Beethoven X: *Es könnte sein!* (It could be!)

Mark Gotham
T. U. Dortmund

Kunpeng Song
Rutgers University

Nicolai Böhlefeld
Salzburg University

Ahmed Elgammal
Rutgers University

Abstract

To mark the 250th anniversary of Beethoven's birth, Deutsche Telekom commissioned a team to develop with machine learning Beethoven's fragmentary sketches for a Tenth symphony. Here, the research part of that team discuss some of the key musical and computational decisions involved, including: making sense of Beethoven's few, ambiguous plans for this work; curating that and other musical data in machine-readable formats; and identifying suitable generation tasks, relevant training materials, and network architectures. Most of all, we consider highly topical themes for AI in the creative arts today, notably on the interaction between not only human and computer, but also living and historical artists.

1 Introduction: *Muss es sein?*

For many reasons, we have inherited an image of Beethoven as an intense, uncompromising visionary. Among the iconic stories, quotes and moments to motivate this impression is the tantalisingly cryptic 'Muss es sein? / Es muss sein!' (Must it be? It must be!) Beethoven penned on one of his last substantial compositions (the op.135 String Quartet of 1826). There has been great speculation about what exactly 'it' is that 'must be', ranging from the distinctly quotidian to the most elevated existential pondering. On the specific end, it seems most likely that Beethoven had in mind an over-due payment he was owed. At the other extreme, those specific answers do not stop commentators from seeing in Beethoven's 'muss sein' a neat encapsulation of his revolutionary spirit. Does the piece have to go this way? Must it have such challenging, radical compositional developments? Yes, it must!

However, it can be misleading to see Beethoven's compositional output as one 'muss sein' after another. Many of the works are so well known that it may seem inevitable that would turn out as they did, but this view can side-line the Beethoven who improvised, who was spontaneous, even witty. Beethoven's sketches remind us of this alternative viewpoint, revealing how Beethoven's creative process involved many options, changes of mind, and paths untravelled.[1] These kinds of questions loom large behind efforts to understand Beethoven's creative process, and especially behind efforts to 'complete' unfinished works. In this exploratory project, we sought to ask what machine learning might bring to bear on such a task. What is it now capable of in the domain of automatic music composition/completion in general, and what forms of human-machine collaborations might be possible and profitable to make best use of the expertise and resources available?[2]

Beethoven's Tenth Symphony (hereafter 'Tenth') provides a fertile testing ground for such a task. There is a small amount of surviving sketch material which is robustly associated with this work: this provides enough to give the project some significant points of departure, but it is sufficiently scant for most scholars to have determined that the work cannot be completed by traditional means. Moreover, what we have from Beethoven is not equally distributed across the prospective symphony – hence Cooper's focus on the first one or two movements in his completion (1988/rev.2012). It is

[1] This is not to say that Beethoven didn't set out with clear plans. Among recent scholarship, Posen (2022) discusses how highly distinctive elements of the final composition, although sometimes expressed in different ways, are present in some of the earliest sketches of the *Eroica* symphony. And for a literary take on all of this, see *The Unbearable Lightness of Being*, in which Milan Kundera considers Beethoven's 'es muss sein' against the idea that 'es könnte auch anders sein' (it could be otherwise).

[2] This is a vigorous area of current debate. See, for instance Briot, Hadjeres, and Pachet (2020); Ben-Tal, Harris, and Sturm (2021); and Sturm et al. (2022)'s recent special issue of TISMIR on AI and Musical Creativity.

compelling to think that with human guidance, machine learning might be able to make a lot out of a little, bringing the realisation of works like this into the realm of the possible, the 'könnte sein'.

This article discusses some of the musical and technical questions and challenges we encountered as frankly as possible. By way of a (very broad and more than slightly simplistic) overview, we seek to build computational resources that can learn from relevant music (by Beethoven and others) simply to produce a greater quantity of sketch-like material, such that the task becomes possible for the composer-completer (previously Barry Cooper working on sketches alone; here Walter Werzowa working with this computationally expanded set of materials) in a way that is not so different from the traditional scholarly completion. As we will discuss, the new process in practice is more iterative and interactive, with musical decisions to be made not only at the 'end' of the journey, but throughout. Above all, we do not aim to produce entire symphonies at the click of a button. Rather, we aim to embrace the potential for a dialogue in which human/s can guide the process, and the computer can retain and enforce some degree of objectivity by virtue of its immunity to some of the perennial challenges for the modern composer-completer: avoiding the influence both of other (e.g., later) music and of their own personal style.

Sketches obviously play an important role in this story and there will likely come a time when we can integrate all the information that we know to be important, working directly with Beethoven's entire body of sketches alongside the equivalent, completed works as part of the process for learning about compositional process. Exciting projects such as 'Beethovens Werkstatt' (click here) show that at least a form of this is eminently realisable, suggesting that the necessary, structured representation and parsing of this information will indeed be available to future versions of this project. In the meantime – and probably even with the final version of that resource – we need to start with old-fashioned, manual, archival work to find and encode 'seed' material for the machine to work with. That being the case, our story begins with some very traditional musicology: the parsing of Beethoven's notoriously ambiguous sketches.[3]

This article aims to shed some light on the whole process, starting with those sketches in a 2-part 'musicological expositions' section (§2). There follows a discussion of the computational side (in 'data-driven developments', §3) before we conclude with some brief 'compositional recapitulations' (§4). We cover details of both the musical and computational processes, while assuming very little prior knowledge of either domain, and are guided throughout by our central, recurring theme of machine learning in the context of both human-computer and current-historic co-creation.

