CHAPTER 10
Cognitive Musicology and Artificial Intelligence: Harmonic Analysis, Learning and Generation[1]

Emilios Cambouropoulos and Maximos Kaliakatsos-Papakostas

10.1 Introduction

A listener is able to discern diverse aspects of music when exposed to musical stimuli: from elementary features of the musical surface (e.g., a discrete note or a chord or a certain timbre), to salient musical patterns (e.g., motives, themes, cadences), and, even, high-level composer or stylistic features. A listener may find, for instance, a particular harmonic progression intriguing, inducing new exciting responses, even though the listener is not able to identify constituent chords and scale degrees, or a melody might sound emotionally moving to a listener, even if (s)he is not able to name the individual notes or intervals. Harmony, melody, rhythm, texture and timbre (among others) are aspects of music that a listener is able to appreciate, encode and remember, despite not having explicit access to underlying components.

Through the centuries, music theorists, analysts, philosophers have attempted to describe (via introspection) and to formalise, core musical concepts and processes, such as scales, chord types, harmonic functions, tonality, rhythmic structures, types of texture, form and so on. In more recent years, computational methodology, and more specifically Artificial Intelligence (AI) has offered new means of precision and formalisation, enabling the development of models that emulate musical intelligent behaviors. This way, musical theories and hypotheses, drawing, not only on traditional music knowledge but, also, on research in music cognition, linguistics, semiotics, logic reasoning, neuroscience and so on, have given rise to actual musical analytic / compositional / performance computer programs that may be tested in a more systematic manner and may give rise to useful computational applications.

Nowadays, Artificial Neural Network (ANN) architectures, and more specifically Deep Learning methods, appear in the minds of many researchers to have superseded the Good Old-Fashioned Artificial Intelligence (GOFAI) paradigm; for many younger AI researchers, Artificial Intelligence is Deep Learning. The hypothesis is that given sufficient amounts of data represented appropriately, and adequate deep learning algorithms, any musical intelligent behavior can be emulated successfully. So, what is the point of developing sophisticated AI programs that employ *ad hoc* knowledge-engineering approaches (drawing on music theory or music cognition) when a generic ANN approach may be at least equally effective (not to mention that it is more flexible and adaptive)?

[1] Cambouropoulos E., Kaliakatsos-Papakostas M. (2021) Cognitive Musicology and Artificial Intelligence: Harmonic Analysis, Learning, and Generation. In: Miranda E.R. (eds) Handbook of Artificial Intelligence for Music. Springer, Cham. https://doi.org/10.1007/978-3-030-72116-9_10

In this chapter we discuss aspects of Cognitive Musicology with a view to presenting reasons why it is relevant in the context of current developments in the field of AI. Otto Laske (1988) states that "cognitive musicology has as its goal the modeling of musical intelligence in its many forms." (p.44). According to Laske, "computer programs serve to substantiate hypotheses regarding musical knowledge and, second, they are the medium for designing structured task environments (such as programs for interactive composition). While it is not a prerequisite for building intelligent systems to have a fully-fledged theory of the activity one wants to support, it is certainly more effective to design such systems on as much theory as one can harness." (p.45) We assert that insights drawn from Cognitive Psychology and, also, music theory, play an important role in building musical models (combining symbolic AI with statistical learning) that can serve both as a means to broaden our understanding of music *per se* and, also, to create robust sophisticated systems for music analysis, composition and performance. Computational modeling, as a means to test musical hypotheses, has enriched musical knowledge in the domain of musical analysis (Meredith, 2016) and music cognition (Temperley, 2012), enabling, at the same time, the development of useful musical systems and applications.

In the next section, we discuss briefly general issues regarding advantages and disadvantages of symbolic vs deep learning methodologies. Then, we present two case studies that show the effectiveness of classical symbolic AI in music modeling (coupled with simple statistical learning techniques): firstly, we examine modeling of melodic harmonisation showing strengths and weaknesses of both the standard AI and deep learning methodologies, and, secondly, we present a creative melodic harmonisation system based on Conceptual Blending Theory (Turner and Fauconnier, 2003) that operates on a high reasoning level allowing sophisticated combination of abstract chord features with a view to generating novel harmonic spaces. In both of these cases we argue that the classical symbolic approach to musical intelligence drawing on music cognition research, coupled with simple statistical learning techniques, provides a reasonable way to address complex phenomena of musical listening, performance and creativity effectively.

10.2 Classical Artificial Intelligence vs. Deep Learning

Before attempting to give reasons for pursuing cognitive musicology research following the more traditional symbolic AI approach (or at least a hybrid AI & statistical learning approach), a brief discussion on core cognitive processes of acquiring knowledge is due. Such processes involve, among other things, abstraction, categorisation, inference, use of prior knowledge, encoding and transmission of knowledge at a high-level symbolic level.

Information abstraction or compression appears to be a universal innate mechanism of cognition and consciousness. It is essential not only for humans in their everyday interaction with the surrounding environment and with other humans, but for other non-human animals as well. Feinberg and Mallatt (2013) maintain that the reduction of information received from the visual sensory system to a more abstract representation of the visible world, facilitates the involvement of memory in decision making; this applies to the auditory and other sensory systems as well. For example, a predator with the ability to form abstractions of the perceived world requires significantly less memory capacity to remember the existence of prey at a specific location while having no visual contact with it. This allows the predator to develop stealth hunting techniques that do not require constant visual contact with the prey, giving the predator an advantage in species evolution. While hunting techniques are irrelevant to music cognition, the aforementioned example shows that abstraction in representing elements in the perceived world, and the advantages in memory requirements that this abstraction yields, is part of a fundamental mechanism in the evolution of species.

Information abstraction, or more formally put, information compression, is a cognitive mechanism constantly at play at all levels of music understanding. Not only the extraction of the musical surface *per se* (i.e., the actual note pitches / durations / beats) from the actual audio signal involves a very sophisticated information abstraction process, but also the extraction of higher-level meaningful structures from the musical surface. This abstraction mechanism allows listeners to move beyond the information layer that the musical surface offers and focus on holistic aspects of the musical stimulus, identifying or appreciating elements that form on higher levels of information organization, such as a harmonic style, or thematic material or a cadential pattern.

Compressed data and information (learned from data or taken as prior knowledge) may be represented by symbols. Symbols point to (signify) something quite complicated in all its fine detail such as a physical object, an event, an idea. Humans use symbols to communicate between them, transferring rich information in a succinct manner (just a few words may convey rich meanings that would otherwise require extremely large and complex data structures to pass on the same meanings).

