
Chapter 4

Understanding musical interaction

4.1 Introduction

When thinking about a discussion between a group of people, we can look at the words (sounds) that were spoken, and make some sense of those; we can look at the focus of the discussion (turn-taking), and how that moves around; we can look at the sorts of things being done with words – giving information, disagreeing, persuading . . .

We can also imagine various ways to record such a discussion, ranging from a sound/video record, to a written (tidied-up) account of who said what, to a summary of the points raised. For a written account, the transcription process can be done by people (on the basis of lengthy education in the language involved), or by current natural language processing techniques that do this more or less accurately.

Interaction between musicians in performance raises similar issues, with important differences too. Here we want ways of analysing the musical material in terms of some building blocks of music; we may be interested in the grammar associated with particular styles, and the rhythmic and pitch organisations employed. Automation of these processes is a challenging field, and in this chapter we will look at some current approaches.

We will look at some of the techniques and algorithms used in the analysis of music, taken as sounds organised in a structured way similar to that of natural language, and also ideas involved in understanding musical discourse.

Current computational technology allows machine-generated music to be incorporated into musical performance and improvisation in interesting and musically productive new ways, and this topic is also covered in this chapter.

Mostly we will use as examples music from **Western tonal music** (WTM) – roughly, music based on the organisation of musical pitch and rhythm common in Western Europe from the 16th century, in classical music up to the early 20th century, and used in most current (Western) popular, jazz and folk music. There are many other musics in the world, raising similar questions, and contemporary composers often work outside this framework.

For those without a basic background in terms such as **metre, rhythm, pitch and key**, you should look at an introduction to music theory, for example the first fifteen very short lessons on the [8notes.com](http://www.8notes.com) website.¹ You should bear in mind that American and British English have some differences in vocabulary, listed as “Alternative terms”.²

¹<http://www.8notes.com/theory>

²<http://dolmetsch.com/introduction.htm>

4.2 Sound, music, score . . .

Here we take music to be primarily an experience of sound in time; but this is not enough to distinguish music from sound in general. There is no general agreement as to what differentiates music from sound. A start to an answer is, for example, given by Roger Scruton:

Music is an art of sound. . . . Nor is it the work of a musician to write poetry, even though poetry too is an art of sound. So what distinguishes the sound of music?

The simple answer is 'organization'. But it is no answer at all if we cannot say what kind of organization we have in mind. . . .

(Scruton 1997, p 16)

In this chapter we look at some of the ways Western tonal music is organised. There are different aspects to this organisation, which are of differing importance in other musics. Some of these are listed below.

Rhythm This concentrates on the time dimension of music, and is seen most clearly in drum and percussion music in general. It underlies nearly all WTM, from Beethoven to the blues.

Pitch An individual note, say one note played on a piano, is associated with a given pitch, associated with the physical frequency of sound waves in the air – each note on the piano or guitar has a different pitch. The way that pitches are organised together is fundamental to many musics.

Intensity The same note or drum beat may be played louder or softer; this dimension of organisation is called **intensity** and is associated with the energy level of the sound waves involved.

Timbre Here we are interested in the difference between notes of the same pitch and intensity played by different instruments, or sung by different people. The physics of this is more complicated; when guitarists change guitar between numbers, it is often about using a different timbre as part of a different atmosphere.

Much music is passed between musicians by playing, listening, and imitating. There are some spoken or sung conventions in describing instrumental music to other musicians, which constitute local specialised languages; see for example bols associated with Indian tala³ (this music is rhythmically complex).

Written notation for WTM is traditionally central to classical musical performance. It reflects in particular the ways in which the pitch and rhythm are organised in this music. It is not important, for this chapter, to be able to imagine music on the basis of seeing such notation (this is something like learning a new language). However, the so-called **score**, which is a (fairly loose) specification of what the musicians are to perform, is important for many kinds of musical activity.

Historically, the task of producing a score involved detailed and laborious work. Nowadays, while we can do a reasonable job of turning spoken speech into written text automatically, the corresponding problem for music, say for producing a score from music improvised at the piano, has not seen such progress.