2 Musicological Expositions

2.1 A 'Pseudo-Sketch' on *Herr Gott Dich loben wir* (the German *Te Deum*)

The rediscovery of sketches is always crucial to a story like this. We have benefited from many such revelations over the years, both for expanding our understanding of known works and for bringing back previously forgotten pieces. (And we have also been flummoxed by more than a few fascinating frauds along the way). The 1980s seem to have been the heyday for re-emergence of material associated with the Tenth. Thanks to the efforts of Barry Cooper and others, we gained a considerably enhanced sense of what musical material and structural plans Beethoven may have had in mind for this work. While we may or may not find further sketches for this work in the future, there seems to be more to unpack in what we already have. This section introduces those sketches and focuses on the emergence of a 'pseudo-sketch': a musical fragment not originally by Beethoven, but which he may feasible have intended to use in this symphony.

The earliest substantial sketch material, clearly in Beethoven's handwriting, and robustly connected to the Tenth dates from 1817/18 (Johnson, Tyson, and Winter, 1985), around the time Beethoven agreed to write a pair of symphonies (the Ninth and Tenth) for the Philharmonic Society of London. In that early sketch, Beethoven sets out his ideas not in musical notation, but rather in text. That sketch is housed at the Beethoven Haus in Bonn, who provide a digital copy of the sketch and a transcription of the text here (BSk 8/56). An version of that transcription is provided here as Table 1.

It is impossible to distinguish here between plans for the Tenth, those for the Ninth Symphony, for ideas realised in other works, and for those abandoned altogether. Even if the work designations were

[3] Table 2 provides a full list of the sketches (along with clickable links to each scan online), and this is a link to our transcriptions as hosted online in a way that supports viewing and playback.

clear, the plans are intrinsically ambiguous. For instance, Beethoven resorts three times to pairs of this idea 'oder' ('or') that. Nevertheless, there are several exciting possibilities to explore.

Firstly, Beethoven's comments allude to the possible use of voices. That is one of the defining characteristics of the Ninth, though presence in the Ninth does not exclude the possibility that they would have also featured in the Tenth. Secondly, Beethoven also raises the prospect of a fugue. That is perhaps separable from the Ninth (notwithstanding the magnificent fugal passages in that work) and more encouraging as a distinctive idea to pursue in the Tenth.[4]

Thirdly, and perhaps most cryptically, Beethoven alludes in this text to 'Bacchanalian' and 'Ecclesiatical' spirits. Does this imply a 'whole world' symphony which 'embraces everything' in the sense that Mahler would famously invoke some decades later?[5] Or was the Ninth symphony the (more) 'Bacchanalian' one, leaving the Tenth as the (more) 'Ecclesiastical' counterpart?[6] This connects with another textual passage in Beethoven's sketches for the Tenth: the 'komm, komm, nimm mich ab zur Verklärung' ('come, come, take me away to the transfiguration', sketch BSk 20; 68, f. 1). Cooper (2003) considers this other sketch as part of two possible hermeneutic readings of Beethoven's plans: a 'heaven-earth duality' (thus extending beyond even Mahler's 'whole world' concept), and an 'attempted recovery of lost innocence' (focal here being the idea of retrospection).

On the heaven-earth duality, while we have very clearly 'earthly' material associated with this work,[7] previously identified sketches do not exhibit much of a 'heavenly' or 'Ecclesiatical' side. It is perhaps significant then that this text also alludes in general to both 'Cantique Eclesiastique' and music in 'den alten Tonarten' (the old church modes), as well as in particular to *Herr Gott Dich loben wir*. This is the German version of the 'Te Deum Laudamus' text for which there is indeed a chant setting (possibly dating from St. Ambrose in the Fourth century) which Beethoven probably knew in some form.[8] This possible source material has not been discussed in previous literature on the piece,[9] and warrants unpacking as a suggestive 'pseudo-sketch'. We include this 'pseudo-sketch' in Table 2 and the online transcriptions as ID0.

Moreover, this melody is eminently consistent with the idea of 'alten Tonarten' ('old modes' or 'tonalities') to which Beethoven also alludes in the same text. Beethoven's apparent plan to use the old modes is compelling for many reasons. First, they were a subject of particular interest to Beethoven (among relatively few other composers) at this time.[10] Second, no surviving musical sketches for the Tenth are indicative of such modal material, so the identification of this theme in the above text constitutes our only 'way in' to that musical territory. Finally, it sheds new light on the idea that Beethoven's retrospective view in this work was limited to his own work, from earlier in his life, 'rather than earlier composers as is the case in other late works' (Cooper, 2003, p.21). This sketch would seem to indicate that Beethoven was engaged in both types of retrospection after all.