The so-called 'deep neural networks' present the important ability to build knowledge on higher-levels of representation. Deep learning is essentially a statistical technique for classifying patterns, based on sample data. It maps a set of inputs to any set of outputs; for instance, in speech recognition, an Artificial Neural Network (ANN) learns a mapping between a set of speech sounds and a set of phonemes. Deep learning systems work particularly well for parsing and classifying raw sensory input data in the visual and/or auditory domain.

The success of ANNs learning from data is restricted by the specific training dataset context. The efficiency of such methods deteriorates when insufficient amounts of training data is available, when the domain of application is shifted and when intelligent reasoning is required for rapid adaptation to new environments (Rosenfeld and Tsotsos, 2018). But why can humans, who also learn from stimuli in their environment, so easily perform domain shift and zero-shot learning (i.e. identification of an element based on specifications that do not appear in the training data)? For instance, how is a child able to identify that an animal is a zebra at first sight, having available only the description that "zebra is a horse with black and white stripes"?

In a study presented by Dubey et al. (2018), a similar question was asked: Why do humans perform well in games they play for the first time - or at least better than machine learning systems trained to perform well at other games? In this study, human participants were asked to play different versions of a platformer retro-style computer game (with similar goals and level design as Atari's "Montezuma's Revenge"); in each version, alternative visual textures were devised for masking the functionality of different components of the game. Aim of this study was to explore the importance of "human priors", e.g. prior knowledge on the functionality of stage components, in game performance. The alternative textures were used as a means to "camouflage" different visual components, gradually disabling visual identification of the function these components imply; e.g., ladders for climbing, enemies to avoid, keys to open doors etc. Some versions also included altered physical qualities (e.g., effect of gravity) and interactions between game agents, but with preservation of the underlying structure, goals and rewards.

Dubey et al. (2018) showed that as the visual interface of the game was altered, the performance of the human players degraded. This fact indicates that prior knowledge is a crucial factor that allows humans to achieve good performances in first-time encountered games. Reinforcement learning systems need to build a model for identifying the functionality of all components of the environment from scratch, after numerous "blind" trial and error simulations; and this hard-earned knowledge is strictly domain-dependent, rendered worthless for new games. There are promising signs that task-agnostic priors can be acquired from data in dynamical systems (Du and Narasimhan, 2019). However, ad hoc modelling remains to this day an effective way of tackling various problems, since it offers the possibility to model elements in an environment through 'hand-crafted' definition of what the priors are.

Human perception of music relies on prior knowledge organised in complex networks of concepts on many levels. The human brain groups together the numerous harmonics of a plucked string into an single integrated tone. Multiple notes occurring simultaneously are grouped into a chord, which is an entity in its own right, carrying functions, meaning, even emotions, that extend beyond the isolated role of each constituent note. The notion of the root of a chord, for instance, is attributed to a note (often missing from the simultaneity that constitutes the chord) depending on complex psychoacoustic phenomena. Such concepts that can be modelled as priors in a computational system; for instance, the General Chord Type (GCT) representation enables automatic encoding of note simultaneities in a form that is close to traditional chord types based on consonance / dissonance optimisation, root and scale degree identification (Cambouropoulos et al., 2014). Of course, such priors can be learned implicitly from data, at the cost of having to collect and annotate enough data. Having such priors, however, explicitly available in symbolic models, makes it possible to develop creative systems that can tackle complex tasks with access to relatively small datasets - or even with toy-example models of musical spaces.

Symbolic AI's strengths lie in the fact that symbolic representations are abstract and can therefore be generalised and transferred in different tasks, and, additionally, due to their affinity to language, they can be easily interpreted and understood by humans. Symbols enable transferring knowledge to other occasions / problems, since knowledge embodied in a symbol is abstract and can be applied to other actual instances; for instance, characterising a newly heard piece as Jazz transfers our broader knowledge of Jazz to that specific piece.

Reasoning involves combining and manipulating symbols allowing arguments and inferences to be made. Symbolic AI implements symbolic reasoning in rule-based or knowledge engineering systems. In such systems, humans must first study, learn and model how certain input relates to specific output and, then, hand-craft the rules into a program. Such rules may rely on distilled knowledge acquired through experience and/or on general cognitive principles. Symbolic AI has a number of drawbacks the most important of which is that, as it requires manual coding, it does not allow dynamic change and it cannot capture the complexity of the real world (see Nilsson 2017 on pros and cons of symbol systems). This problem can be partly addressed by introducing statistical learning that allows some amount of adaption of rules to specific contexts.

Deep learning models, on the contrary, are flexible, adaptive, easy to build as they do not require fully fleshed-out models, and they are resilient to noise or incomplete information. They have, however, limitations particularly in high-level cognitive tasks where generalisation is required; ANNs tend to fail disastrously when exposed to data outside the pool of data they are trained on. As they are black-boxes, it is not clear how they work, what is learned and how intermediate activation patterns may be interpreted. Additionally, they are data-hungry requiring huge amounts of data to capture even simple concepts and need enormous computing power and storage space (see Marcus 2018 for a critical appraisal).

Recently, attempts are made to reconcile symbolic systems, that are strong in abstraction and inference, with deep learning that excels in perceptual classification (Garnelo and Shanahan 2019). Combining symbolic AI with deep learning may assist addressing the fundamental challenges of reasoning, hierarchical representations, transfer learning, robustness in the face of adversarial examples, and interpretability (or explanatory power).

The different approaches of Symbolic AI and ANNs in music research are discussed by Geraint Wiggins (1999) and Petri Toiviainen (1999) in the same volume on Music and Artificial Intelligence (Miranda, 1999); also see Chapter 9, by Geraint Wiggins, in this volume.

In this chapter, emphasis is given to the advantages of traditional symbolic computational modeling of musical tasks. Building computational systems that rely on cognitive-based and/or music-theoretic-based systematic descriptions of processes involved in mapping certain input to certain output, enables the development of sophisticated models that may have both theoretical and practical merits. In terms of theory, our understanding of music *per se* is enriched, traditional assumptions are tested, empirically-derived cognitive principles evaluated and new musical knowledge is acquired. As knowledge is explicit in such AI models, sophisticated practical systems can be created that allow intelligent interaction with musicians / users though the manipulation of meaningful symbolic representations (e.g., educational systems, compositional assistants, interactive performers, content-based music search engines, and so on). Such systems make use of prior sophisticated knowledge acquired through years (or even centuries) of experience and introspection, and, also, capitalize on findings resulting from empirical work in music cognition. This way sophisticated models can be built relatively quickly combining diverse components on different hierarchical levels of organisation. Additionally, symbolic systems reinforced with simple statistical learning capacities, can adapt to different contexts based on relatively small training datasets allowing this way a certain degree of flexibility. Furthermore, such models can bridge different conceptual spaces enabling the invention of novel concepts not present in the initial input spaces. All these qualities will be discussed in more detail in the following sections, focusing on computation models in the domain of musical harmony (analysis and generation).

