<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>About TreeTorika: Rhetoric, CAAC and Mao (Book chapter)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Lindborg, PerMagnus</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2008</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10220/40408">http://hdl.handle.net/10220/40408</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>© 2008 IRCAM.</td>
</tr>
</tbody>
</table>
About *TreeTorika*: Rhetoric, CAAC and Mao

PerMagnus Lindborg

This chapter examines computer assisted analysis and composition (CAAC) techniques in relation to the composition of my piece *TreeTorika* for chamber orchestra. I describe methods for analysing the musical features of a recording of a speech by Mao Zedong, in order to extract compositional material such as global form, melody, harmony and rhythm, and for developing rhythmic material. The first part focuses on large-scale segmentation, melody transcription, quantification and quantization. Automatic transcription of the voice was discarded in favour of an aural method using tools in Amadeus and Max/MSP. The data were processed in OpenMusic to optimise the accuracy and readability of the notation. The harmonic context was derived from the transcribed melody and from AudioSculpt partial tracking and chord sequence analyses. The second part of this chapter describes one aspect of computer assisted composition, that is the use of the rhythm constraint library in OpenMusic to develop polyrhythmic textures. The flexibility of these techniques allowed the computer to assist me in all but the final phases of the work. In addition, attention is given to the artistic and political implications of using recordings of such a disputed public figure as Mao.

**Introduction**

In *TreeTorika*, I deal with rhetoric through the recordings of Mao Zedong’s speeches, travelling further down the avenue I explored in other works such as ReTreTorika for quartet and computer, *ConstituOral* for loudspeakers and the trio *Mao-variations.* In these three pieces, transcriptions of the voice are taken as “found” musical material (as in *objet trouvé*). In the first piece, bits of un-edited recording fuse with the saxophone and the ensemble, making Mao’s voice very much a part of the sonic image. In the later pieces, the original material is segmented, transformed and re-composed, thus departing from the source. The challenge with *TreeTorika* was to propose a faithful transcription of the recording in material terms and at the same time to create a composition of maximal integrity in poetic terms. When the source material stems from such a complex

---

1Before these pieces, I had written works which took aspects of classical rhetoric as points of departure, *e.g.* *Khorsa* for orchestra and *Jag åt bokstaven A som trillade at ur ditt öra* for duo. *TreeTorika* was followed by *SynTorika* for quartet and computer and by *Le mammouth englouti* for loudspeakers.

155
personality as Mao Zedong, I try to make it possible to listen to the music without being reminded about what he was.

Rhetoric as a metaphor for composition

As a musician, I am interested in prosody: the way someone forms and phrases a vocal delivery. Rhetoric is not about what is being said; it is about how something is said. Oratory is not about clarity in public address; it is about manipulating an audience. Rhetoric is the study of oratory and oratory is the subject of rhetoric. There is no such thing as written oratory. There can be preparatory notes for a speech, and an article can be read aloud in plenum – but good writing can make for bad oratory. Furthermore, rhetoric is associated with a Greek tradition, codified by Aristotle, and oratory with a Roman tradition, codified by Herennius.

I am interested in the way speakers use the rhetorical situation, kairos: the particular moment when someone has to speak in public. Kairos demands that the orator (lawyer, teacher, politician...) gauge the situation and respond to it by adapting to an adequate mode of delivery. The art of speech-making lies in understanding the dynamics within the speaker+topic+listener system and using this knowledge to affect the listeners’ mindset. Great orators craft the situation to their advantage: it is about convincing the listeners. Rhetoric can be a metaphor for composition. But does it mean that a work of music is oratory? The answer is no. As much as I like to think that TreeTorika is capable of moving the listener, it does not attempt to convince people to adopt any particular point of view. It is an abstraction: a musical drama involving certain aspects of rhetoric.2

The orator

Mao spoke Hunanese, which many Chinese quite clearly find hard to follow. Several writers have noted his tendency to use a strong rural dialect even after coming to power in Beijing in 1949. His pronunciation must have sounded connotated to urban ears, but he did not give it up and was obviously concerned about showing “peasant pathos” in order to maintain a strong association with his original rural powerbase, at least up to the end of the 1950s. It can be argued that Mao Zedong, a singularly powerful individual with relatively few public appearances, did not see oratory as a crucial means of exercising his power; his preference lay with writing and calligraphy [5].3 After leaving Shanghai in 1927, Mao spent most of the next two decades in poor rural regions, first involved in guerrilla warfare and then in Yan’an. At the time of the Guomintang retreat and the Communist seizure of power in 1949, Mao was the undisputed leader and from that point onwards, his every word was recorded [13] [12].

