Flexible Generation of Musical Form: Beyond Mere Generation

Arne Eigenfeldt
School for the Contemporary Arts
Simon Fraser University
Vancouver, Canada
arne_e@sfu.ca

Oliver Bown
Art and Design
University of New South Wales
Sydney, Australia
o.bown@unsw.edu.au

Andrew R. Brown
Queensland College of Art
Griffith University
Brisbane, Australia
andrew.r.brown@griffith.edu.au

Toby Gifford
Queensland Conservatorium
Griffith University
Brisbane, Australia
t.gifford@griffith.edu.au

Abstract

Despite a long history of generative practices in music, creation of large-scale form has tended to remain under direct control of the composer. In the field of musical metacreation, such interaction is generally perceived to be desirable, due to the complexity of generating form, and the necessary aesthetic decisions involved in its creation. However, the requirements for dynamic musical generation in games, installations, as well as in performance, point to a need for greater autonomy of creative systems in generating large-scale structure. This position paper surveys the complexity of musical form and existing approaches to its generation, and posits potential methods for more computationally creative procedures to address this open problem.

Introduction

Galanter's definition of generative art (2003) – "any art practice where the artist uses a system… which is set into motion with some degree of autonomy contributing to or resulting in a *completed work of art*" (italics ours) – assumes a fully finished artwork. Furthermore, implicit in this definition is that the system may involve human interaction, in that the system need only *contribute* to the final work. In practice, human involvement, whether through algorithm design, direct control by an operator or interaction with a live human performer, has remained an active presence in the dynamic generation of music.

One reason for this is that generating entire musical compositions entails the development of musical form, a highly complex task (Berry 1966). Form, discussed more fully below, involves the complex interaction of multiple musical structures in order to logically organise the work's progression in time. Strategies are required to organise these structures so as to "provide reference points for the listener to hold on to the piece, otherwise it may lose its sense of unity" (Miranda 2001).

Musical metacreation (MuMe) is an emerging term describing the body of research concerned with the automation of any or all aspects of musical creativity (Pasquier *et al*. 2016). It looks to bring together and build upon existing academic fields such as algorithmic composition (Nierhaus 2009), generative music (Dahlstedt and McBurney 2006), machine musicianship (Rowe 2004) and live algorithms (Blackwell *et al*. 2012). There have been many musically successful MuMe production systems that have generated complete compositions, and therefore generated long-term musical structure. As MuMe does not exclude human-machine interaction, these systems have tended to rely upon human-machine partnerships.

Although MuMe considers itself to be a subfield of computational creativity (CC), some definitions of the latter exclude a large part of the former. Colton and Wiggins (2012) suggest that the degree of creative responsibility assigned to a CC system *may* include the "development and/or employment of aesthetic measures to assess the value of artefacts it produces" as well as "derivation of motivations, justifications and commentaries with which to frame their output". Veale (2015) on the other hand declares that any works that do not meet these self-reflective criteria are to be deemed "mere generation". For many creative practitioners, however, and perhaps musicians in particular, "merely" generative systems can still play a hugely important role in human-computer co-creativity, and despite this lack, generative software can still be considered actively creative (Compton and Mateas 2015). Much innovation has been achieved with MuMe systems that make no claim to be fully autonomous, and as noted in the reflections on the first Musical Metacreation Weekend (Bown *et al.* 2013), the delegation of large-scale musical structure to a system is challenging, to the point that many composers felt the need to remain "in the loop" and in order to maintain control over form interactively.

In some instances, interaction with the generative system is significantly restricted by design; for example, composition may entail managing a surfeit of parameters that constrain the system's choices (e.g. Bown and Britton 2013). Alternatively, the presentation may require decisions in absence of a human, such as a continuously running installation (e.g. Schedel and Rootberg 2009) or interactive media (e.g. Collins 2008). For these very practical reasons, along with intellectual and aesthetic reasons, the need to automate long-term structure is a pressing issue in the MuMe community. This paper will describe musical form and the difficulties in its creation, present some existing methodologies for its creation and outline what we feel are some novel approaches to the problem, with particular attention to adapting generative techniques for formal design to dynamic situations.