In summary, with no new sketches, and no musical notation at all, we can identify a new 'pseudo-sketch' which arguably constitutes an additional theme to work with, and which contains highly interesting material that apparently connects with wider plans and ideas associated with this work. But the process of 'retrieving' this is entirely non-computational. A similar process is necessary for the musical sketches. Although they are closer in format to something that we can engage with

[4] This is simultaneously enhanced and complicated by the question of what fugal subject (or subjects) to use. The textual sketch under discussion may suggest a fugue on *Herr Gott Dich loben wir* (to which we will turn in a moment), though a later sketch (HCB BSk 20/68, fol.1v), offers another fugue subject in B-flat major, also apparently intended for this work. See ID4 of the source reference list (Table 2) and online transcriptions.

[5] 'The symphony must be like the world. It must embrace everything.' Mahler to Sibelius, Helsinki, 1907 (Blaukopf and Blaukopf, trans Baker 1991, p. 204).

[6] '(more)' because there is a religiosity of sorts in the Ninth via the text (Schiller's *An die Freude*), which includes deistic comments relevant to the following discussion of the 'Te Deum'.

[7] Most notable, perhaps, is a rousing theme (HCB BSk 20, fol.1v), that Beethoven was ultimately to realise as the 'Gratulationsmenuett' (WoO3), complete with trumpet and drum fanfare. Click here for a score on IMSLP and see ID5 of our summative Table 2 (references) and transcriptions online.

[8] Naturally, there are many versions of this tune specifically linked to the Te Deum in not only Catholic but also Lutheran contexts, including in several Bach cantatas (BWVs 16/1, 119/9, 120/6, 190/1, 190/2, and 190a/2).

[9] Indeed, this part of the text is not included in the Beethoven Haus' online transcription/translation; it was confirmed for us by Prof. Christine Siegert who we thank for this and other assistance.

[10] These modes feature in important late works such as the 'Heiliger Dankgesang' movement of the op.132 string quartet (1825), which Beethoven explicitly labels as being 'in der Lydischen Tonart' (in the Lydian mode).

computationally, they are still initial ideas, hastily written down by a composer not famed for the neatness of his handwriting. We turn to that subject now.

2.2 The Scherzo-Trio, *ossia* 'A single note is worth a thousand words'

A musical 'motive' refers to a kind of prototypical theme, simpler than any specific manifestation, which lies behind the 'real' music. The idea is common in the analysis of existing music, and it can be instructive in parsing the initial ideas for a work 'to come', particularly for composers like Beethoven who would frequently sketch out musical ideas initially in this way.

Cooper (2003) argues persuasively for a motivic reading of the Tenth in terms of a progression from an 'open' melodic configurations on scale degrees $\hat{3}$-$\hat{2}$-$\hat{5}$,[11] to closure on either $\hat{3}$-$\hat{5}$-$\hat{8}$ or that most quintessentially conclusive of musical melodies, $\hat{3}$-$\hat{2}$-$\hat{1}$.[12] Cooper sees the 'open' version of the motive as the primary compositional 'problem' to be 'solved' in the symphony, and links this to a wider metaphorical reading as briefly mentioned above. He considers the 'closed' form ($\hat{3}$-$\hat{2}$-$\hat{1}$) to be 'already achieved in a first-movement sketch' (the main remit of his attention, as discussed) but also notes the significance of that $\hat{3}$-$\hat{2}$-$\hat{1}$ appearing 'confidently and conspicuously right at the head of' the material Beethoven had in mind for a possible finale theme (the 'Gratulationsmenuett', see ID5 in Table 2 and the online transcriptions).[13] Viewed one way, then, we have a symphony progressing from an opening $\hat{3}$-$\hat{2}$-$\hat{5}$, to a finale strongly based on a repeated emphasis of the closing $\hat{3}$-$\hat{2}$-$\hat{1}$.

One of the earliest, longest, and clearest pages of musical sketch material associated with the Tenth symphony sets out the primary melodic material apparently intended for a Scherzo-Trio pair in c minor and C major respectively.[14] The Trio is explicitly identified with the word 'Tr[i]o' about halfway down the page (probably in Beethoven hand) and can be un-controversially transcribed with a melody centred on scalic motion from C-G, then G-C, and C-F. If there's a motive to be had here, it's the melodic outlining the interval of a perfect fourth. We go from scale degree $\hat{1}$ down to $\hat{5}$ (C-G), from $\hat{5}$ back up to $\hat{1}$ (G-C), and then from $\hat{1}$ to $\hat{4}$ (C-F).

The 'Scherzo' which appears above that Trio it is not so named by Beethoven (that text on the page is not in Beethoven's, but rather in Schindler's hand). Then again it is explicitly set out in 3/4 time, at a 'Presto' tempo, and in partnership with this explicitly labelled 'Trio', so Schindler's assessment seems relatively uncontroversial.

More typically for Beethoven, the process of transcribing this Scherzo is no straightforward, neutral matter. As was the case for the titles, so too in the musical notation is the Trio relatively clear, but the Scherzo is more cryptic. Indeed, this Scherzo is replete with note stems for which the note head is either ambiguous or missing, and which thus leave much to be deduced in terms of the exact pitches. Among the many possibilities here, and in contrast to transcriptions by some others, we opt for a reading of the opening ascent as G-C, B-E♭, D-G (this giving a fourth-up-step-down pattern).