The recordings I studied are of seven major speeches given between 1949 and 1956 [7]. I have not been able to find earlier audio material. Although it is disputed [2] I think Mao was a skilled orator. His delivery is not flamboyant or aloof, but rather, from a musical point of view, based on rhythmic stability and lively melodious prosody with no extreme fluctuations in dynamics or register. The pauses between phrases are carefully

---

2For a study of classical rhetorics, see [1], [4].

3In this he differs from politicians such as Martin Luther King, Olaf Palme or Benito Mussolini, people for whom the scene, the microphone and the camera was the platform for a public mandate.
measured for maximal effect, and the energy gradually mounts over time. Coughing, grunting and throat clearing are not muted, but become an integral part of Mao’s vocal style, seemingly boosting the alpha male image of the Great Helmsman, and somehow adding power and virility (rather than the opposite) to his persona.

At the beginning of TreeTorika, I included Mao’s guttural noises in the transcription. They appear in all the speeches I have studied. The same persistent coughing which occurred several times simultaneously with different rounds of applause made me realize that the speech had been edited. Apparently, a dedicated and unscrupulous Chinese Communist Party sound engineer had engaged in a bit of cut and paste. Clearly, if the present recording is different from the actual event, it may be a construct in overall duration and even large-scale form. Only a study of the raw recording could indicate what influence such manipulations had on Mao’s rhetorical style, but this is beyond the scope of the present text. Let us then say, once one accepts the recording as “found material”, the difference between recording and “real” speech is irrelevant. Moreover, who can say if the engineer’s version isn’t better than Mao’s?

Preliminary observations on the recording

The recording which forms the basis for TreeTorika is the speech Mao gave at Tiananmen Square in Beijing on October 1st, 1949, in which he declares the communist People’s Republic of China. The duration of the recording is just under 20 minutes; it has a narrow band of frequency and is quite noisy. The modern listener is estranged from the reality of the event by the low fi qualities of the recording.4

Before going any further, let me emphasize that my work is not about the text and its political context, but rather it focuses entirely on a musical understanding of Mao’s oratory.

In any rhetorical situation, the large-scale form must be understood as a dialogue between speaker and listeners. I will start by making some general observations on the present recording. There are two significant sounds: Mao’s voice and the rounds of applause of the audience. The pattern of their interaction may be observed in a sonogram, shown in Figure 1.

![Figure 1. Sonogram of the speech.](image)

The register of Mao’s voice is that of a tenor, mainly covering a range from a low E (MIDI 53) to a high G (MIDI 67), reaching a high C at the end. Particular emphasis on a

---

4However, this is an incentive for using noise expressively, which I did in the pieces mentioned in the Introduction. Samples of Mao Zedong’s voice appear “clean” i.e. de-noised, “raw”, “de-voiced”, deformed by granulation, time-dilated or filtered in various ways. The more transformed the sample is, the less strongly it connotates the source.

5Sound examples of this chapter are available online at [http://www.pnmpm.tk](http://www.pnmpm.tk), under /publications then /TreeTorika.
word is often given by a long glissando, covering more than an octave from a lower register around B♭ (MIDI 46). The way the last few words of a section drown in intense applause indicates that the audience was familiar with Mao's timing. Although, on the other hand, it is not difficult to imagine that officials on the podium gave cues for ovations.

In the sonogram in Figure 1, the rounds of applause appear as clearly demarcated bands. There are 44 in total, out of which 24 are “long”, around 8 seconds in duration, and 18 are “short”, around 4 seconds. The remaining two, approximately 16 seconds long, appear at the very end. Interestingly, rounds of applause, either long or short, are with remarkable consistency of the same duration. This probably has to do with the Party sound engineer’s post-production discussed earlier. The interaction pattern gives clues to the large-scale form: the orator dominates the beginning, applause towards the end is denser and throughout the speech, the phrases of the orator get shorter. All in all, these lines of force create a forward leaning drive. After some of the loud rounds of applause, the voice moves at a significantly slower rate and a lower pitch. It then works its way with increasing intensity to a local climax followed by another loud round of applause. This pattern is repeated six times, each time lasting approximately three minutes. It suggested for my composition a division of the speech into six segments followed by a short coda.

