
```

Overall success on Bach examples

This algorithm was run on the 48 examples in Bach's 48 preludes and fugues.

- Mostly analysis is correct, stopping at at sub-bar level.
- Sometimes gets Bach's given metre, or grouping of bars.
- Some "interesting mistakes".

Metrical ambiguity

The algorithm can be seen as a way to parse a particular sort of metrical structure.
It does not allow ambiguity, but sticks with an initial analysis, even when later evidence mounts up against this.

The algorithm can be extended to maintain different metrical hypotheses, where one wins out, perhaps displacing earlier leading candidates.

In an English Country Garden

This is analogous to "garden path" sentences in natural language, where parts of sentences are understood one way during incremental analysis, only for that analysis to be discarded later on when more information is available.
Compare:

The man who hunts ducks . . .

completed as . . .

The man who hunts ducks out on weekends.

The deliberate play on metrical (and other sorts of) ambiguity in music is more prevalent in music than in natural language, and is part of what makes music interesting (and difficult to process by machine).

Metrical garden paths

For an example where the initial metrical analysis is contradicted by subsequent rhythmic input, look at paper by Peter Vazan, Michael F. Schober, "Detecting and resolving metrical ambiguity in a rock song upon multiple rehearing".

`http://www.icmpc8.umn.edu/proceedings/ICMPC8/PDF/AUTHOR/MP040242.PDF`

Overlaying different metrical patterns?

Some other problems in metrical analysis are illustrated by the following example from Ravel, with simultaneous presence of different temporal organisations. How could adapt the earlier ideas to analyse music like this?
Listen to repeated pattern of four equal notes at the start of Ravel's "Rapsodie Espagnole".

Now listen on, and note the time signature which is not what the listener expects from hearing the initial section of music.

More metrical ambiguity

Summary

- Converting between representations.
- Metrical organisation in WTM
- Metrical “parsing” and metrical ambiguity.