This is compatible with the sketch, though if the only goal were to transcribe Beethoven's handwriting as faithfully as possible from the ink on the page, without further deduction, we would likely settle on an alternative reading, probably including an E♭ in place of the D in b.5. The goal here has been to decipher what Beethoven might have had in mind, or at least what might 'work' with the wider processes at work. The sketch is the focus, but there's other, surrounding musical evidence to be had both within the piece (in the motive discussed above), and elsewhere in Beethoven's oeuvre. As the 2/4 theme's parallel with the *Pathétique* sonata indicates,[15] and as many commentators have noted, Beethoven does seem to have been interested in re-using the same material across works at this time.

We begin within the piece – indeed, within this single page of sketch material and the potentially motivic outlining of fourths. As discussed, the Trio proceeds by filling in fourths from $\hat{1}$-$\hat{5}$, $\hat{5}$-$\hat{1}$, and $\hat{1}$-$\hat{4}$, successively; this already connects positively to this reading of the Scherzo, which centres on successive leaps of a fourth, arranged in a 'sequence'.

[11] See especially the unmistakable similarity between ID1 in Table 2 and the online transcriptions) and the slow movement of Beethoven's early *Pathétique* sonata, op.13. This is key to the question of 'retrospection'.

[12] Click here for an introduction to scale degrees in the *Open Music Theory* textbook (hereafter 'OMT').

[13] For the status as a possible finale, see also Brandenburg (1984), p.112.

[14] Click here to view the sketch online and see ID11 of the summative Table 2 and online transcriptions.

[15] See sketch ID1 of Table 2 and the online transcriptions, as discussed.

Zooming out, there is a strong connection here to the thematic basis of the opening of the symphony's 2/4 Andante theme (the *Pathétique* figure, discussed above) and its characteristic $\hat{3}$-$\hat{2}$-$\hat{5}$ motive. This $\hat{3}$-$\hat{2}$-$\hat{5}$ shape is evident on those specific scale degrees in bars 4–6 of this Scherzo and is also naturally present throughout that opening, which is built on a sequence of that same figure several times. This Scherzo theme could thus be thought of as emerging backwards out of the 2/4 theme via a little sequential *Fortspinning* (composing out), to use the usual musicological term, or perhaps better in this case, *Rückspinnung* (composing backwards).

In short, the dramatic Scherzo seems still to adopt the opening motiv (the one requiring 'resolution'), while the paired Trio (ever an opportunity for reducing tension) turns those fourths into a conjunct melodic line that leads more towards the 'resolved' finality of the $\hat{3}$-$\hat{2}$-$\hat{1}$ finale. In this reading, the Scherzo-Trio pair provides a pivotal centre for the multi-movement work and its motivic process.

Moreover, the figure is also highly redolent of certain other, highly significant works of Beethoven's late period. This pattern seems to recur in every possible variant of this 'fourth and step' motivic configuration. Witness, for instance, the highly relevant case of the late-period piano sonata (op.110). The last movement features two fugal passages based a common fugal subject which is presented 'straight' at the first time of asking (from b.27) and then in inversion (from b.137). Motivic unity has been a long-standing focus of Beethoven scholarship and many scholars have pointed specifically to the op.110 sonata as an example of unity across a multi-movement work, including Cooper,[16] in discussion of those connections in the Tenth. Yet this is the first time that the Scherzo has found a place in this discussion, both within the potential motivic trajectory of the Tenth symphony, and as part of this wider nexus of significant late works. And this reading depends heavily on the proposed transcription, with the tessellated fourth-step pattern. In summary, it all comes back to that first process of transcribing, even if it takes a thousand words to make the case for a single note.

3 Data-driven developments

All these deductions and decisions still only get us to the very beginning of a potential work: a small collection of themes as starting points to work with; 'seeds' to prompt further generations. Moreover, while computational methods such as cluster analysis could provide some complementary insights into the kinds of motivic connections made above, they have yet to have any significant, decisive role.

Only when we have made the many and varied decisions about textual and musical starting points can we begin to incorporate computational methods for generating new material. Broadly, this involves creating machine learning architectures (note: plural) that will learn from corpora (again, plural) of existing, 'relevant' material to provide potential solutions to well-formed musical questions. Among the ways of approaching this, we separate the tasks and the architectures into dedicated networks for:

- Continuation: extending a fragmentary melody into a longer one;
- Harmonisation: adding additional (e.g., accompanimental) parts to a melody;
- Transition: providing 'bridging' material to connect the preceding and following music;
- Orchestration: distributing the core material across the full orchestral forces.

And just as we have different architectures for separate tasks, so too do we need dedicated corpora for each stylistic categories, separating 'march' from 'minuet', for instance. There can be a shared, underlying corpus applicable to all with a large range of broadly relevant material (i.e., Western tonal music), but individual, context-specific corpora provide significant fine-tuning for each stylistic task at hand. It is important to note that deciding what counts as 'relevant' is once again a musical, subjective, context-specific judgement-call not so different from those discussed above.