After dividing the speech, I then focused on small differences in vocal sound quality between the segments and decided to expand on them in the orchestration. The first segment is iskhnos, “even” in character; the second is grand; the third is energetic but contained; numbers four and five are grumpy but dramatic; the sixth is triumphant, deinios.

Figure 2. An early sketch of the large form, with preliminary ideas for orchestration that influenced the methods of analysis.

A preliminary sketch is shown in Figure 2, where segments are grouped so as to create a piece in three movements (an enlarged version of Mao-variations, the central movement of which was composed using material from the same Tienanmen Square speech). This initial tripartition was later discarded in favour of a grouping in four parts (1+2, 3, 4+5, 6+coda). Nonetheless, the general direction, speed and timbre of the segments remain essentially the same as drafted. The segments were composed in chronological order, except for the third one, which was finished last. The methods of analysis were developed in accordance with the requirements of the segment at hand.
Analyses

After the large-scale segmentation, the succeeding analyses deal with the voice: first, the focus is on the phrases and in particular the alternation between speaking and silence; second, the vocal line is transcribed into a melody; third, the phrases and the melody are used to determine tempo and quantification; fourth, harmonic material is extracted from the vowels. By quantification, I mean dealing with rhythm, as opposed to quantization, which I relate to harmony. Figure 3 outlines the processes and dataflow as well as the supporting applications.

![Diagram](image)

**Figure 3.** Overview of analysis and composition phases.

Marking the phrases

The low quality of the recordings, Mao’s vowel-rich pronunciation, sliding vowel changes and barely articulated consonants pose problems for the transcriber. I initially worked with AudioSculpt spectral differentiating methods to generate automatic segmentation, but I did not obtain satisfactory results. By relying on my own ears I could be more precise. In addition, a “manual-aural” method forced me to listen repeatedly and in great detail to the sounds. This allowed me to integrate the musicality of Mao’s voice. This side effect of such an (admittedly tedious) method of analysis – although having the merit of gratifying the musician in the musicologist – later helped my orchestration work.⁶

---

⁶Fabien Lévy writes in [6] that he preferred doing the “transcription/quantification by hand, despite the slowness of the work, printing the chord sequence onto graph paper and then copying it again slowly. In fact, it is this manual operation that permits me to understand, to control – in short, to listen to this primary material with my inner ear, in order to render it musical subsequently.” I sympathise entirely with the approach, willingly using “slow” techniques in the transcription, while on the other hand preferring automatic rhythm quantification as far as possible. However, like Lévy, I did a fair amount of manual adjustment to rhythms and chord spellings in order to accommodate impractical notation, guided by orchestration techniques, intuition and experience.
The software Amadeus provides comfortable navigation in long soundfiles and editing of markers, labelled with different names and colours. The entire soundfile, a short extract of which is shown in Figure 4, eventually contained some 800 different markers, e.g.:

- *m* = *Mao*, the start of a vocal phrase
- *si* = *silence*, i.e. the end of a phrase
- *app, APP* = *normal applause* and *loud applause*
- *appsi, APPsi* = silence after applause
- *x* = Mao coughing or clearing his throat
- *> = accent*, syllable given particular weight

![Figure 4](image)

*Figure 4.* Excerpt from segment 4 with markers showing start and end of two short phrases, followed by clapping (during which Mao coughs).

Exporting the markers as a text file and using the data in OpenMusic is straightforward (see the upper left part of the patch in Figure 12).

We may now visualize the markers in various ways. For example, let us take a closer look at the output from segment 2, the longest stretch of sound uninterrupted by rounds of applause which lasts for approximately two and a half minutes. The middle (thicker) curve in Figure 5 plots the duration of the spoken part of phrases in this segment against their absolute onset time (as given by *m* markers). Below it is another curve showing the duration of *quasi*-silences (as given by *si* markers). The uppermost curve shows the duration of whole phrases (i.e. “enunciation followed by silence”). These curves represent the partial results from the patch in Figure 12 (*markers-analysis*, left side of the patch).