Defining musical form

Musical form is "the result of the deployment of particular materials and processes" (Whittal 2016). Our immediate perception of music is its *surface*: the relationship between individual events: for example, a melodic phrase, a given chord, a drum beat, the signal processing on a sound recording. The selection and resulting relationships between these individual objects at a given point in time can be considered the music's *design* (after Salzer 1962). These surface elements almost always undergo some sort of organisation over time, and such methods involve the creation of *musical structures*: for example, the combination of melodic phrases into a longer melody, the organisation of harmonies into a repeating progression, the combination of related drum beats into an eight-bar phrase, the control over time of a signal processing parameter, such as the slow opening of a filter. *Form* is the consequence of the structural relationships between the various musical elements.

Kramer (1988) suggests that one difficulty in conceptualising form is due to its inherent role in organising time. While theories exist dealing with rhythm and meter (e.g. Xenakis 1992; Lester 1986), "more difficult to discuss are motion, continuity, progression, pacing, proportion, duration, and tempo", all aspects to do with musical form. Schoenberg expressed the complexity of form, stating that it requires "logic and coherence" in order for a musical composition to be perceived as being comprehensible, but that its elements also should function "like those of a living organism" (Schoenberg and Stein 1970). The dilemma for young composers, and MuMe practitioners, is achieving a balance between a strict structure that appears logical, with organic elements that engender surprise.

Forms that are not based upon functional tonality's goal-directed nature – such as the non-developmental and non-teleological structures often found in ambient music, world music, and specifically Stockhausen's *Momentform* (1963) – still require subtle and deft handling: "the order of moments must appear arbitrary for the work to conform to the spirit of moment form", yet they must not *be* arbitrary (Kramer 1978). For example, many works of Stockhausen depend upon *discontinuity* for their structural effect, but the points of division require careful selection:

> *Ending a permutational form is nearly always a matter of taste, not design. While the listener may be satisfied with a sensation of completion, the composer knows that though a series of permutations may eventually be exhausted, it does not automatically resolve. The ending's essential arbitrariness has to be disguised* (Maconie 1976).

Certain formal relationships have proven more successful than others, and these relationships became standardised: from simple procedures – such as ternary, rondo, and canon – to more complex relationships, such as sonata. All of these can be considered *architectural forms*, which pre-exist, and to which structures and surface features can be "poured into": in other words, a *top-down* approach.

The opposite method has been to allow the material to define its use: form from material (Boulez *et al.* 1964). Such organic procedures have found great success in much twentieth century art-music, including improvisation, and can be considered a *bottom-up* approach. Narmour (1991) provides a useful discussion on the interaction and opposition of these two approaches in musical composition.

While recent research on form in pre-20th century tonal music has provided new insights (Caplin 1998), none has appeared with the same depth and scope for contemporary non-tonal music. Kramer's examination of *Momentform* (1988) offers compelling views on non-teleological music, while also noting that 20th century composers' rejection of traditional (i.e., top-down) approaches to form based upon expectation have forced new non-universal formulations of large-scale organisation: "continuity is no longer part of musical syntax, but rather it is an optional procedure. It must be created or denied anew in each piece, and thus it is the material and not the language of the music" (Kramer 1978). Another theme in late 20th Century musicology has been an increasing respect for non-Western, or non-art-music structures. Consideration of long-term form in Indian, Indonesian and African music, for example, or in contemporary electronic dance music, broadens the scope of this enquiry.