10.3 Melodic Harmonisation: symbolic and subsymbolic models

Various methods for music analysis and generation - and specifically harmonisation - following the ‘classical’ AI approach have been presented during the past decades. The first score generated by a computer, the *Illiac Suite* string quartet composition in 1957 (Hiller and Isaacson, 1979), included a mix of rule-based approaches and Markov transition matrices for composing cantus firmus music, rhythmic sequences and four voice segments. The first attempts in building cognitively-inspired computational models should probably be attributed to the pioneering work of Christopher Longuet-Higgins, a cognitive scientist that proposed among others, a key-finding (Longuet-Higgins and Steedman, 1971) and a meter-finding (Longuet-Higgins and Lee, 1984) model that processes notes in a left-to-right fashion based on fundamental music theoretic and cognitive concepts (collected essays can be found in Longuet-Higgins, 1987).

Among the most complete approaches to modelling four-part chorales in the style of Johann Sebastian Bach was presented by Ebcioglu (1988) in the CHORAL expert system, which comprised 270 rules to represent the knowledge for harmonising a melody. A review of such purely hand-crafted rule-based approaches can be found in Pachet and Roy (2001), while a more recent study of such systems was presented by Phon-Anmuaisuk et al. (2006). Rule-based methods are useful for examining the nuts and bolts of a musical style and for studying how several components or ‘musical concepts’ are interrelated towards forming what listeners identify as a coherent musical style.

A more generic approach to rule-based modelling of harmony, is to model a wider range of harmonic genres through generative grammars. Rohrmeier (2011) and Koops et al. (2013) have presented grammars that model tonal harmony and Granroth-Wilding and Steedman (2014) develop grammars that describe Jazz style harmony. Grammars offer a clear and powerful interpretation of how high-level harmonic concepts are hierarchically organised and what their relations and functions are. Harmonic grammars so far enable primarily describing musical surfaces in terms of chord symbols; generated symbols, however, cannot be rendered to actual musical surfaces. It is, also, still difficult to adapt such grammars to diverse styles, since their formulation is based on specific alphabets of chord labels and (manually-constructed) hierarchical relations between them.

The methods discussed so far are not adaptive, in a sense that specific rule-sets represent specific musical styles; representing new styles would require to come up with new sets of rules. Probabilistic generative models can capture probabilities of occurrence of specific elements in a dataset, therefore offering a way of adaptation to specific styles. Among the most popular probabilistic AI approaches are Hidden Markov Models (HMM). Regarding harmonisation, and specifically melodic harmonisation, HMM model conditional relations between chords, melodic notes or other information of interest (e.g., chord functions). After learning from data, new harmonies that reflect learned characteristics can be generated by sampling from the distribution of such learned conditional probabilities, or traversing paths of optimal probabilities over a given set of conditions (e.g., composing the optimal harmonic path over a given melody).

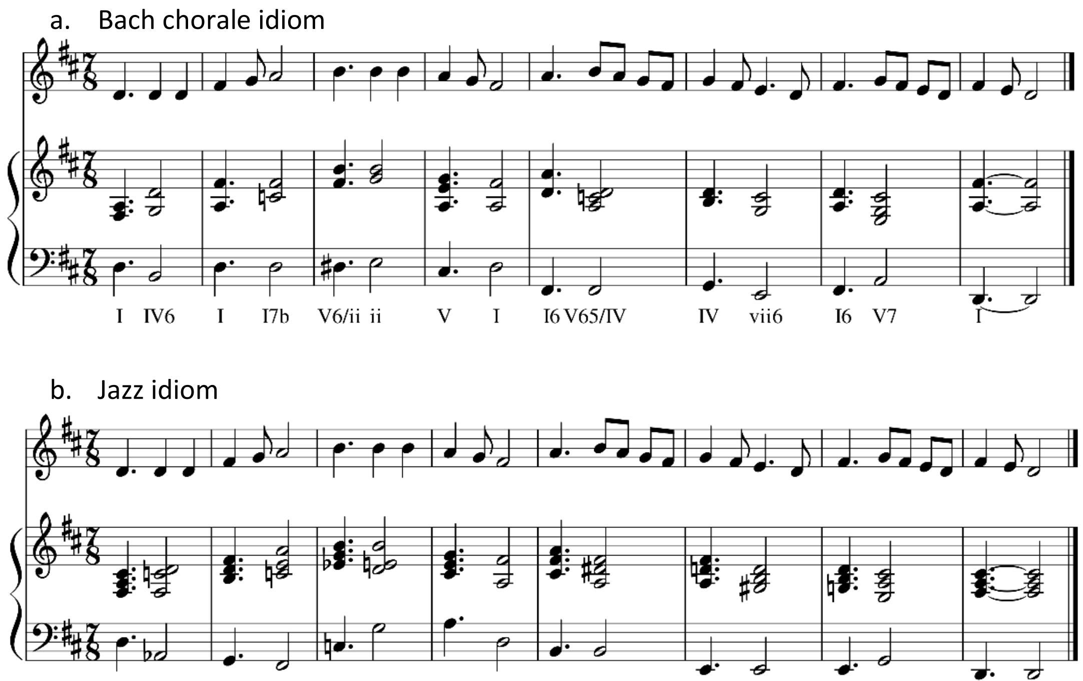
Among many examples of employing HMMs for melodic harmonisation, the approaches of Allan and Williams (2004) and Raphael and Stoddard (2004) incorporated a dual HMM for four-part harmonisation: role of the first HMM was to define a coarse functional layout over a given melody, while the second HMM assigned specific chord symbols given the functional labelling and the melody. The idea of layering additional information in HMM-based models was also discussed by Raczynski et al. (2013), where information about local tonality was incorporated as conditions for defining chord symbols over a given melody. In the original "MySong" application (Simon et al., 2008), users could sing a melody and then select a mixing rate between classical and jazz harmonies; two HMM models trained on classical and jazz music data respectively where then be combined into a single model for generating the desired harmonic mix.

Music composed by humans incorporates meaning on many structural levels. For instance, tonal music includes sections, periods, phrases with sub-phrases, conveying the essence of closure on different levels. An important weakness of the Markov-based models is that they cannot capture long-term structure since their conditional context includes information on fixed-size window frames in time. Even though the context can be increased (Markov models of higher order; i.e., conditioning their prediction based on information further back in the past), this increase quickly leads to highly specialised models that actually model specific segments in the training data, rather than stylistic properties in the data.