![Figure 5](image)

*Figure 5.* Durations of enunciations and silences in the second segment.

From the diagram, we see that the longest phrases (around four seconds) appear before the middle point. At the outset, silences are long – even longer than Mao’s enunciations – but towards the end, the silences are very short, for a gulp of air. As Figure 6 shows, the proportion between length of phrase and length of silence over the section is quite consistent (around 2:1), except at the end.
Interestingly, there is a tendency, in particular in the third and sixth segments, for the orator to alternate between long and short phrases, thus creating a rhythm on the phrase level. An example is shown in Figure 7.

Figure 7. The third segment (here, the second bit out of eight) shows a regular pattern of alternating long and short phrases. The thick line indicates the duration of \( m \) (Mao) bits, the dotted line below indicates the \( si \) (silence) bits. Compare this visualization with Figure 10, which shows the pitch and amplitude for the notes in the same section.

Transcribing the melody

The next step of the work was to transcribe the vocal line as a melody, treating each syllable as a note.\(^7\) Mao’s dialect, Humanese, is a language of tones, and in particular the many glissandos pose problems for the transcriber. In *TreeTorika* (with the exception of segment 3), I interpreted sliding syllables as two or three fixed notes.\(^8\)

The earlier remark about the advantages of a “manual-aural” method over an automatic segmentation method applies here as well. A Max/MSP patcher, shown in Figure 8, supported the transcription process. A syllable is visually and aurally determined by selecting a portion in the *waveform~* window. While it is looped, the user determines the pitch either by relying on the *fiddle~* estimation or by checking the sound against a note on the keyboard (if necessary adjusting it by a quartertone).

\(^7\)While ancient Chinese was essentially monosyllabic, *putonghua* (the common modern Chinese spoken language) is to a high degree semantically polysyllabic. However, the pronunciation, giving each syllable almost identical weight, makes it sound monosyllabic. As can be expected, this happens more in formal than in casual speech.

\(^8\)I am aware that this is a limitation and I intend to improve the method in future work.
Figure 8. Aural transcription assisted by a simple Max/MSP tool. For this syllable, fiddle~ estimates the pitch be a quartetone higher than B4, but my ear says it is F#4... More often than not, the fiddle~ had problems detecting the fundamental, perhaps because of the noises present in the recording. The amplitude of each note, determined automatically, is given by fiddle’s velocity detection (here 64.7).

As with the phrase markings discussed above, a low-level familiarity with the “found” material allowed me to integrate the musicality of the source material. After having worked through all the syllables in a section, the program saves onset, offset, pitch and velocity for all the notes in a text file, which is subsequently transferred to OpenMusic (as shown in the upper right half of the patch in Figure 12). Maquettes such as the one in Figure 9 were used at different stages to validate the transcription. From this point on, I refer to the voice transcription simply as the “max-notes” melody.

Figure 9. A maquette of segment 3, section 2 used to check the “Max/MSP notes”. It contains the n markers from Amadeus transcribed as downbeats, the voice transcribed as melody and a soundfile excerpt.

Interesting conclusions about Mao’s vocal delivery can be drawn from a thorough study of the data. Figure 10 shows normalized pitch and amplitude curves for a section.
They appear linked; indeed for 85% of the notes the correlation is 0.9 or greater. The succeeding much shorter section has an even higher degree of correlation. One may notice a pattern wherein a phrase typically consists of several short notes leading to a longer, louder note.

![Figure 10. Highly correlated pitch and amplitude in vocal line in segment 3, section 2. Compare with Figure 7.](image)

Regarding the phrases, segment 3 is the most regular part of the speech. It is also the segment where the sections have the highest values for correlation between pitch and amplitude. By contrast, segment 5 does not show a consistent degree of correlation. Towards the end, the correlation breaks down, as hinted in Figure 11. At this point, less than half of the notes show a correlation of 0.9 or more. This means that in segment 5, the make-up of the notes is more varied than in segment 3; i.e., there are more “high-soft” and “low-loud” syllables. In other words, one might say that the orator is “wilder” or that he uses a larger expressive palette. Listening to the recording supports this observation.