	H	L		I	L	I	I	L	I	H
	1	2	3	6	9	12	17	22	23	31
meter	x			x					x	x
tempo	x								x	x
attack	x	x	x	x	x	x	x		x	x
density	x	x	x		x	x	x		x	x
harmony	x			x		x	x		x	x
motivic	x	x	x	x	x	x		x	x	x
repetition	x						x		x	x
texture	x	x	x	x		x	x	x	x	x
orchestration	x	x	x	x	x	x	x	x	x	x
register	x	x	x	x	x	x	x	x	x	x
loudness	x			x	x		x		x	x

Fig. 1. Park's analysis of discontinuities within the first 31 measures of Debussy's *De l'aube à midi sur la mer*. At left are the form delimiting parameters, above are the four architectonic levels: High, Intermediate, Lowest, (none), and the measures in which alterations occur. X indicates a discontinuity in the music for that parameter.

Music theorist Richard Parks has produced an interesting analysis of Debussy's music (1989) that provides a clue to the composer's unique organisational methods regarding structure. Parks suggests that form-defining parameters include meter, tempo, successive-attack activity, sonorous density, harmonic resources, thematic/motivic resources, repetition/recurrence, quality of texture, orchestration, register, and loudness. He then examines a variety of compositions by Debussy, and partitions the works based upon the locations of simultaneous alterations of these parameters: the discontinuities (see Fig. 1). The resulting delimiters demonstrate how a multiplicity of structural boundaries interact to create larger formal structures.

Although post-structuralist thinkers challenged how form might be considered in music (Nattiez 1990; Dahlhaus 1989), their contribution mainly concerns the potential *meaning* and *reception* found within a work's form. The former has no bearing in our discussion here; the latter has allowed new viewpoints, specifically those involving cognition and musical perception (e.g. Meyer 1956; Sloboda 1991; Deutsch 2013) to become considerations in generative music system design and use. Consistent between pre-20th century, modernist, and postmodernist concepts has been the role of modulating tension as a means of structural and formal design. Perceptual models of musical tension have recently been formulated (Farbood 2012), and, more generally, cognitive models for composition have themselves been evaluated (Pearce and Wiggins 2007); however, the latter have only been used to generate short musical excerpts, and not complete works.

Musical composition requires the organisation of material in such a way that it provides enough surface variation to maintain the listener's interest, while providing enough structural repetition in order to avoid overwhelming the listener with new material. This continuum can be compared to information theory's compressibility (Shannon and Weaver 1949), or Galanter's complexism theory (Galanter 2008), with the "sweet spot" being the formally-balanced, aesthetically pleasing musical work. Structures to control surface features in music can be generated – by humans or computationally – without too much difficulty, as shown in the wealth of interactive music production systems; knowing *when* to apply and alter such structures takes a great deal more sophistication and contextual knowledge and understanding.

Generative Music

Algorithmically generated music has a long and rich history: from Mozart's musical dice game (Hedges 1978), aleatoric compositions of Cage, Cowell and Stockhausen (Nyman 1999), through to compositions done in part by a computer program (Hiller 1970). These have been variously described as algorithmic composition (Cope 2000), generative music (Eno 1996), procedural music (Collins 2008) and, more recently, as musical metacreation (Bown *et al.* 2013). Previous computational models of musical structure include *Cypher* (Rowe 1992), *Experiments in Musical Intelligence* (Cope 1996), *GESMI* (Eigenfeldt 2013), the use of statistical prediction (e.g. Conklin 2003), the use of machine learning techniques (e.g. Smith and Garnett 2012), and agent negotiation (e.g. Eigenfeldt 2014). Sorensen and Brown (2008) explored human-guided parametric control over structure in the *MetaScore* system.

MuMe for interactive media faces the challenge of adapting to an *a priori* unknown and unfolding dramatic structure. As Karen Collins notes, in the context of video-game composition:

> *procedural music composers are faced with a particular difficulty when creating for video games: the sound in a game must accompany an image as part of a narrative, meaning sound must fulfill particular functions in games. These functions include anticipating action, drawing attention, serving as leitmotif, creating emotion, representing a sense of time and place, signaling reward, and so on* (Collins 2008).