Overall, we elected to restrict ourselves to music either by Beethoven, or that he could possibly have known, and to err on the side of 'more is more' in an attempt to provide enough data for machine learning to be effective. All the same, we re-emphasise that at least for the narrower, topic-specific corpora for 'Scherzi', 'Fugues' and so on, although we can be be guided by clear-cut metadata (like titles) and musical features (like time signatures), ultimately many judgement calls are required.

[16] Cooper (2003), p.8 after Dahlhaus and Kinderman.

3.1 Continuation

One of the most basic tasks we might need to accomplish is the extension of a short, fragmentary sketch into a longer melody. Our 'continuation' model is designed for this task, taking its inspiration and broad structure from previous successes in the related field of natural language processing in text.

Natural language models are typically designed to take a sequence of words and predict those that follow, with the goal of generating realistic sentences. Transformer-based models such as Vaswani et al. (2017) have proven particularly powerful in this domain, using a 'self-attention' mechanism to learn correlation among members of the input sequence. Dai et al. (2019)'s 'Transformer-XL' enhances this further with a memory mechanism which stores vectors of previous segments for reuse in processing future ones. This seems to enable better performance in capturing a longer range of dependency among thousands of events.

Naturally we are not the first team to adapt this approach to a musical task. We acknowledge and benefit in particular from: Oore et al. (2020)'s sparse event-based representation for converting musical data from (or via) the MIDI standard to the Transformer-XL's token sequence format;[17] Huang et al. (2018)'s algorithm to reduce the intermediate memory requirement; and Wu, Wang, and Lei (2020)'s further modification of the Transformer-XL approach to achieve longer generations.

As with many machine learning contexts, we seek to build a system that performs well partly by imitating something of the human style of learning. There are countless kinds of relationships between tones in musical sequences and likewise many ways to learn the nature of those structures. While the details of how this is accomplished may differ significantly, and while it is often hard-or-impossible to interpret the exact nature of what 'rules' the system has 'learned',[18] we can at least recognise in this an in-principle priority of the human approaches to teaching melody that have been taught for centuries, for instance in the 'reduction' of melodies into simpler prototypes. In both the musical and machine learning contexts, 'attention' is a watchword: one musical priority of the model must be to identify and attend to 'the important moments' and Vaswani et al. (2017) proposes a method for 'multi-head-self-attention' in which each of several attention heads may focus on different musical features.

3.2 Stylised continuation

The basic continuation model described above thus extends the musical seed within the style and/or theme of the seed itself. This works iteratively, extending the original sequence one token at a time. As such, the influence of the initial seed (first few tokens) becomes attenuated over time. This is a limitation wherever the goal is to create long sequences of music meaningfully related to the original seed (i.e., theme). In order to overcome this, we developed a stylised continuation model to enforce an explicitly selected choice of theme (theoretically any, though typically the original seed).

Again, we benefit from previous work in this area of transferring musical style from a chosen seed/theme to the output sequences, notably from: Choi et al. (2020) who encode musical style using a transformer autoencoder,[19] and Vaswani et al. (2017) who remove the necessity of temporal compression. In this latter, the encoder takes in a musical sequence alongside a set of empty vectors. It then outputs 'bottleneck' vectors which contain compressed information from the input sequence. In the training phase, the decoder takes in a musical segment and predicts the next token based on both the previous tokens in that sequence and the 'bottleneck' vector representation from the encoder. In the evaluation phase, the decoder extends a seed by taking the musical seed and iteratively predicting subsequent musical tokens. The extended musical sequence is thus influenced by the 'bottleneck style vectors' and their attempt to capture aspects of the style in question. For this project, we built architectures using multiple bottlenecks, with the aim of storing a wider range of stylistic information (in the form of sequences) more effectively.

[17] Any such route via MIDI means a loss of musical information including pitch spelling. Because recent work in machine learning for harmonic analysis has shown that pitch spelling can significantly improve results (Micchi, Gotham, and Giraud, 2020; Nápoles López, Gotham, and Fujinaga, 2021), future work on automatic composition may also do well to consider alternatives.

[18] We look forward to the forthcoming special issue on 'Explaining music with AI: Advancing the scientific understanding of music through computation' (*Music & Science*, ed. Meredith, Volk and Collins)

[19] In our model both the encoder and decoder are composed of a stack of Transformer-Decoder layers.

3.3 Additional Lines: Harmony and Counterpoint

Beethoven's sketches for this work consist mainly of single-line melodies and similarly sparse textures. As such, we also need a model that will take in these sparse inputs and provide accompanying material. This can take many forms, distinguishing minimally between the addition of contrapuntal voices and of block chord harmonies. Counterpoint can be learned through training on every permutation pair of parts in relevant music. For instance, in a 3-voice fugue, we can learn from how part-pairs 1–2, 2–3, and 1–3 interact as three separate entries. Particularly useful here are datasets setting out these musical parts on separate lines.[20] As in all cases, we can segment whole works into shorter contiguous segments, for instance dividing by internal rests. This can be constrained further, for instance, to handle the special role of the lowest voice separately.

For a more general form of 'harmonisation' (defined slightly broader than in opposition to counterpoint), the task is a less specialised form of accompanying the sparse single-line with additional material. We consider it operationally useful here to think of the task of harmonisation by analogy to machine translation. Translation involves taking in one language and outputting the corresponding idea in another. Harmonisation be conceptualised in a similar way: taking in a melody and outputting a corresponding structure in harmonic terms.