![Figure 11. Less correlated pitch and amplitude in segment 5, section 5.](image)

**Beat speed**

After having extracted phrases and notes, the next step is to “tame” the data in order to make an easily readable score, while maintaining a high degree of accuracy. The crux of the problem is finding a good pulsation speed, in relation to which quantification can be made. The demand for high accuracy has to be reconciled with the musicians’ demand

---

9The sum-square-difference between normalized curves is used as correlation measure.
for an adequate learning and use of resources.\footnote{For composers, logistics such as limited rehearsal time can be a frustrating real-world parameter to deal with. When it comes to complex rhythms, my experience is that the result is highly dependent on notation style.} Every performer employs some form of "inner clock", with subdivisions ticking away. Since a given duration sequence may be notated in any pulsation speed (with or without changing meters), the composer can act on the efficiency of the inner clock(s) set in motion.

The perception of a sequence of durations as rhythm is a psycho-motoric phenomenon. As Vijay Iyer points out, "we connect the perception of musical motion at the ecological level to human motion. This suggests that musical perception involves an understanding of bodily motion — that is, a kind of empathetic embodied cognition" [3]. A different notation signals a different rhythmic context, even if such a context is not explicitly stated. Implicitly and subtly, the musician’s inner rhythm influences the listener’s perception of the music.\footnote{Dealing with the relations between the composer’s notation and the musician’s “inner clock” is fundamental to my work. See for instance my piece for percussion Danses Condensées where the durations of individual notes in a ten-second phrase are kept unaltered over 29 repetitions, being at the same time notated within a superimposed process of a large-scale accelerando going from “non-beat” proportional notation, via slow and irregular beats to increasingly high beat speeds.}

**Figure 12.** The main analysis patch, with beat speed analysis to the left and melody transcription to the right. The bpf-lib with "candidates" is shown enlarged in Figure 14 (although for another section of the piece.)
The beat speed analysis is made in the main patch shown in Figure 12. First, the left part of the patch is used to choose an optimal beat speed and calculate a sequence of measures. During a second step, the right-hand side functions are used to quantify the “max-notes” melody according to the BPM (beats per minute) and measure markings determined in the first step. In addition, the BPF (breakpoint function) marked “melody’s pitch distribution” gives us information about the orator’s use of the ambitus of his voice.

The phrase markers are used to assist in the choice of beat speed for each section. I developed an algorithm that determines the suitability of BPMs according to accuracy and readability. Each “candidate BPM” produces an approximated sequence of measures. The fitness value according to accuracy uses the sum square distance between the “raw” and the approximated duration sequences. The algorithm for calculating readability favours simple time signatures such as 4//4 and disfavours short and oblique ones such as 7//16. The patch which calculates the suitability is shown in Figure 13.

![pm-bpm-calc](image)

**Figure 13.** Subpatch for beat speed calculation. The input is a list of “raw” onset times and the output is a list of fitness values for different beat speeds.

The resulting values are displayed as curves in a bpf-lib, such as the one shown in Figure 14. In general, the curve for accuracy climbs steadily over the range, since the time resolution increases with BPM speed. The readability curve is more salient, with peaks and valleys. The algorithm also gives the average of the first two curves once normalized; the peaks of this curve indicate optimal beat speeds.
Figure 14. Beat speed candidates for segment 3, section 2. The thick curve displays the “average”. Occasionally the context influences the choice. In this example, the peak at BPM=102 was selected instead of the highest fit at BPM=90, because it allowed a consistent quasi accelerando over the whole segment.

As one would suspect, very slow beat speeds produce high readability and low precision, while the other extreme gives low readability and high precision. In the most musically relevant region, between 70 and 140 BPM, the peaks are narrow, which can be explained by the fact that the fractional rhythms are closer to whole numbers. This reveals that choosing a beat speed off-peak, even by a little, may result in awkward rhythm notation as well as imprecision. In TreeTorika, this imprecision, i.e. the difference between timings in the recording and those in the score, is purely abstract, since it is never heard in performance. By contrast, where soundfiles and transcriptions are played in unison, e.g. in ReTreTorika and SynTorika, the notation is crucial.

Quantification

A few points should be mentioned here on the quantification method, which uses the OMquantify function. Rather than having notes discarded – and it is not easy to identify which ones – I spent time tweaking the subdivisions filter for each section. For example, I typically avoided having both quintuplets and septuplets in a section. If a good visual result cannot be reached, the option is to lower the tolerance parameter. To facilitate the work, there are two OMquantify in parallel: one giving a “high resolution” rhythm for reference and one that is tweaked for optimal readability. Figure 15 shows the intermediary output displayed in a poly object.