Musical descriptions of drama are often connected with temporal structure – indeed for some music theorists "structural and dramatic factors are fundamentally inseparable" (Suurpää 2006). Other emotive descriptors for music such as tension, relaxation, anticipation and surprise are variously described as operating in the moment (Hindemith 1970; Huron 2006), across phrasal structures (Narmour 1990; Huron 2006; Negretto 2012) or across the structure of sections, movements, and entire pieces (Schenker 1972; Lerdahl and Jackendoff 1983).

MuMe for interactive media has focused, up to now, on reactivity (e.g. Eigenfeldt 2006) or on generative techniques operating over short timescales, to suit an externally supplied dramatic contour (e.g. Hoover *et al.* 2014). Current techniques include selecting from pre-composed content (Collins 2008) or algorithmic manipulation of symbolic scores (Livingstone *et al.* 2010) on receipt of a signal from the host system. What remains conspicuously absent is the dynamic generation of longer-term temporal structures.

As Nick Collins notes, "it is rare to see engagement from algorithmic composition research with larger-scale hierarchical and associative structure, directedness of transition, and interactions of content and container" (Collins 2009). Perhaps in response to this perceived dearth of fully formed generative works, Collins pursued a brief research direction involving a multi-agent generative acousmatic system, *Autocousmatic*, which created complete electroacoustic works (Collins 2012), discussed more thoroughly later.

Example Practices in Generating Form

Despite these difficulties, designers of MuMe systems have made attempts to control structure through generative means. We outline some of these approaches, both bottom-up and top-down.

Bottom-up: Perceived Structure through Self-Organisation

Many MuMe systems have relied upon the human performer, whether an improvising musician or the designer operating the machine directly, to "move the system along" (e.g. Lewis 1999; Pachet 2004). The complex interactions between human and machine can give rise to an organic self-organisation (Blackwell and Young 2004; Beyls 2007).

Some systems have attempted to impart a musical form upon the improvisation. Within *The Indifference Engine* (Eigenfeldt 2014), agents generate individual formal structures upon initiation that provide density and activity goals over the course of the work. These structures are continuously adapted based upon how they perceive the evolving environment, which includes a human performer. Within the *JamBot* (Gifford and Brown 2011) target complexity levels can be managed to vary or maintain sectional characteristics that dynamically balance the texture of human and generated parts.

Musebots (Bown *et al.* 2015) are autonomous musical agents that interact in performance, messaging their current states in order to allow other musebot to adapt. Recent musebots have been developed that broadcast their intentions, and not just their current state, thereby allowing other musebots to modify their own plans[1].

Top-down: Architectural models of structure

Adopting a more architectural approach within generative music has required pre-generation of formal structures in varying degrees. *GESMI* (Eigenfeldt 2013) creates complete electronic dance music tracks, using structural rules derived from a supplied corpus. Formal repetition is the first structural element generated, using a Markov-model learned from the example music, with surface features later filled in. Due to the clear repetitive phrases found within the original styles, *GESMI*'s forms are entirely believable.

Lerdahl and Jackendoff's much discussed Generative Theory of Tonal Music (1983) offers a tantalising model for top-down generation using musical grammars; unfortunately, it has never been successfully implemented in a production system – most likely due to its dependence upon 19th century functional tonality – and only a limited number of times as an automated analysis system (Hamanaka *et al.* 2006).

Cope (2000) models musical tension at several hierarchical levels through SPEAC: statement, preparation, extension, antecedent, consequent. Cope hand tags his corpus with these labels based upon harmonic tension, and uses these tags when selecting from the corpus in his recombinant methodology, enabling the generation of high-level templates that can be filled in later.

It is also possible to impart formal structures upon self-organising material. *Coming Together: Notomoton* (Eigenfeldt 2014) uses a multi-agent system exploring such organisation through agent negotiation. While the surface variation resulting from the agent interaction provides surface interest, variety in macrostructure is ensured through the use of an algorithm initiated at the beginning of the performance that segments the requested performance's duration into sections, replete with varying goals for the defining musical parameters.