For this 'harmonisation' task, our model contains stacks of transformer encoders and decoders in the structure of a standard NLP machine translation model. The encoder takes in the melody line and outputs memory vectors. The decoder then takes in the full music and predicts the next musical event based on both the previous one and the memory stored in the encoder.

While we can often apply machine learning architectures initially developed for language and other fields directly to music (assuming suitably comparable encoding of the data), there are sometimes significant differences. For instance, in translating text, it is routine to change the word order to achieve a syntactically sensible output in the destination language. In the case of a comparable 'translation' between melody and harmony, such a reordering would be inappropriate. As such an additional control is needed here to enforce the preservation of the music in its original order.

3.4 Transition

The transition model serves to provide a bridge between two segments of generated music. This may be to connect two separate parts together, or to replace a short passage within some generated music that is felt to be inadequate for whatever reason. Here the 'BERT' masked language model (Devlin et al., 2019) provides a suitable tool. BERT (Bidirectional Encoder Representations from Transformers) is typically trained by randomly concealing 15% of the events (e.g., words) and attempting to predict what ought to be there based on the remaining 85%.

Once again, this can be applied to musical tasks as long as we are sensitive to the special circumstances in the nature of the domain. Generating musical transitions is different and challenging in two ways. First, instead of masking out 15% randomly, the masked passage consists of a contiguous block of consecutive events (such as a single unsatisfactory bar). Second, instead of predicting the masked-out words all at once, it predicts events one by one, iteratively. This requires a different attention map masking mechanism. The attention mask of our transition model is an nxn square matrix, where n is the number of musical events in the training sequence. A contiguous span of cells is initially filled with -inf (mapping to 0 after the softmax function) and the associated information for this range is thus masked-out and inaccessible.

This is one of many contexts in which gaps can emerge in how we count musical time. The NLP-based default is to count in terms of *events* (e.g., words). While this can fit the musical domain (i.e., notes), we will often wish to use other time measurements such as *bar* (US: measures) of a fixed duration (in symbolic units), but with an an unknown number of events. Among the possible solutions, when training a model for filling a fixed length in bars, we calculated the average number of events in the corpus for the given number of bars.

[20] While the setting of one instrumental stave per contrapuntal voice is routine for much chamber music, contrapuntal music for the keyboard involves more voices than staves. As such, multiple voices are combined on staves and often also cross between them. Aljanaki et al. (2021)'s 'MCMA' is a notable dataset here in that it re-notates keyboard music specifically in the open stave format for this purpose.

3.5 Orchestration

While composers approach the task of writing for orchestra in different ways, many chose to write a 'short score' first, containing most of the material, and then to transfer this to the full orchestra as a second step. For computational purposes, we can operationalise much of this simply in terms of assigning each note in the 'short score' (here: generated material) to one or more instruments in the orchestral score. Specifically, we can continue to use an existing representation of notes, adding to that existing information (pitch, timing, ...) their instrument assignment.

We further operationalise this task in terms of 'masking', using a similar BERT model to that used for the transition task. Where the transition model masks *all information* for a short span of events, the orchestration model masks only the *orchestration information for all events*. Once again, masked cells are initialised to -inf., so that this information is blocked.

In training this model, we need we matching pairs of 'short' and 'full' orchestral scores. Somewhat relevant material for this exists in the form of 'piano reductions' of symphonies, however, these reductions are of limited use for our purposes because they do not correspond directly enough to the generated material. For instance, they introduce the constraint of suitability for performance on the piano specifically (which is irrelevant for our purposes).[21]

In creating training material, we prefer instead to reduce orchestral scores to a kind of 'short score' using deterministic algorithms. In so doing, we can control the approximate density of the output, ensuring that it corresponds to the equivalent density of the newly generated material and thus a relevant context for learning. Once again, this involves many musical decisions about which elements are essential, and which are expendable.

For generation, once again, while we can make use of pre-existing architectures to some extent, there are also domain-specific differences. First, while there is a broadly consistent set of core instruments in the classical orchestra, the actual set of those used varies slightly from movement to movement. For instance, Beethoven uses trombones in only some movements of his Fifth, Sixth, and Ninth symphonies, and not at all in the others. Second, the pitch range of notes varies for different instruments. We explicitly encoded those range constraints for each instrument as a precaution, despite the likelihood of the model learning this by itself. Finally, as with most of our musical contexts, orchestration is highly sequential: if a melody began on a particular instrument, it is likely to remain there until a structural break point like a cadence. As such, and unlike a typical BERT, our orchestration model is trained to predict the instrument events iteratively, one by one.

4 Compositional recapitulations

At the end of this process, we have deduced the musical material to use, and generated continuations and expansions of various kinds. In some senses, the composition proper begins at this point, with human decisions over which generated music to use and which tasks require re-generating.

The opportunity to 'double back' is perhaps particularly useful and interesting. In the traditional form of this task, a composer leaves a set of sketches that constitutes the totality of available material (short of discovering more). Here, while we begin with those initial sketches, we can generate extensions, choose from among them, and then repeat the process, taking the extension itself as a kind of (larger) sketch from which to generate further extensions. In short, there is a back-and-forth dialogue between human and computer, with the human leading the process overall, while keeping themselves bounded by the material generated by a system that only knows the 'relevant' repertoire.