One way to overcome the 'locality' effect in the prediction of Markov-based models is to employ a hierarchical stratification of Markov models, where layers on top capture information about what model parts should be used in lower layers. Thereby, dependencies further away in time are captured in the top layers, while the generalisation capabilities of low order Markov is preserved in the bottom. For example, Thornton (2009) presented a Hierarchical Markov model for harmonic generation, where the top model would define the succession or repetition of chord-generating hidden states. Hierarchical relations in combination with probabilistic modelling is also achieved with Probabilistic grammars, which offer a way to learn and model alternate hierarchical properties of strings of musical sequences (Abdallah et al., 2016). Additionally, more complex, probabilistic models have been proposed that incorporate information about the metric position in the chord (Dixon et al., 2010) or voice (Whorley et al., 2013) decision process – for melodic and four-part harmonisation respectively.

Another approach to overcome the locality problem of Markov-based models is to "tie" the generative process on structure-inducing landmarks that indicate harmonic structural closings or phrase endings, i.e. cadences. To this end, methods have been proposed that focus on generating chord sequences that end with proper cadences. An approach that has been examined by Yi and Goldsmith (2007) and Borrel-Jensen and Hjortgaard Danielsen (2010) is to incorporate a special cadence evaluation scheme for rating/discarding entire melodic harmonisations generated by a Markov-based system. Other approaches examined learning chord sequences from start to end (Allan and Williams, 2004; Hanlon and Ledlie, 2002), making sure that the conditional probability for "starting" the chord sequence would allow only valid endings with proper cadences. If positions of intermediate cadences that determine lower-level phrases are known, then Markov models with constraints can be employed (Pachet et al., 2011).

A simple approach for composing melodic harmonisations under this scheme was presented by Kaliakatsos-Papakostas and Cambouropoulos (2014), where constraints are added at phrase boundaries ensuring appropriate cadential schemata at structurally important positions; intermediate chord progressions are filled in according to the learned chord transition matrices. This method is incorporated in the CHAMELEON melodic harmonisation assistant (Kaliakatsos-Papakostas et al., 2016; 2017) that is adaptive (learns from data), general (can cope with any tonal or non-tonal harmonic idiom) and modular (learns and encodes explicitly different components of harmonic structure: chord types, chord transitions, cadences, bass line voice-leading). This system preserves the merits of rule-based systems in its overall hierarchical modular outlook and at the same time it is enhanced with statistical learning mechanisms that extract context-sensitive harmonic information enabling adaptation to different harmonic idioms. Two examples of melodic harmonisation of a traditional Greek diatonic melody in the Bach chorale and azz idioms is presented in Figure 10.1. In CHAMELEON, cadence locations are given by the user (or assumed by default only at the end of the harmonisation); automatic methods, however, for identifying potential intermediate phrase boundaries and cadence positions can be employed.


a. Bach chorale idiom

b. Jazz idiom

Figure 10.1: Harmonisation of the traditional Greek melody *Milo Mou Kokkino* (repetition of phrases omitted) by CHAMELEON in: a. Bach chorale idiom, and b. Jazz idiom. Voice-leading is incomplete and erroneous as only the bass line movement is modelled.

Symbolic AI modelling requires manual encoding of information in the form of explicit rules about note / chord relations. These rules can be probabilistic and, therefore, adaptive to peculiarities of specific datasets; nonetheless, these rules still capture specific aspects of the richly diverse and hierarchically structured information that is incorporated in musical surfaces. In relation to the harmonisation methods discussed above, in most cases the composition process ends at the point where chord symbols are generated. Converting chord symbols to actual notes, or implementing voice leading with the generated chord symbols, requires further complicated models that take into account auditory stream and voice separation principles, segmentation and phrase structure, metric structure and harmonic rhythm, dissonance and consonance, as so on. Probabilistic approaches have been examined for determining the bass voice in a generated chord sequence (Makris et al., 2015), but the problem of proper and complete voice leading is still very complicated even in the extensively studied case of the Bach chorales (see problems with voice-leading in Figure 10.1a). Therefore, symbolic models do not offer a ‘holistic’ description of the music they model, since they are only able to model and generate specific aspects of information that is described by explicit representation. A holistic musical model that is general enough to describe diverse musical styles and sophisticated enough to generate high quality musical surfaces in different idioms is still an elusive goal. Deep learning techniques seem to promise a faster way to accomplish such holistic behavior or at least give the illusion that this goal is easier to achieve with ANNs.

Deep learning and ANNs have also been used for generating harmonisations. DeepBach (Hadjeres et al., 2017) is an example of combining LSTM (looking both forward and backward in time) and feedforward components for capturing different modes of musical information from the score. This system learns from Bach chorales that are encoded using information about each voice separately, including also information about the metric structure (time signature and beat position), key signature and locations of the fermata symbol (indicating phrase endings). After DeepBach is trained with multiple Bach chorales that include annotations for all the aforementioned information modes, it can generate new Bach chorales either from scratch or by completing certain parts of a given score (e.g., harmonise a melody or fill specific parts on a given score). To generate from scratch, DeepBach needs an input that contains information about the metric structure, key signature and locations of fermatas. With this input, the system follows a process similar to Gibbs sampling: it initially generates random notes for each voice and during many iterations (on a magnitude of tens of thousands) single random notes from random voices are selected and readjusted with a sampling process, based on the joint distribution reflected by the system according to each current setup of notes. As iterations evolve, the initially random setup of notes slowly converges to a new setup that follows the style of Bach chorales. Score-filling is performed in a similar manner, but the notes given in the input are not subject to change by the sampling process.

Even though there are errors in the end results, DeepBach is able to learn important elements of high-level information including chords and cadences. An impressive aspect of how ANNs learn is the ability of this system to learn high-level features implicitly, meaning that such information is not encoded explicitly in the data, but it emerges inside the latent variables of the network. Any time DeepBach composes a new piece, either from scratch or by filling a given excerpt, it effectively ‘explores’ the space of all possible Bach chorales, with a (considerably high) degree of accuracy on many diverse levels (chords, chord progressions, cadences etc.). This is achieved by starting from different random note setups (in the part it is expected to fill) and then converging with random sampling to a new piece that reflects the overall characteristics of a Bach chorale.