³<http://kksongs.org/talamala.html>

When we look at musical interaction later, we will need the notion of a score which gives the performing musicians some shared understanding of what musical actions are involved at different moments. If you want to see what an orchestral score looks like, look at Beethoven's Symphony no. 5.⁴

We will now look at some of the issues in machine analysis of music; in particular, ways of getting machines to help with recognition of the local structure of music in terms of rhythm and of pitch.

4.3 Rhythm, beat and metre

By **rhythm** we mean a pattern of accentuation of events in time, for example the 'short, short, short, long' at the start of Beethoven's Fifth symphony. These are concrete patterns, which suggest and are supported in WTM by longer-term structuring of accents in time. Here we look at two of the ways in which music in WTM is structured at the medium scale.

4.3.1 Beat

One ability needed when making sense of musical interaction is a way to recognise where the beat is when listening – this is where the listener will tap her finger or toe, and the regular beat that coordinates most dance. Most people do this naturally, but this is not easy for computer analysis. This basic ability is a basis for the coordination of musical performance where different musicians play in an ensemble, or where virtual computer musicians produce music in time with other musicians, human or virtual. This is the problem of **beat tracking**.

Some computer-generated music produces music where the beat is perfectly regular: within the limits of the technology, beats occur exactly every 0.75 seconds, for example. This is fine if this is what is intended. But this is unnatural for human musicians, who find keeping time alongside music produced this way reduces the possibility of shaping the tempo, which is basic to musical performance. It also means that the human must follow the computer. In a small ensemble of human musicians, the players listen to each other to coordinate the beat, which will vary by some amount during playing; the lead in shaping the tempo is taken by different players at different times.

For human listeners, it is not hard to pick up the beat when listening to music in a known style in real time. For machine beat-tracking in real-time, the input can be taken in two ways:

1. as audio input;
2. by tracking the physical gestures of the performers, rather than the sounds produced.

The latter can simplify the problem, because the places of the start of notes in time are usually then made obvious, and this is the main information used in beat-tracking algorithms.

⁴<http://www.dlib.indiana.edu/variations/scores/bgp5237/large/>

The most usual format used here is that of MIDI (Musical Instrument Digital Interface), which started out as a way of describing key press and release actions on piano-like keyboards. Nowadays MIDI versions of many instruments are available, and there are several software libraries supporting manipulation of MIDI data, for example in Java.⁵

Some sample .wav files used for testing beat-tracking algorithms can be downloaded by following the instructions from an ISMIR conference.⁶ These give an idea of the variation of genre that such algorithms attempt to cover. You might like to try tapping along with some of these. It is not the case that everyone taps along in the same place – for example sometimes it is possible to tap twice as fast as someone else, and both make musical sense. We will look at this issue again later in this chapter.

There are some simple cases, such as some music with loud regular percussion (drum) beats, where attending to these sounds is the key to detecting the beat. But other cases, such as jazz guitar, can play without strong accents on the beat, with some accents off the beat (syncopation), and the human listener still follows the musical beat.

There are several very different algorithms that have been used here. Let us look at the case where we have MIDI input.

A system of virtual musical agents is described by Wulfhorst et al. (2003), where a simple beat-tracking algorithm based on MIDI tracking of the notes played by other musical agents is given. Suppose an agent keeps a record of recent events (note onsets) by time, and suppose the agent has an initial expectation of how frequent the beats will be. The agent can consider mapping these events as radial lines on a circle, for a notional beat at times $t, t + I, t + 2I, \dots$

- the angle on the circle corresponds to the times with respect to the beat, so that events occurring time I apart are at the same angle.
- more recent events are weighted more strongly (or events from a small time window are used).

If I is a good estimate of the time between beats, we expect to see events clustering at angles around the circle; if I is a bad estimate, the events will not show a pattern.

So by looking at a number of candidate intervals around the expected beat, the agent can find the best estimate to the current beat — and then use this estimate to make a musical gesture itself in relation to the current beat.