Our guiding principle was that the final work would consist entirely of generated material, but that our composer would have no limits on how atomically to treat that material (e.g., one bar at a time), nor any constraints on how often to recombine and request new generations for a particular context. In a way, this is all highly analogous to the task of human-only composition. As the composer, scholar, and noted completer of unfinished work Deryck Cooke puts it: 'what we call inspiration must be an unconscious creative re-shaping of already existing materials in the tradition' (Cooke, 1959, p.171). We take inspiration from this attitude and from the possibility of developing methods for AI-assisted composition that enable, enhance and extend human creativity and exploration.

[21] Note that the 'short score' of material intended ultimately for orchestra is different from a piano reduction. For more on this distinction, click here for a chapter on 'Transcription from piano' in *OMT*.

Acknowledgments and Disclosure of Funding

Thanks to Deutsche Telekom for initiating and funding this project, and to all involved, notably:

- Matthias Röder, project lead/manager
- Walter Werzowa, composer
- Robert Levin and Christine Siegert, advisors
- The many student assistants who helped build corpora.

Thanks also to all anonymous readers (through AIMC and otherwise) for their comments on earlier versions of this text.

References

Aljanaki, A.; Kalonaris, S.; Micchi, G.; and Nichols, E. 2021. MCMA: A symbolic multitrack contrapuntal music archive. *Empirical Musicology Review* 16(1):99–105.

Ben-Tal, O.; Harris, M. T.; and Sturm, B. L. 2021. How Music AI Is Useful: Engagements with Composers, Performers and Audiences. *Leonardo* 54(5):510–516.

Blaukopf, K., and Blaukopf, H. trans. Baker 1991. *Mahler: His Life, Work, and World*. Thames & Hudson.

Brandenburg, S. 1984. Die Skizzen zur Neunten Symphonie. In Goldschmidt, H., ed., *Zu Beethoven*, volume 2. Berlin: Verlag Neue Musik. 88–129.

Briot, J.-P.; Hadjeres, G.; and Pachet, F.-D. 2020. *Deep Learning Techniques for Music Generation*. Computational Synthesis and Creative Systems Series. Springer.

Choi, K.; Hawthorne, C.; Simon, I.; Dinculescu, M.; and Engel, J. 2020. Encoding musical style with transformer autoencoders. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20, 1899–1908. JMLR.org.

Cooke, D. 1959. *The Language of Music*. Oxford paperbacks. Oxford University Press.

Cooper, B., and Winter, R. S. 1992. Correspondence: Beethoven's Tenth Symphony. *Journal of the Royal Musical Association* 117(2):324–330.

Cooper, B. 1985. Newly identified sketches for Beethoven's Tenth Symphony. *Music & Letters* 66(1):9–18.

Cooper, B. 2003. Subthematicism and metaphor in Beethoven's Tenth Symphony. *Ad Parnassum* 1.

Dai, Z.; Yang, Z.; Yang, Y.; Carbonell, J.; Le, Q.; and Salakhutdinov, R. 2019. Transformer-XL: Attentive Language Models beyond a Fixed-Length Context. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2978–2988. Florence, Italy: Association for Computational Linguistics.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* 4171–4186.

Huang, C.-Z. A.; Vaswani, A.; Uszkoreit, J.; Simon, I.; Hawthorne, C.; Shazeer, N.; Dai, A. M.; Hoffman, M. D.; Dinculescu, M.; and Eck, D. 2018. Music transformer: Generating music with long-term structure. *International Conference on Learning Representations*.

Johnson, D.; Tyson, A.; and Winter, R. 1985. *The Beethoven Sketchbooks: History, Reconstruction, Inventory*. California studies in 19th century. University of California Press.

Micchi, G.; Gotham, M.; and Giraud, M. 2020. Not all roads lead to Rome: Pitch representation and model architecture for automatic harmonic analysis. *Transactions of the International Society for Music Information Retrieval* 3(1):42–54.

Nápoles López, N.; Gotham, M.; and Fujinaga, I. 2021. AugmentedNet: A Roman numeral analysis network with synthetic training examples and additional tonal tasks. In *International Society for Music Information Retrieval Conference (ISMIR)*.

Oore, S.; Simon, I.; Dieleman, S.; Eck, D.; and Simonyan, K. 2020. This time with feeling: Learning expressive musical performance. *Neural Computing and Applications* 32(4):955–967.

Posen, T. 2022. From "Radical Blunders" to Compositional Solutions: A Form-Functional Perspective on Beethoven's Early Eroica Continuity-Sketches. *The Beethoven Journal* 35.

Sturm, B. L.; Uitdenbogerd, A. L.; Koops, H. V.; and Huang, A. 2022. Editorial for TISMIR special collection: AI and musical creativity. *Transactions of the International Society for Music Information Retrieval* 5(1).

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Wu, X.; Wang, C.; and Lei, Q. 2020. Transformer-XL based music generation with multiple sequences of time-valued notes. *ArXiv* abs/2007.07244.