Another approach for exploring harmonisations has been presented by Google in the ‘Bach doodle’ application (Huang et al., 2019b) that appeared on the interface of the popular search engine in 2019 (on J. S. Bach’s birthday). This approach is based on two-dimensional convolutional neural networks that have been very effective in image recognition tasks, since they can capture two-dimensional spatial patterns. The ‘Bach doodle’ is based on a music adaptation of the Coconet (Huang et al., 2019a), that is able to find patterns in the two-dimensional time-pitch space. Even though this approach does not process temporal information in the way that DeepBach does; that is, time is considered only in the context of two-dimensional pattern rather than accumulated dynamics within in a sequence, the principles of music generation are similar: new compositions are explored by sampling on probability spaces that are formed by combining learned convolutional filters. DeepBach and the Bach doodle are two among numerous examples of deep learning systems that exhibit the impressive ability to infer high-level features from musical surfaces and generate new music by effectively exploring new possible materialisations of musical surfaces based on the learned latent spaces.

A trained ANN involves symbol manipulation in the sense that the input data are encoded as strings of symbols and labels. Musical knowledge is manually incorporated in the training data; for instance, in DeepBach, notes are encoded explicitly in separate voices, metrical structure and tonality are given, and phrase structure is explicitly annotated. The more information is annotated explicitly in the training data, the better the resulting learning and performance of the system. This, however, comes at the expense of making the initial representation and symbolic preprocessing more complex, compromising thus the simplicity of the Deep Learning approach. Deep learning is not miraculous. It requires meaningful data to learn from. Humans may learn Bach chorale harmony simply by being exposed to Bach chorales. Human listeners, however, have prior knowledge regarding beat, metre and rhythm, have the ability to separate auditory stimuli into separate streams, knowledge regarding tuning and scale systems, can parse sequences of notes to smaller phrases, have a latent understanding of consonance and dissonance, and a whole system of chord hierarchies; all this knowledge plays a role in learning Bach chorale structure through mere exposure. A computational system (either symbolic AI or deep learning) needs one way or another such information. Bringing closer together symbolic reasoning and connectionist approaches may be a good way to deal more effectively with highly complex data such as musical data (abstract, multi-parametric, hierarchical, multi-layered).

10.4 Inventing New Concepts: Conceptual Blending in Harmony

Conceptual blending is a cognitive theory developed by Fauconnier and Turner (2003) whereby elements from diverse, but structurally-related, mental spaces are ‘blended’ giving rise to new conceptual spaces that often possess new powerful interpretative properties allowing better understanding of known concepts or the emergence of novel concepts altogether. In the context of the COINVENT project (Schorlemmer et al., 2014), a formal model has being developed inspired by category theory, wherein two input spaces (I1 and I2) that share a number of common properties (Generic space) are combined into a new blended concept (Blend) after incompatibilities, inconsistencies and contradictions have been eliminated (Confalonieri et al., 2018). As an illustration of the model’s potential, the proof-of-concept computational creative assistant CHAMELEON that performs melodic harmonization and blending has been developed (Kaliakatsos-Papakostas et al., 2017; CHAMELEON, 2020).

What concepts are there to be blended in music? Focusing on harmonic structural blending (rather than cross-domain blends between; e.g., text and music, image and music, or physical motion and music), musical concepts are taken to be generalisations of harmonic entities and relations, derived from a corpus of harmonic annotated data via statistical learning. This data-driven approach ensures that learned concepts reflect adequately characteristics of diverse harmonic idioms. From each independent harmonic space (e.g., modal, common-practice tonal, jazz, atonal, organum, etc.), represented by a set of characteristic annotated music pieces, the following structural characteristics are learned and explicitly encoded: chord types, chord transitions (probabilistic distributions), cadences (i.e., chord transitions on designated phrase endings at different hierarchic levels), and voice-leading (i.e., bass line motion in relation to melody, bass-melody distance, chord inversion). This structural information sometimes corresponds to standard musicological linguistic terms (e.g., 'cadence', 'perfect cadence', 'dominant', 'leading-note', etc.), bringing the learned musical concepts closer to the standard notion of 'concept' in the domain of cognitive linguistics. Such features drawn from diverse idioms may be combined so as to create new harmonic blended styles; for instance, tonal cadences may be assigned to phrase endings and modal chord transitions may be employed for filling in the rest of the phrase chords (Figure 10.2).

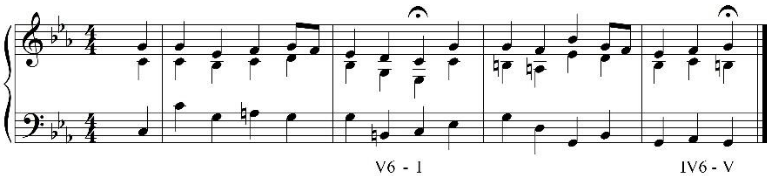


Figure 10.2 Bach Chorale melody harmonised in medieval Fauxbourdon style with inserted tonal cadences.

Take for instance the concept of *perfect cadence* in common-practice harmony and the *phrygian cadence* of renaissance music. The former has some salient features appearing in all instances found in, for instance, the Bach chorale dataset: leading note resolved upwards to tonic, seventh (in dominant chord) resolved downward by step, leap from root of dominant chord to tonic chord in the bass line. The later contains always a downward leading tone moving by semitone to the tonic, and an upward movement of the seventh degree by tone to the tonic. If these two cadential concepts are imported in the formal blending model (as I1 and I2), the highest rating blend in terms of preserving salient features from both input cadences and also ensuring that the resulting chord types are acceptable in these idioms (e.g. major triad, minor triad, major seventh chord etc), is the *tritone substitution* progression which is commonly found in jazz music (it contains both upward and downward leading notes). In this case, by blending two established harmonic concepts, a new concept is invented that has not been seen in the training data (Zacharakis et al., 2017).

The above cadence blending process is generalised to any two input chord transitions, allowing the creative blending of entire chord transition matrices from different idioms. Let us assume a 'toy' harmonic space where, within a major tonality, only three chords exist, namely, the tonic, subdominant and dominant seventh chords. It is clear that such 'toy' chord transition spaces of, for instance, C major and F# major tonalities have no common chords and do not communicate, so it is not possible to move from one space to the other. Can the proposed chord transition methodology 'invent' new transitions and possibly new chords that may connect the two spaces in a meaningful way? The chord transition blending methodology is applied to all the chord transitions in the C major and F# major tables, i.e. each chord transition in the first matrix is blended with each chord transition in the second matrix producing a list of resulting blends. The resulting blends are ranked according to criteria that take into account the number of salient features from each input space preserved in the blend and the balance of the contribution of each input chord transition in the blend. The highest ranking chord transition blends include a sort-of tritone substitution or German sixth chord transition (i.e., G7 → F# or D*b*7 → C), and the diminished seventh chord (i.e., B°7→C or E#°7→F# where E#°7 is enharmonically identical to B°7). The first blended transition establishes a connection between existing chords of the two input spaces, whereas the second proposes a new chord, a diminished seventh chord, that constitutes a bridge between the two tonal spaces - see detailed description in (Kaliakatsos-Papakostas et al., 2017). If more transition blends are allowed, the resulting transition table is augmented and populated with more connections between the two spaces.