The clustering in the example on the right suggests that there are events happening regularly at each quarter beat. This happens often (think of percussion **filling-in**), and suggests that there is more going on than a level repeated beat; we now look at this aspect.

4.3.2 Metre

Typically in WTM the underlying pulses of music are organised hierarchically; in the case above, the tapping level beat is subdivided into four. Beats themselves fit into larger repeated units, usually in groups by 2, 3 or 4, and they themselves in

⁵<http://www.midi.org/aboutmidi/index.php>

⁶http://www.music-ir.org/mirex/2006/index.php/Audio_Beat_Tracking

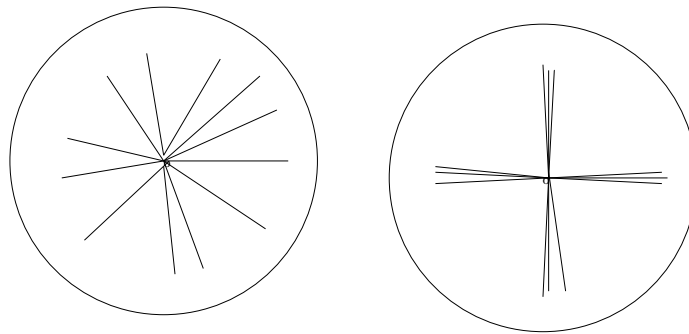


Figure 4.1: Bad (left) and good (right) choice of period

larger groupings. The bar (measure), marked in standard notation by a vertical line, is usually one or two levels above the beat; bars themselves fit into regular patterns too, as in the 12-bar blues. This hierarchical structure is called the **metrical** structure of the music. The so-called **time signature** written at the start of a piece indicates this hierarchy at the level of bars and beneath.

Here is an example of rhythmically simple music, based on a dance pattern. The rows of dots beneath indicate levels of metrical grouping. The beat is usually heard at row 2, which is divided in four. The beats group together two by two. In each case, the larger constituent grouping is taken to give the more important temporal grid – not necessarily where accents fall most often, as syncopated music shows. The relative importance, or salience, of different levels of metrical organisation has been studied extensively by psychologists (London 2004). It has been shown that the most salient level of grouping in a metrical hierarchy tends to have an inter-onset interval of around 600 ms. This means that groupings that are separated by around 600 ms in time are more likely to be perceived as the beat (Parncutt 1994).

0:
1:
2:
3:
4:

Figure 4.2: Metrical hierarchy in a Bach gavotte

There is ambiguity at level 3 for many listeners: while there is agreement that the level two units are grouped in twos, it is ambiguous how that grouping is done – either the first two units at level two are combined, or the second and third. You can

listen to this and decide what you think (here played rather quickly).⁷ It is a challenge to find these metrical levels automatically, in a way that fits with natural human abilities; in cases like a waltz, we find it fairly easy to recognise that there are three beats in a bar, and to tell where the first, strong, beat occurs. Again, there are different approaches to this task, which is in general harder than beat tracking. The interested reader can look, for example, at David Temperley's book on musical cognition (Temperley 2001*b*).⁸

There are many other ways to organise rhythm in music – see the brief overview mostly within WTM.⁹

4.4 Pitch and key

We now look at how different pitches play a role in WTM, and how computer analysis can support interactions that depend on this organisation. Each note on a piano or guitar corresponds to a different pitch, and these instruments are designed to play only a subset of the possible pitches (unlike the human voice, or violin). The way notes on the piano are laid out with white and black notes, and a pattern repeating after seven white notes, comes from the particular organisation of pitch in terms of **keys**.

Keys are organised in two main ways (major and minor) around a key note (called the **tonic**); the tonic acts as a pitch centre or home note, where the melody typically ends. Listen to these musical examples.¹⁰

Take a look at this accurate account.¹¹ The bluffer's very short guide is here.¹²

Knowing the current key is important in musical interaction – for example, in folk music, when someone wants to follow without a break to a new tune of their choice, they will announce the key (rather than the tune), and the musicians then have the right context to accompany the new melody.