We note that the potential to influence self-organisation is an active research area within computer science: guided self-organisation (Prokopenko 2009).

Towards Organic Top-down / Bottom-up Form Generation

For purposes of dynamic generation – for example, music for online games, generative video, or more structured musical improvisation – architectural form's inflexibility provides little attraction or utility; conversely, the more organic self-organisation model is extremely difficult to control. Instances of dynamic musical generation in acoustic situations have tended to involve improvisation (Hill 2011), although some efforts involving generative methods have recently appeared (d'Escrivan 2014). As such, there are no existing models available for computational dynamic generation of which we are aware.

Whether approaching the problem of dynamic generation from a top-down or bottom-up perspective, human interaction has remained conspicuously present. In order to design generative musical systems that can produce flexible long-term temporal structures that adjust to the dynamic situations of gaming and generative multimedia, it is necessary to remove human interaction, and provide more autonomy to the system. Such a solution is necessary for more powerful MuMe systems, while at the same time approaching a true computationally creative system that will no longer be merely generative.

Because music has a large rule set – albeit rules that tend to have been agreed upon *after* the creative acts – some initial success has been achieved by directly codifying rules (e.g. Ebcioğlu 1988), or learning them through analysis (Conklin and Witten 1995). MuMe researchers do have access to large databases of symbolic music representations[2] which may produce further success in this direction; however, the material as it is provides potential use for melodic, harmonic, or rhythmic generation, but little use for structural generation, as such analysis has not yet been

[1] http://musicalmetacreation.org/musebots/videos/

[2] see http://metacreation.net/corpus-1/ for a list of such corpora.

automated, despite promising beginnings (Kuhl and Jensen 2008).

The use of aesthetic agents within music has been proposed previously (Spector and Alpern 1994; Pearce and Wiggins 2001; Collins 2006, Galanter 2012), and their complexities noted. However, the higher one rises in the musical hierarchy (i.e. toward generation of complete musical compositions), the more one relies upon aesthetic judgment: it becomes more difficult, if not impossible, to evaluate creative output, since there are no optimal solutions in these cases (Pasquier *et al.* 2016) rendering even the judgment of relative suitability awkward.

Collins' *Autocousmatic* (2012) uses critical agents within an algorithmic compositional system. Complete fixed media works are generated based upon rules derived using machine-learning algorithms trained on exemplar works of acousmatic music. Formal aspects are derived from the database of works, using a top-down method. Several versions of a work are generated with varying surface details; the agents then analyse the candidate generations, comparing them to a single exemplar work, and the best version is selected.

When the completed generated works were evaluated by human composers, a recurring criticism centered around "problems of structure", "structural designs", and "issues... to do with larger forms". Firstly, one must acknowledge that the critical agents are unable to derive enough high-level knowledge from low-level feature analysis, so the perceived formal limitations are understandable; however, there are more fundamental issues in play. The agents are only able to compare the generative material to existing examples, rather than any intentionality of the compositional system at any point in the compositional process: as such, musical context for the decisions as they are being made is completely lost. In addition, the top-level architectural structures learned by the system are dislocated from lower level organisations and so the interdependence between hierarchical levels in creating a well-formed musical structure is absent.

Autocousmatic is not a real-time system, so the opportunity for selection – albeit completely automated selection – from a pool of extant generations exists. Performance systems, and those concerned with dynamic situations, eliminate such possibilities. Even when "big data" approaches, such as those of *Autocousmatic* and Collins' more recent work (2016), become conceivable in real-time, it is doubtful that they will solve the issue of dynamic generative structure for musical CC.

While we argue for the continued necessity of bringing artistic domain-specific knowledge to bear on any successful generative system – especially those that attempt to generate formal structures – we acknowledge the open problem of how such structures can be created dynamically through computational means.