Figure 15. Two versions of quantification for a melody segment (segment 4), before optimal parameters had been found. By the third measure, the upper one (tweaked) is off the mark.

\footnote{This range is closest to heart beat rates and presumably of the orator’s as well.}
Extracting harmonic material

Since this article focuses on rhythm, the description of how partials and chord sequence analyses from AudioSculpt were used to create harmonic fields will be brief. To put it simply, the partial tracking approximates a sound with a large number of break-point curves while the chord sequence analysis reduces the data between specified markers into a small number of steady pitches, i.e. a chord. I find that the qualitative difference between the two analyses can be exploited expressively. Thus, the partial tracking is the more “objective” analysis, in the sense that the scope of interaction within AudioSculpt is limited, and once the data is exploited in OpenMusic, I tend to accept the result more or less “as it is”. Therefore, the method is suitable for generating large numbers of small notes, e.g. a cloud texture, such as the piano part in segments 4 and 5, shown in Figure 16.

![Figure 16](image)

**Figure 16.** Excerpt from piano part, measures 347 ff.

Partial tracking was also used for slow-moving textures with “blurred” synchronization. The left hand side of the patch in Figure 17 generated material for the strings in the second section of the first segment, and the right hand side for accordion and clarinet (joined by oboe in the brief third section).

By comparison, the chord sequence analysis can be more “subjective”, in the sense that the active specification of marker points involves decision-making already during analysis. Even when the markers are generated automatically, the detection threshold is a strong parameter which allow the composer to control the result. In *TreeTorika*, I found it useful to work with two concurrent analyses. The patch in Figure 18 shows how data were extracted from an SDIF file containing analysis data from AudioSculpt where many markers had been specified (up to one for every syllable). On the right side is another analysis file from the same section, but processed with fewer markers.

The output from the first would then be used to create a fast-moving chordal layer, while that of the other would provide material for a stable background. Examples can be found in the concluding tutti section of the second segment, shown in Figure 19, and throughout the entire final segment.
Figure 17. Patch used to filter and quantify a partial tracking from AudioSculpt. Note that quantification is entered in relation to the BPM chosen for the section.

Figure 18. A patch applying harmonic quantization and filtering to chord sequence analysis data. The function drop-246 performs a chord spacing technique typically used in big band orchestration to distribute chords containing small intervals (good-sounding on a piano) to e.g. a sax section. To the right, a “bass” line (which could be scored for piccolo flute or gongs or some other instrument) is produced with the virt-fun or the best-freq functions.
Figure 19. Excerpt from the end of segment 2, woodwinds and brass (measure 118 ff). The chords for each group of instruments were generated from two concurrent chord sequence analyses. The saxophone plays the “max-notes” transcribed melody.

Quartetone quantization

The AudioSculpt analysis produces an agglomerate of pitches, a “raw chord”, potentially containing a large number of pitches. I typically wanted chords to consist of 6 to 15 notes. To reduce the size of the raw material, some notes were filtered out if they could be considered too faint or too short, depending on the context. The patches above employ the function quartumtones. It works by allowing a certain number (between 15 and 20 percent per agglomerate) of those pitches that are furthest away from on pitch temperate intonation to be approximated to tempered quartetones, while the rest become semitones. The psychological concerns discussed above in the passage about rhythm notation are valid in relation to quartetone harmony, because the spelling of chords influences rehearsal efficiency.\(^{13}\) To my ears, chord sequences are the most convincing and understandable when quartetones are in minority. The way the quantization is made in TreeTorika associates a fidelity to the sound of the recording with demands for unproblematic ensemble intonation.