Notes which differ in that the fundamental frequency of one as a sound wave is twice that of the other are given the same name, and not distinguished as tonic notes (they are said to be an octave apart). It is natural when men and women sing a melody together that they sing an octave apart, yet are understood as singing 'the same notes'.

There is a geometric way to think of the organisation of pitches that leads to an interesting approach to key detection. In this approach, different pitches appear several times, because their role is ambiguous depending on which key they are analysed as belonging to. By looking at the assignments of pitches which occur, to their possible places in a 2 or 3-dimensional structure, and looking for the assignment which gives the most compact representation in that structure, a candidate can be computed as the most likely key for the set of pitches concerned.

A recent system that exploits this approach is given by Mardirossian & Chew

⁷<http://www.we7.com/#/track/French-Suite-No-5-in-G-BWV-816--Gavotte!trackId=2179255>

⁸<http://www.link.cs.cmu.edu/cbms/index.html>

⁹[http://reference.findtarget.com/search/Meter%20\(music\)/](http://reference.findtarget.com/search/Meter%20(music)/)

¹⁰<http://ask.metafilter.com/6444/Music-Just-what-exactly-is-key#132028>

¹¹[http://wapedia.mobi/en/Key_\(music\)](http://wapedia.mobi/en/Key_(music))

¹²http://www.soundfeelings.com/products/music_instruction/eproducts/key.htm

(2008),¹³ where Table 1 shows the arrangement of pitches in a two-dimensional array.

4.5 Musical grammar and style

4.5.1 Grammar

Given some understanding of how pitch and rhythm are organised, we can now think of how music is structured at higher levels. One approach here is to make use of notions from **grammars**, as applied to both natural languages like English, and computer languages like Pascal.

A grammar gives a set of rules by which combinations of some entities (words, symbols, sounds . . .) can be categorised as OK or not, according to the grammar. When a compiler complains about a syntax error, it uses the syntax rules of the language at hand to determine whether the input really is a statement in that language or not. Similarly a musical grammar can be used to characterise music in a given style, say – what is it that makes one piece a blues piece and another not?

Grammars have been very productive in providing ways to manipulate natural and computer languages, and they can play an important role in machine understanding and generation of music also. A grammar can be used not only for recognition, but also for generation: given a grammar for, say, Java, it is possible to generate automatically well-formed Java programs from a specified subset of the language. There is no guarantee that these programs do anything interesting, but we can ensure that there are no syntax errors; if some randomness is built into the generation process, then this gives some sort of fair sample of the possible programs. Musical grammars have been used for both recognition and generation.

An example of this is Mark Steedman's grammar for the blues, at the level of chord sequences (Steedman 1996).¹⁴ The initial grammar rule given there is this:

12bars → I I7 IV I V7 I

It says that the twelve bars can be made up of six successive sections, based on the chords:

I I7 IV I V7 I

(in C major, this is C, C with flattened 7th, F, C, G with flat 7th, returning to C). Other rules allow variations by substituting different chords, in given contexts; these substitutions are known (in different form) to blues musicians, and are here incorporated into a grammar. The discussion in the paper on different grammar formalisms is beyond the scope of this chapter, but the ideas are important. Given a sequence of chords in this way, the grammar can be used to see if this is a possible blues sequence.

This grammar does not operate on the level of the individual notes played and sung, but already assumes a more abstract understanding in terms of successive chords in a given key – this is called the **harmonic** level of description.

¹³<http://en.scientificcommons.org/42491827>

¹⁴<ftp://ftp.cogsci.ed.ac.uk/pub/steedman/music/batat.ps.gz>

A much more elaborate approach to WTM using grammars is given by Lerdahl & Jackendoff (1983); in principle this applies across the board, with metrical as well as harmonic and melodic aspects included. It is, however, given somewhat informally, and does not get us easily to a practical way to parse WTM using all these aspects.