The special purpose-made melody in Figure 10.3 contains a remote modulation from C major to F# major and back again to C major. This rather awkward melody cannot be harmonized correctly by the learned Bach Chorale harmonic style as chord transitions in C major cannot cope with the transitions to/in the F# major region (and vice-versa). Even if a key-finding algorithm indicates the exact positions of the modulations so that the relevant keys may be employed in the appropriate regions, the transitions between the regions would remain undefined (random chord transitions). Chord transition matrix blending of the sort previously discussed, creates meaningful connections between the two tonal regions and the melody can be harmonized correctly by the blended transition table (Figure 10.3).

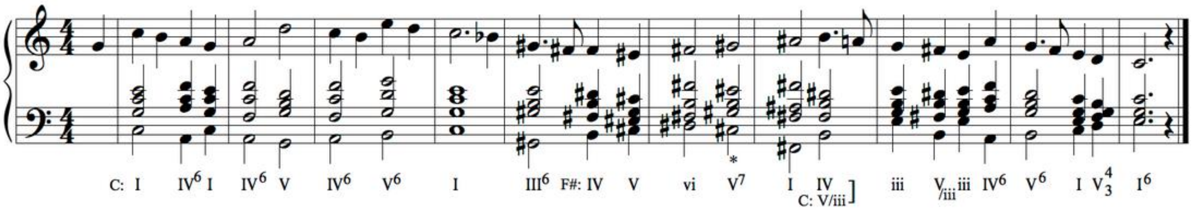
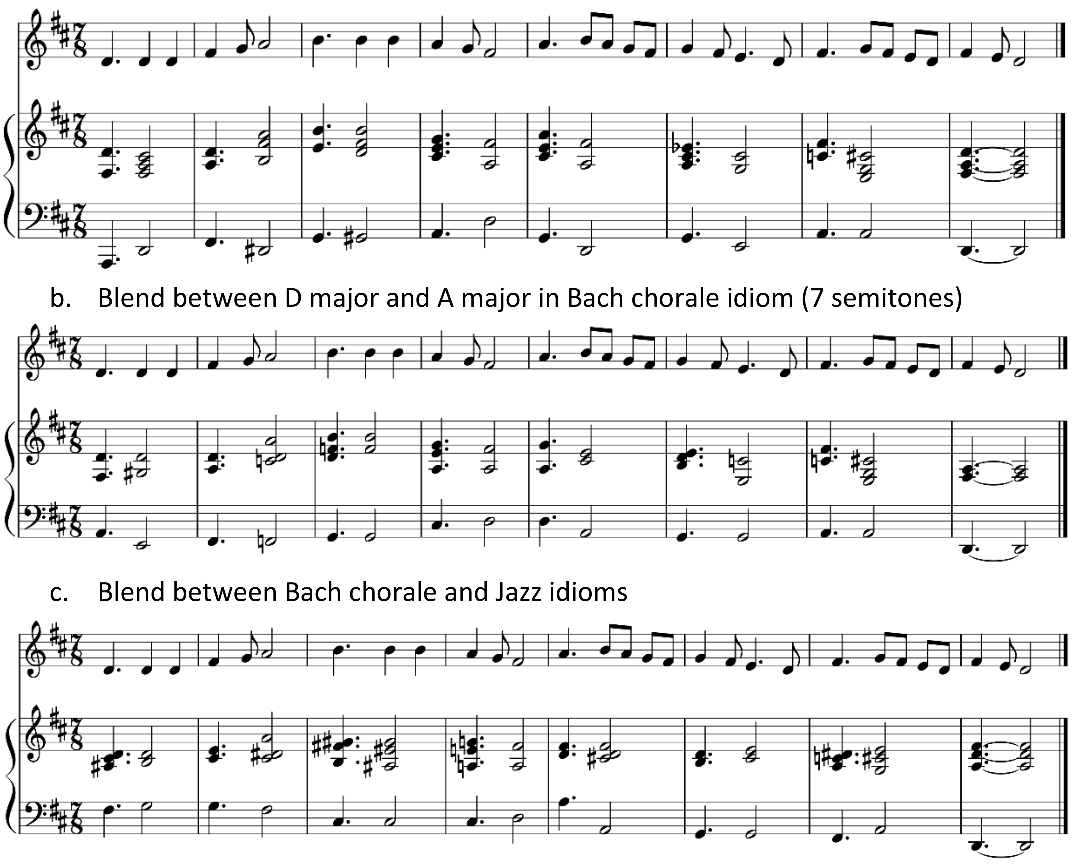

C: I IV6 I IV6 V IV6 V6 I III6 F#: IV V vi V7 I IV iii V/iii iii IV6 V6 I V4 3 I6
*
C: V/iii

Figure 10.3. Melody with distant modulation between C major and F# major is successfully harmonized by CHAMELEON after applying blending between the two tonal spaces (from Kaliakatsos-Papakostas et al., 2017, Figure 8b).

Blending different tonal spaces (different keys) in the same harmonic style can be used creatively for introducing chromaticism and more advanced harmonies that go beyond the initial tonal spaces. The traditional Greek melody *Milo Mou Kokkino* in D major can be harmonised in many different ways, with blended variations of the Bach Chorale major harmonic idiom in various shifted tonalities. In Figure 10.4a and 10.4b, D major is blended with G# major (tritone distance between keys) and D major is blended with A major (7 semitone distance). In these examples the harmony deviates from common practice functional progressions towards free chromaticism. The produced chords cannot always be explicitly identified as belonging to one of the blended spaces. It is also interesting that the blended tonal spaces can produce such a diverse range of forced harmonic chromaticism, with elements of tonal mixture and chords of ambiguous functionality, even though the melody is purely diatonic (without any chromatic elements). In this case, blending produces novel harmonic spaces that go well beyond the initial diatonic input spaces.

An example of blending different harmonic spaces, namely Bach chorale tonal with Jazz is shown in Figure 10.4c. This example illustrates a harmonisation that is neither plain tonal (as in Figure 10.1a) nor Jazz (as in Figure 10.1b); this harmonisation has a distinct character with a mixture of simple and extended triads, and shows a high degree of originality in relation to the more well established contributing idioms.

a. Blend between D major and G# major in Bach chorale idiom (6 semitones)



b. Blend between D major and A major in Bach chorale idiom (7 semitones)

c. Blend between Bach chorale and Jazz idioms

Figure 10.4. Harmonisation of the traditional Greek melody *Milo Mou Kokkino* (repetition of phrases omitted) by CHAMELEON in: a. Blend between two major tonalities 6 semitones apart, b. Blend between two major tonalities 7 semitones apart, and e. Blend between Bach chorale and Jazz idioms.