\(^{13}\)While TreeTorika’s harmonic fields are based on spectral techniques, the orchestration does not emphasize instrumental anonymity, as in purist spectral composition orchestration.
Composing with rhythm constraints

The last part of this chapter deals with one aspect of the composition, namely the development of polyrhythmic textures using the Rhythm Constraints library in OpenMusic [10] [11]. The OMRC library allows the user to create rhythm sequences, and includes functions to interface with the pmc search engine.\(^{14}\)

I will outline how the algorithm works. First, the user defines the basic building blocks, which in OMRC parlance are called “rhythm cells”. The engine attempts to build a rhythm sequence of a predetermined length by chaining rhythm cells. Second, to control the chaining of cells, the user specifies rules. If no rules are present, the output will be random. The rules are of two different kinds: Boolean rules are true/false tests, which a candidate element either passes or fails; heuristic rules are “soft” tests, which assign fitness values (sometimes called estimates) to candidates having passed the Boolean tests. When attempting to add a cell to the rhythm sequence (or “chain”) the engine first sorts all the candidates according to fitness, and tries them in order; this is called a “best-first search”. If the engine runs out of acceptable rhythm cells and still has not reached the desired sequence length, the cell tried last is dropped and the second-highest scoring candidate is tried in that position. If the chain still cannot reach the goal, the third-best is tried, and so forth. If the engine runs out of candidates, the search is directed to the second-highest scoring candidate in the preceding position. The process of advancing as far as possible and backtracking only if necessary is called a “depth-first search”. This means that the user can affect the output by determining both the material and the rules.

One significant feature of the library is the possibility to define rules to specify the interaction vertically, between layers, as well as linearly, between cells. Sandred’s work leads to the ingenious implementation of a conception of rhythms as hierarchical structures, and paves the way for highly interactive composition.

Rhythm constraint composition in TreeTorika

In movements 3 and 6, I wanted four instrumental layers (marimba, vibraphone, piano and pizzicati strings) to provide a dense texture. The preceding analysis has shown that the pulse speeds in these sections were determined to be between 80 and 126 BPM. This meant that the texture could be woven from triplets, semiquavers and quintuplet semiquavers, interspersed with longer notes. Layers would be synchronized on phrase starts (corresponding to Mao’s phrase starts). Figure 20 shows a patch using OMRC to generate the sequences.

As Sandred pointed out, the technique of search-by-constraints makes two different approaches possible: one where the user puts more information into the building blocks, and one where there is more information in the rules. In TreeTorika, the emphasis is on the former. In order to connect accompanying rhythms with the melody, the rhythm cells are extracted directly from the voice transcription. Material and rules are valid for one section at a time (plus a few measures corresponding to the applause that follows.

\(^{14}\)The original code for the pmc-engine was written by Mikael Laurson for PatchWork and ported to OpenMusic by Orjan Sandred. The principles of a problem space search guided by rules date back the 1950s with Herbert Simon’s and Allen Newell’s General Problem Solver.
Figure 20. The patch used to create polyrhythmic texture for TreeTorika, segment 6, using the rhythm constraint library. The 10/4 value to the heuristic rules sets the duration of the accelerando and ritardando in various layers.

in the recording). Rhythm cells are made up of two values, taken sequentially from the durations of the notes of the melody, in the sub-patch make-rhythms (Figure 21). For example, if the first four notes have durations a, b, c and d, the resulting set of rhythm cells will be ((a b) (b c) (c d)). The set is sorted according to the difference in duration between the two values in each cell, so that “short-long” pairs are at one end and “both-same” pairs are at the other.\textsuperscript{15} Depending on the context, a certain number of either kind of rhythm cell pair can be employed for each section.\textsuperscript{16}

The “standard rules”, shown in Figure 22, assure that rhythmic layers are of equal length, that they are different, that the cells do not repeat, and so on. They also require that all layers have cells starting on a downbeat, i.e. a phrase onset.

I found it rewarding to use several of the simpler (and weaker) tendency rules, rather than fewer (and stronger) of the true/false rules. In TreeTorika, I could work with

\textsuperscript{15}After experimenting with complicated sorting functions, I found that a simple difference is sufficient.  
\textsuperscript{16}Longer rhythm cells (with three elements or more) may have promoted a closer connection to the melody, but making sequences while upholding rules demanding synchronized downbeats and no homorhythms between three concurrent layers proved impossible.
relatively short sections (typically 15-40 seconds), and this had considerable practical advantages. Much effort goes into trial-and-error, and a small change to a rule may produce a startlingly different result, which the composer may accept or reject (using rules not yet formalised). Tendency rules were employed, for example, to build a layer with mostly “short-long” rhythmic values, or one with few quintuplets, or one privileging the use of triplets whenever available, and so on.