Transcription of Beethoven's comments in BSk 8/56

<table>
<tr><td>[links:]

Adagio Cantique
From[m]er Gesang
in einer Sinfonie
in den alten Tonarten.</td><td>[Mitte:]

entweder
für sich allein
oder als Einleitung
in eine Fuge
“Herr Gott Dich loben wir
alleluja”</td><td>[rechts:]

vieleicht auf diese weise die
ganze 2te Sinfonie charakteri=
sirt wo alsdenn im lezten</td></tr>
<tr><td colspan="3">[über die ganze Zeile:]

Stück oder schon im adagio
die Singstimmen eintreten
die orchester Violinen etc werden beym lezten Stück verzehnfacht.
Oder das adagio wird auf gewiße weise im lezten
Stücke widerholt wobey alsdann erst die Singstim[m]en
nach u nach eintreten – im adagio text
griechischer Mithos Cantique Eclesiastique
im Allegro Feyer des Bachus.</td></tr>
</table>

Table 1: A reproduction of Beethoven's comments in BSk 8/56 as discussed in the main text and set out to broadly resemble the original.

(P.T.O. for primary sources and Table 2.)

Primary sources

The primary source material for this project is, of course, Beethoven's own sketches. The relevant parts are hosted in three archives:

1. Beethoven-Haus Bonn, Sammlung H. C. Bodmer:
   - HCB BSk 20/68 (SBH 673), fol. 1r–2v
   - HCB Mh 86 (SBH 645), fol. 1r–1v
   - HCB BSk 8/56 (SBH 652, SV 161)
2. Staatsbibliothek Preußischer Kulturbesitz Berlin, Musikabteilung:
   - aut. 9/1, fol. 1v–5r
   - Artaria 201, p.124-5
3. Vienna, Gesellschaft der Musikfreunde:
   - Archive A50, p.12

Table 2 below provides more specific details of the sources, complete with folio reference, approximate year, URL for accessing the digitised sketch image online, and sources for a corresponding transcription in existing literature.

The exceptions are Vienna A50 for which there is not currently a digitised version online and HCB BSk 8/56 for which there is no transcription (it has not previously identified as a musical source as discussed in the main text).

ID	Source: full reference	Specific Folio and Clickable URL	Year	Transcription in the literature
0	HCB BSk 8/56	BSk 8/56	1817/18	N/A
1	Artaria 201, p.124–5	p.125	1822	Brandenburg (1984), p.111
2	Artaria 201, p.124–5	p.125	1822	Brandenburg (1984), p.111
3	HCB BSk 20/68, fol.1r–2v	fol.1v	1822	Brandenburg (1984), p.111
4	HCB BSk 20/68, fol.1r–2v	fol.1v	1822	Brandenburg (1984), p.111
5	HCB BSk 20/68, fol.1r–2v	fol.1v	1822	Brandenburg (1984), p.112
6	HCB BSk 20/68 fol.1r–2v	fol.2v	1822	Cooper and Winter (1992), p.325, ex.1
7	HCB BSk 20/68 fol.1r–2v	fol.1r/10	1822	Cooper and Winter (1992), p.326, ex.2
8	HCB BSk 20/68 fol.1r–2v	fol.1r/10	1822	Cooper and Winter (1992), p.326, ex.2 (ossia)
9	Vienna, A50, p.12	p.12 (no URL)	1824	Cooper and Winter (1992), p.327, ex.4
10	HCB Mh 86, fol.1r–1v	fol.1r–1v	1812	Cooper and Winter (1992), p.327, ex.5
11	aut. 9/1, fol.1v–5r	fol.1v/1–1v/4	1825	Cooper (1985), p.12, ex.1
12	aut. 9/1, fol.1v–5r	fol.1v/5–2r/5–6	1825	Cooper (1985), p.12, ex.2
13	aut. 9/1, fol.1v–5r	fol.3r/1–3	1825	Cooper (1985), p.14, ex.4
14	aut. 9/1, fol.1v–5r	fol.3r/5	1825	Cooper (1985), p.14, ex.5
15	aut. 9/1, fol.1v–5r	fol.3r/4	1825	Cooper (1985), p.14, ex.6
16	aut. 9/1, fol.1v–5r	fol.3v/1–5	1825	Cooper (1985), p.15, ex.7
17	aut. 9/1, fol.1v–5r	fol.4r/2	1825	Cooper (1985), p.15, ex.8
18	aut. 9/1, fol.1v–5r	fol.4v/1–2	1825	Cooper (1985), p.16, ex.9
19	aut. 9/1, fol.1v–5r	fol.4v/3–4	1825	Cooper (1985), p.16, ex.10
20	aut. 9/1, fol.1v–5r	fol.4v/6	1825	Cooper (1985), p.16, ex.11
21	aut. 9/1, fol.1v–5r	fol.5r/1	1825	Cooper (1985), p.17, ex.12
22	aut. 9/1, fol.1v–5r	fol.5r/2–3	1825	Cooper (1985), p.17, ex.13

Table 2: Details of the surviving sketch material for the Tenth, complete with folio reference, approximate year, URL for accessing the digitised sketch image online, and corresponding transcription in existing literature on this topic. Our transcriptions are provided in an online supplement with the same ID numbers at this address (click here). This also supports playback.