4.5.2 Style

There has been a lot of work on determining musical style automatically, from sound or from symbolic description. Statistical and machine learning techniques are proving to be good at this recognition task, by looking for patterns in the music. (A survey of some work using audio is given by Tzanetakis & Cook (2002).) This does not give a way to **generate** music in a given style.

David Cope has shown how it is possible to generate music in the style of a given composer, or genre, by a combination of analysing a number of pieces in the target style to look for common patterns (in this case, pairs of combinations of notes that listeners are likely to take as similar, with small changes in rhythm or pitch outline, for example). These become indicators of the style, but for the overall shape that the music should take, he supplies a hand-written grammar. Thus we can imagine new piano blues by combining a grammar, like the one mentioned above, with patterns of piano notes repeated across performances by a particular blues piano player.

See David Cope's homepage,¹⁵ especially the section on Experiments in Musical Intelligence, with MP3s of machine-generated music in various styles, using the techniques he describes (Cope 1996, 2001).

4.6 Musical discourse

Let's now consider what might be used to support musical interactions mediated by machine. There are several scenarios where people want to do this sort of thing.

Accompaniment Systems to provide accompaniment to a soloist where the accompaniment and soloist are following a musical score. In this case, the system needs to track tempo changes made by the soloist, and deal with some amount of deviation from the scored notes, as a human accompanist would do. There are some impressive systems that achieve this; see this short article¹⁶ and Christopher Raphael's 'Music plus one' page.¹⁷

Distributed ensembles There is also a desire to allow musical collaboration between people and virtual musicians who are in physically distant locations, perhaps as part of some larger production. The technology used for remote meetings in general is an obvious place to start, but there are possibilities for more sophisticated interactions that support musical interaction better. For a taster of this work, visit the telematic Circle.¹⁸

Real-time transformation of musical material With powerful signal processing available, musical sounds can be transformed in real time in a myriad of ways. Thus the sounds of a human performer can be mutated during performance, by

¹⁵<http://artsites.ucsc.edu/faculty/cope>

¹⁶www.jstor.org/stable/20055566

¹⁷http://xavier.informatics.indiana.edu/~craphael/music_plus_one

¹⁸<http://www.deeplistening.org/site/telematic>

someone controlling these transformations, or fed as input into other music sources. What is more interesting here is to have these transformations under the control of the performer themselves, say by using some physical gesture that triggers a change, or doing this by some feature of the music itself. A very brief description of the former is here.¹⁹

New interfaces to (new) instruments There are new possibilities of ways of making sounds from traditional instruments, or from new instruments that may only exist in virtual, digital form. The physical acts involved in making music, by singing, blowing a wind instrument, and so on are part of the experience of making music which feel 'left out' with a push-button interface. Look at the short description of various interfaces at the start of the Stanford CCRMA course web page.²⁰

Human/machine improvisation People have been working for some time towards the idea of having virtual musicians join in with human musicians, especially in music with a significant element of improvisation, where the music does not follow a score or a memorised model, but is produced spontaneously during performance. As we have seen, in music in the WTM tradition the improvised music is expected to make sense in its rhythmic and harmonic context (although this does not hold for all music, by any means). Given that human performers have some knowledge of the style in which the improvisation will evolve, it makes sense for a machine participant to have knowledge also of the style, both to follow the playing of others, and to generate suitable music itself. Having a grammar of the style is one way to achieve this.

In mixed ensembles of this sort, the aim is to enable shared guidance of the improvised music, so that the human is not subservient to the machine, nor the machine to the human. One interesting approach is the MIMI system from the University of Southern California. This page has some videos of its performance while improvising with a human in several styles.²¹

4.7 Exercises

1. Beat tracking

Some music is intended to have a regular underlying pulse, but other forms of music are organised without any such perceptible regularity; sometimes there is a mixture; sometimes the beat is not obvious initially, but once established is easy for people to follow.

Listen to the start of the following extracts, and see how easy or hard it is to tap along regularly with the music. What features of the music help or hinder in this task? How could a computational listening device make use of the helpful features?