Deep learning techniques can be used for generating morphs between different spaces. Why employ symbolic AI techniques if this is possible? 'Interpolated' music generation has been explored by leveraging the spatially interpretational capabilities of the latent space in the Variational Autoencoder (VAE) (Kingma and Welling, 2013). Examples in image generation (Gulrajani et al., 2016) have shown that a continuum of new images can be generated that include intermediate-morphed characteristics between two input images. This continuum is constructed by a specialised training process that includes two phases steps, divided by an intermediate sampling step. Similar to the 'vanilla' Autoencoder, one goal of the training process is to perform accurate reconstruction of the input, while a parallel goal is to construct latent representations that follow a Gaussian distribution. Given enough data, the latent space obtains continuous characteristics and, thereby, sampling between any two points is made possible. Images generated by points on the line that connects any two points in the latent space, exhibit the effect of morphing between the characteristics of the images that correspond to the two extremes.

In music, both interpolation and extrapolation from two given excerpts has been examined by (Roberts et al., 2018). For instance, if the melodies corresponding to two extreme points in the latent space were a major and a minor melody, sampling from interpolated points between the latent representations of the inputs would generate new melodies with intermediate levels of major and minor characteristics. Sampling from latent points that are closer to, for instance, the major excerpt would generate a melody that is closer to a major melody than sampling closer to the minor end. Even though this system learns from data that represent musical surfaces, the ‘morphing continuum’ that is being formed between any two points in the latent space includes high-level information as, for example, tonality.

Exploring spaces in-between two learned spaces is possible, as with the Variational Autoencoder; in case the two learned spaces are two musical styles, morphing between the two styles is made possible. There are, however, some shortcomings with this morphing approach. Firstly, extensive amounts of data are required for learning two styles, with a view to creating the continuous latent space. Secondly, high-level features in the latent spaces are not transparent, in a sense that it is not clear which features are represented by which latent space variables. For instance, it is not possible to force such a system to generate a major melody without providing an example of how a major melody looks like. The beta-VAE variation (Higgins et al., 2017) potentially allows disentanglement of prominent features, but again, concrete features are not necessarily clearly divisible.

Except from the above shortcomings, there is also an inherent limitation: musical materialisation of latent space points only happens by rehashing material in the musical surface of the (numerous) examples that were encountered during training. Therefore, such systems are able to interpolate (and even, to some extent, extrapolate) between musical styles, but they are able to do so only by reproducing elements of what already exists in the training data. This approach does not enable the creation of new concepts that allow creative connections between two seemingly disjoint spaces. The creation of such concepts is possible using Conceptual Blending, which allows the creation of combinational components that connect disjoint spaces, with very few (if any) training and with transparent access to what concepts are combined (however, at a cost in hand-crafting the relations between low-to-high-level features).

10.5 Conclusions

In this chapter, recent research in the domain of melodic harmonisation and computational creativity has been presented with a view to highlighting strengths and weaknesses of the classical cognitively-inspired symbolic AI approach (often in juxtaposition to contemporary deep learning methodologies). A modular melodic harmonisation system that learns chord types, chord transitions, cadences and bass line voice-leading from diverse harmonic datasets is presented. Then, it is shown that the harmonic knowledge acquired by this system, can be used creatively in a cognitively-inspired conceptual blending model that creates novel harmonic spaces combining in meaningful ways the various harmonic components of different styles. This system is essentially a proof-of-concept creative model that demonstrates that new concepts can be invented that transcend the initial harmonic input spaces. It is argued that such original creativity is more naturally accommodated in the world of symbolic reasoning that allows links and inferences between diverse concepts at highly abstract levels. Moreover, symbolic representations and processing facilitate interpretability and explanation that are key components of musical knowledge advancement. Finally, reconciling symbolic AI with deep learning may be the way forward to combine the strengths of both approaches towards building more sophisticated robust musical systems that connect sensory auditory data to abstract musical concepts.

10.6 References

Abdallah, S., Gold, N., & Marsden, A. 2016. Analysing symbolic music with probabilistic grammars. In *Computational music analysis* (pp. 157-189). Springer, Cham.

Allan, M., Williams, C.K.I. 2004. Harmonising chorales by probabilistic inference. *Advances in Neural Information Processing Systems 17*, MIT Press, 25–32.

Borrel-Jensen, N., Hjortgaard Danielsen, A. 2010. Computer-assisted music composition – A database-backed algorithmic composition system. B.S. Thesis, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark. B.S. Thesis.

Cambouropoulos, E., Kaliakatsos-Papakostas, M. & Tsougras, C. 2014. An Idiom-independent Representation of Chords for Computational Music Analysis and Generation In *Proceedings of the Joint ICMC-SMC*, Athens, Greece.

CHAMELEON. 2020. http://ccm.web.auth.gr/chameleonmain.html (Accessed on 22 April 2020)

Confalonieri, R., Pease, A., Schorlemmer, M., Besold, T. R., Kutz, O., Maclean, E., & Kaliakatsos-Papakostas, M. (Eds.). 2018. *Concept invention: Foundations, implementation, social aspects and applications*. Springer.

Dixon, S., Mauch, M., & Anglade, A. 2010, June. Probabilistic and logic-based modelling of harmony. In *International symposium on computer music modeling and retrieval* (pp. 1-19). Springer, Berlin, Heidelberg.

Dubey, R., Agrawal, P., Pathak, D., Griffiths, T. L., & Efros, A. A. 2018. Investigating human priors for playing video games. arXiv preprint arXiv:1802.10217.

Du, Y., & Narasimhan, K. 2019. Task-Agnostic Dynamics Priors for Deep Reinforcement Learning. arXiv preprint arXiv:1905.04819.

Ebcioglu, K. 1988. An expert system for harmonizing four-part chorales. *Computer Music Journal* 12 (3), 43–51. ISSN 01489267.

Fauconnier, G., & Turner, M. 2003. *The way we think: Conceptual blending and the mind's hidden complexities* (reprint ed.). New York, NY: Basic Books.

Feinberg, T. E., & Mallatt, J. 2013. The evolutionary and genetic origins of consciousness in the Cambrian Period over 500 million years ago. *Frontiers in psychology,* 4, 667.

Garnelo, M., & Shanahan, M. 2019. Reconciling deep learning with symbolic artificial intelligence: representing objects and relations. *Current Opinion in Behavioral Sciences*, *29*, 17-23.