Beyond Mere Generation - New Directions

We recognise the potential for the use of machine-learning to build aesthetic-agents in the real-time evaluation of generative music, with the understanding that they will require domain-specific knowledge in their assignment. We propose building on Collins' approach, with agents trained on specific corpora of exemplar music, Kramer's notion of discontinuities as form-defining elements, and recent research in musebot communications to express intentions and goals.

Musebots, described earlier, allow designers to create autonomous musical agents that interact in a collaborative ensemble with other musebots, potentially created by other designers. A particularly exciting aspect involves the notion that developers must decide *how* the musebots should interact, and *what* information is necessary to produce meaningful musical interaction. Musebots offer the potential to create complex musical surfaces and structures in which the organisation is emergent rather than attributed to a single clever programmer. Concepts of formal design have been raised already: initial musebot ensembles followed either a self-organising model, or a reactive model in which one musebot "took the lead" in determining sectional change. They have thus far avoided the requirement of large-scale formal structures by limiting their performances to five to seven minute compositions.

Musebots communicate their current states and, potentially, their intentions; however, as with all creative acts, intentions are not always achieved. Having dedicated musebots actively *listening* to music as it is being generated would allow for aesthetic decisions to be made as to when formal changes will need to be made, thus exemplifying a bottom-up perspective informed by high-level knowledge. These agents could be trained on specific styles, using standard MIR feature analysis (Tzanetakis and Cook 2000), having learned *why* formal changes occurred in the corpus. The example music would be hand-annotated by experts – rather than relying upon inexact machine analysis – at points of structural and formal change. The agents could learn to recognise a discontinuity – using models proposed by Parks (1989), for example – as well as examining the musical features prior to this break. How long is unvarying continuity acceptable until change is required? Or, at what point is boredom about to be felt by the listening agent (Eigenfeldt 2014)? This knowledge could then be used during generation, allowing the musebots to produce material using current methodologies (e.g. Eigenfeldt, Bown, and Carey 2015), while the listening agent could suggest when structural changes need to occur.

However, this model requires careful selection in determining the specific corpus, and locations within that corpus, on which the listening agents would be trained for the specific generation desired. There is no universal standard pertaining to musical form; how much repetition and variation is preferable in electronic dance music is significantly different than in free improvised music, or Debussy, for example.

Large-scale structure could be instantiated through the use of shape-negotiation; negotiation has already proven to be a useful method of organising musical agents (Eigenfeldt 2010). These shapes could be applied to a variety of musical structures over the course of a performance. Musebots have already been created that react to high-level attributes of valence and arousal (Eigenfeldt *et al*. 2015). Individual agents can ignore, agree, or offer alternatives to the formal contours; rather than having a single agent issuing predetermined orders (or following the directions of a human operator), these shapes can be proposed, accepted, and altered by prescient musebots.

An important aspect in terms of the dynamic generation of longer-term temporal structures would be the potential for the shapes themselves to be modified by the agents in real-time. Rather than interpreting these shapes directly, breakpoints could be assumed to be individual goals at proposed formal divisions, with the agents determining individual trajectories toward these agreed upon goals. These divisions could be considered suggestions, and the bottom-up listening agents could provoke revisions to these breakpoints. We recognise this as being a form of dynamic time warping (Keogh and Ratanamahatana 2005).

This multi-agent approach also allows for the maintenance of alternative interpretations and corresponding generative options. It provides flexibility to changing circumstances required by dynamic and interactive systems such as interactive games or improvised music performance systems. This approach has been used for generative rhythms and formalised as the Chimera Architecture that simultaneously tracks a collection of viable scenarios for musical continuation (Gifford and Brown 2009).