Tendency rules provide an interesting way to create directionality and I optimised them for each section. Figure 23 is a rule which forces the search engine to pay attention to rhythm cells of a short duration in the beginning of a section, and of a long duration towards the end of a section. The application of this rule leads to a thinning out of notes, which may globally be perceived as a ritardando in that layer.
Figure 23. A tendency rule, creating a ritardando in layer 3. (Detail of the sub-patch hr-3-rit of Figure 20.)

Orchestration

In the last step involving OpenMusic, the rhythmic layers were combined with harmonic fields extracted from the melody and orchestrated for mallet instruments using a “chord-splitting” function. The output was continuously checked against the soundfile, as in Figure 24.

Figure 24. Maquette for the rhythmic layers for segment 3, section 2. The poly contains the transcribed melody after quantization and transposition, together with two rhythm layers after harmonization. One of the layers has been subject to the “chord-splitting” function, creating two-beater alternating patterns to be played by marimba or vibraphone. The corresponding passage from the recording is at the top.

Finally, the various compositional elements were collected in a poly object and transferred to Finale, where minor editing such as chord spelling, as well as substantial and detailed orchestration work were carried out.
The melody derived from the orator’s voice is present throughout *TreeTorika*, but does not always carry a soloist function. The saxophone, leading from the start (see Figure 25) and throughout the first segment, is also featured as a soloist in the concluding sections of segment two and three, and leads the entire sixth and last segment. The second most important instruments are the side drums (gran cassa), which represent the audience’s applause, in dialogue with the orator.

In the first segment, the saxophone plays the melody *quasi cadenza* with a sparse accompaniment. The situation is turned upside-down in the second segment, when the full ensemble takes over, playing a monolithic chorale drowning out the melody. Setting out on a lighter note, the third segment gradually picks up mass and movement, before ending in aggressive saxophone solo phrases. The fourth segment lead by trombone and strings is macabre and dark. The mood continues into segment 5, where the melody is reduced to five unspecified pitches for cowbells, before tension is released in a slowly ascending line. The final segment starts on an upbeat and fuses aspects of the music of the preceding segments. A heated exchange between saxophone, ensemble and bass drums leads to a cataclysmic coda.

![Figure 25. First measures of the score, woodwinds.](image)

**Conclusion**

With *TreeTorika*, I attempted to remain analytically faithful to a given structure (in this case, a sound recording) while preserving my compositional freedom. Upon hearing the piece, listeners do not immediately associate it with rhetoric and are surprised to learn how much the composition draws on one orator, on one unique speech. I hope I reached my goal and managed to make the piece be heard as music and not as a transcription of a piece of oratory. After seven pieces based on Mao’s voice, people have asked me whom the “next dictator project” will be centred on... Stalin, perhaps? The darker side of human behaviour attracts many people, and I am no exception in this respect. More importantly, if my work can illuminate and assist in understanding the workings of *musical* power within the field of politics, my research will have come a long way.

174
In future work (involving a less than tyrannical speaker) I will concentrate on improving analytical methods in OpenMusic patches, e.g. with the inclusion of a suitability value for *metric stability* in the beat speed optimization algorithm, and with the integration of the algorithm into *OMKant* [8] [9]. My algorithm is simple, and may not be used beyond the present task, although it could eventually be integrated into *OMKant* as a user defined function. I am also pursuing Max/MSP tools to assist in the transcription of melody, rhythm and chords, using editable automatic segmentation to benefit from the advantages of real-time interactive analysis.

**Acknowledgements**

*TreeTorika* for chamber orchestra [1111-sax-1111-acc-2prec-pf-2111] (22’, 2006) was commissioned by The Ensemble Ernst with the support from Komponistrådet (Composers’ Council), Norway. At this point, it is passable to mention that the title does not refer to dendrochronology (although a working-title was the old carpenter’s adage “measure twice, cut once”) but is a contraction of the initials of those to whom the composition is dedicated, conductor Thomas Rimul and The Ensemble Ernst, and of the word “rhetoric”. At the first performance on October 14, 2006, at the Ultima Festival in Oslo, Lars Lien performed the saxophone part.

I thank Jean Bresson for his close readings of the drafts of this article, and Chan Hing-yen and Joyce Beeuwan Koh for stimulating discussions about music and rhetoric.
Bibliography