Granroth-Wilding, M., Steedman, M. 2014. A robust parser-interpreter for jazz chord sequences. *Journal of New Music Research 43*(4), 355-374.

Gulrajani, I., Kumar, K., Ahmed, F., Taiga, A.A., Visin, F., Vazquez, D., Courville, A. 2016. Pixelvae: A latent variable model for natural images. arXiv preprint arXiv:1611.05013

Hadjeres, G., Pachet, F., Nielsen, F. 2017. Deepbach: a steerable model for bach chorales generation, Proceedings of the 34th International Conference on Machine Learning-Volume 70, 1362–1371.

Hanlon, M., Ledlie,T. 2002. Cpubach: An automatic chorale harmonization system. http://www.timledlie.org/cs/CPUBach.pdf

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., Lerchner, A. 2017. beta-VAE: Learning basic visual concepts with a constrained variational framework., *International Conference on Learning Representations*, 2(5), p. 6.

Hiller, L.A., Isaacson, L.M. 1979. *Experimental Music; Composition with an electronic computer*. Greenwood Publishing Group Inc.

Huang, C.Z.A., Cooijmans, T., Roberts, A., Courville, A., Eck, D. 2019a. Counterpoint by convolution. arXiv preprint arXiv:1903.07227 .

Huang, C.Z.A., Hawthorne, C., Roberts, A., Dinculescu, M., Wexler, J., Hong, L., Howcroft, J. 2019b. The bach doodle: Approachable music composition with machine learning at scale. arXiv preprint arXiv:1907.06637 .

Huron, D. 2016. *Voice leading: The science behind a musical art*. MIT Press.

Kaliakatsos-Papakostas, M., Cambouropoulos, E. 2014. Probabilistic harmonisation with fixed intermediate chord constraints*, Proceedings of the joint ICMC–SMC 2014*, Athens, Greece.

Kaliakatsos-Papakostas M., Makris D., Tsougras C., Cambouropoulos E. 2016. Learning and creating novel harmonies in diverse musical idioms: An adaptive modular melodic harmonisation system. *Journal of Creative Music Systems* 1(1).

Kaliakatsos-Papakostas, M., Queiroz, M., Tsougras, C., Cambouropoulos, E. 2017. Conceptual blending of harmonic spaces for creative melodic harmonisation. *Journal of New Music Research* 46 (4), 305–328.

Kingma, D.P., Welling, M. 2013. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114

Koops, H.V., Magalhaes, J.P. and De Haas, W.B., 2013, September. A functional approach to automatic melody harmonisation. In *Proceedings of the first ACM SIGPLAN workshop on Functional art, music, modeling & design* (pp. 47-58). ACM.

Longuet-Higgins, H. C., & Steedman, M. J. 1971. On interpreting Bach. *Machine Intelligence*, 6, 221-241.

Longuet-Higgins, H. C., & Lee, C. S. 1984. The rhythmic interpretation of monophonic music. *Music Perception,* 1(4), 424-441.

Longuet-Higgins, H. C. 1987. *Mental processes: Studies in cognitive science*. Cambridge, MA: MIT Press.

Laske, O. E. 1988. Introduction to cognitive musicology. *Computer Music Journal*, 12(1), 43-57.

Makris, D., Kaliakatsos-Papakostas, M. A., & Cambouropoulos, E. 2015. Probabilistic Modular Bass Voice Leading in Melodic Harmonisation. In *Proceedings of ISMIR 2015* (pp. 323-329).

Marcus, G. 2018. Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*.

Meredith, D. (Ed.). 2016. *Computational music analysis* (Vol. 62). Berlin: Springer.

Miranda, E. R. (Ed.). 2013. *Readings in music and artificial intelligence*. Routledge.

Nilsson, Nils J. 2017. "The Physical Symbol System Hypothesis: Status and Prospects." *SpringerLink*, 2007, 9–17. https://doi.org/10.1007/978-3-540-77296-5_2.

Pachet, F., Roy, P. 2001. Musical harmonization with constraints: A survey. Constraints 6 (1), 7–19.

Pachet, F., Roy, P., Barbieri, G. 2011, Finite-length markov processes with constraints, Twenty-Second International Joint Conference on Artificial Intelligence.

Phon-Amnuaisuk, S., Smaill, A., Wiggins, G. 2006. Chorale harmonization: A view from a search control perspective. *Journal of New Music Research* 35 (4), 279–305.

Raphael, C., & Stoddard, J. 2004. Functional harmonic analysis using probabilistic models. Computer Music Journal, 28(3), 45-52.

Simon, I., Morris, D. and Basu, S., 2008, April. MySong: automatic accompaniment generation for vocal melodies. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 725-734). ACM.

Temperley, D. 2012. Computational models of music cognition. *The psychology of music*, 327-368.

Thornton, C. 2009, Hierarchical markov modeling for generative music, In *Proceedings of the International Computer Music Conference* (ICMC2009).

Toiviainen, P. 2013. Symbolic AI versus connectionism in music research. In *Readings in music and artificial intelligence* (pp. 57-78). Routledge.

Raczynski, S.A., Fukayama, S., Vincent, E. 2013. Melody harmonization with interpolated probabilistic models*. Journal of New Music Research* 42 (3), 223–235.

Roberts, A., Engel, J., Raffel, C., Hawthorne, C., Eck, D. 2018. A hierarchical latent vector model for learning long-term structure in music. arXiv preprint arXiv:1803.05428

Rohrmeier, M. 2011. Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music* 5 (1), 35–53.

Rosenfeld, A., & Tsotsos, J. K. 2018. Bridging cognitive programs and machine learning. arXiv preprint arXiv:1802.06091.

Whorley, R. P., Wiggins, G. A., Rhodes, C., & Pearce, M. T. 2013. Multiple viewpoint systems: Time complexity and the construction of domains for complex musical viewpoints in the harmonization problem. *Journal of New Music Research*, 42(3), 237-266.

Wiggins, G., & Smaill, A. 2013. Musical Knowledge: what can Artificial Intelligence bring to the musician? In *Readings in music and artificial intelligence* (pp. 39-56). Routledge.

Yi, L., Goldsmith, J. 2007. Automatic generation of four-part harmony., Laskey, K.B., Mahoney, S.M., Goldsmith, J.,(Eds.), BMA, *CEUR Workshop Proceedings*, 268.

Zacharakis, A., Kaliakatsos-Papakostas, M., Tsougras, C., & Cambouropoulos, E. 2017. Creating musical cadences via conceptual blending: Empirical evaluation and enhancement of a formal model. *Music Perception*, 35(2), 211-234.