- (a) Byrd *Pavane*²²
- (b) Bach *Two part invention*²³
- (c) Debussy *Clair de lune*²⁴
- (d) Boulez *Dialogue de l'ombre double*.²⁵

¹⁹<http://www.infomus.org/Research/Mapping.html>

²⁰<https://ccrma.stanford.edu/courses/250a/lectures/survey/>

²¹<http://www-rcf.usc.edu/~mucoaco/MIMI/>

²²<http://www.mfiles.co.uk/mp3-downloads/Earle-of-Salisbury.mp3>

²³<http://www.mfiles.co.uk/mp3-downloads/invention2part-no4-el.mp3>

²⁴<http://www.mfiles.co.uk/mp3-downloads/debussy-clair-de-lune.mp3>

²⁵http://www.last.fm/music/Pierre+Boulez/_/Strophe+VI

2. Hierarchical metre

- (a) Often rhythms are organised around a hierarchy, above and below the level of the beat. For example, the following music is written with an underlying beat of about 100 beats per minute, with the beat subdivided on 4, while the beats are organised in groups of three, where the first of each group of 3 is stronger. Listen to the start of this and listen out for this structure.

Albeniz, *Leyanda*.²⁶

- (b) Usually beats are grouped in groups of 2, 3 or 4. Grouping in 5 or 7, for example, is unusual. Listen for the unusual grouping here – the conductor’s movements reflect the grouping (but not very obviously). Tchaikovsky’s *6th symphony, second movement*.²⁷

Also try Pink Floyd, *Money*.²⁸

3. Pitch organisation

Western tonal music takes the pitches it uses from the notes playable on a piano and guitar. Other traditions make different distinctions, fewer in some folk traditions, more in Indian music.

Listen to this piece which uses twice as many pitches as normal, by having pianos so that the notes on one are slightly higher than the notes on the other.²⁹

Because we are not used to making sense of this sort of music, it can sound ‘out of tune’. Does this music ‘make sense’ to you? What, if anything, does it mean for something to “make sense” as music?

4. The musical dice game

A method of putting together a large number of possible pieces of music from basic building blocks was produced in the 18th century. There is information about the game here.³⁰

You can use the first implementation to generate a piece; you will need to be able to play midi sound files to listen to the result.

This can be compared to a way of generating poetry by selecting from possible first lines, possible second lines, and so on — Queneau did a version in French; there is an English version on-line.³¹ Notice that this respects the form of a sonnet, in terms of syllables and rhyme scheme.

In each case, the object is to produce an output which is well-formed (as a minuet, or as a sonnet). Explain how the building blocks are designed so that they will fit together appropriately. In the musical case, this should involve the organisation of metre, and of pitch and key – other organisational properties may be relevant also.

4.8 Further reading

For mathematical, rather than computational background, the work of Fauvel et al. (2003) provides a readable account of the interaction between music and mathematics. It is accessible for those with some knowledge of musical notation and comfortable with non-specialised mathematics.

²⁶<http://www.mfiles.co.uk/mp3-downloads/leyenda.mp3>

²⁷<http://www.youtube.com/watch?v=2SEDU8AyxVU&feature=related>

²⁸http://www.last.fm/music/Pink+Floyd/_/Money

²⁹<http://www.youtube.com/watch?v=EU85bUyDPWs>

³⁰<http://webplaza.pt.lu/public/mbarnig/pages/dicemus.html>

³¹<http://www.bevrowe.info/Poems/QueneauRandom.htm>

Two important research centres for work in this area are IRCAM in Paris³² and the Stanford CCRMA.³³ Both have several research groups, and wide interests, as well as strong links to musical performance.

Current research is scattered across several areas and journals. *The Computer Music Journal*³⁴ has many relevant articles, and is available on-line.

³²<http://www.ircam.fr/?L=1>

³³<https://ccrma.stanford.edu/>

³⁴<http://www.mitpressjournals.org/cmj>