To summarise, convincing musical intelligence involves coordinating and solving many micro-problems in order to achieve musical *coherence*. Generating sequences of events – whether low-level melodic shapes or high-level formal outlines – is a more elementary task of simple generation. We can use corpus-based strategies involving statistical or rule-based learning and endlessly generate content that sounds similar to other content, but when we attempt to insert originality – for example, in combinatorial creativity, by combining one kind of melodic style with an unrelated song structure, and trying to make these things 'fit' – we encounter problems of coherence that aren't necessarily answered by looking at the corpus. We posit that these problems may inherently require forms of evaluation that take them into a domain beyond Veale's "mere generation" and into computational creativity proper, where the only way to determine the value of an output is through its analysis.

Much MuMe research has already looked at whether this is indeed the case; for example, Blackwell and Young's swarm/self-organisation approach (2004) looks to see how far non-evaluative structuring processes can be taken. The problem remains open, and we suggest that MuMe researchers should continue to pursue generative approaches to complex structure, following either the scientific tradition of attempting to create autonomous systems that implement a theoretical hypothesis, or in the artistic tradition building interactive systems that attempt musical coherence with a human performer.

We feel that the musicology of Parks and others provides a strong starting point to these investigations and is a productive way forward. It is grounded in a level of analysis that is sufficiently abstract to apply to all music. What is potentially of great interest here is that it affords a tie-in with the kinds of linguistic reasoning that is present in other areas of computational creativity (Perez and Sharples 2001; Veale 2012). If we begin to think of systems that form their own concepts of musical structure, then we can imagine them building a language from which a logic emerges. This logic would define the coherence of the music, and could have generative potential through metaphors and other linguistic constructs. It would be a mid-level language, meaning not at the musical surface, but also not necessarily at the level of our actual use of language (i.e., a mentalese representation). It would also be highly subjective, adaptive to the individual's own experience, just as statistical learning approaches are, but very different to statistical learning in terms of generative process – it would involve analytical problem solving in an iterative generate-and-test cycle. This implies an approach where we would ask, for any given musical form, or corpus of musical forms: can a non-trivial conceptual language be constructed for which this music is coherent? Or given a set of such solutions, what are the generative properties from which coherent music can emerge?

References

Berry, W. 1966. *Form in Music* (Vol. 1). Prentice-Hall.

Beyls, P. 2007. Interaction and Self-organisation in a Society of Musical Agents. *ECAL Workshop on Music and Artificial Life*, Lisbon.

Blackwell, T., Young, M. 2004. Self-organised music. *Organised Sound*, 9(02), 123–136.

Blackwell, T., Bown, O., Young, M. 2012. Live Algorithms: towards autonomous computer improvisers. *Computers and Creativity,* Springer Berlin, 147–174.

Boulez, P., Noakes, D., Jacobs, P. 1964. Alea. *Perspectives of New Music* 3(1), 42–53.

Bown, O., Britton, S. 2013. Methods for the Flexible Parameterisation of Musical Material in Ableton Live. *Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE)*, Boston, 24–26.

Bown, O., Eigenfeldt, A., Pasquier, P., Martin, A., Carey, B. 2013. The Musical Metacreation Weekend: Challenges Arising from the Live Presentation of Musically Metacreative Systems. *AIIDE*, Boston, 27–34.

Bown, O., Carey, B., Eigenfeldt, A. 2015. Manifesto for a Musebot Ensemble: A platform for live interactive performance between multiple autonomous musical agents. *International Symposium of Electronic Art*, Vancouver.

Caplin, W. 1998. *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*. New York: Oxford University Press.

Collins, K. 2008. *Game Sound: An Introduction to the History, Theory and Practice of Video Game Music and Sound Design.* MIT Press.

Collins, N. 2006. Towards Autonomous Agents for Live Computer Music: Realtime Machine Listening and Interactive Music Systems. *Doctoral dissertation*, Cambridge.

Collins, N. 2009. Musical Form and Algorithmic Composition. *Contemporary Music Review*, 28(1), 103–114.

Collins, N. 2012. Automatic Composition of Electroacoustic Art Music Utilizing Machine Listening. *Computer Music Journal*, 36(3), 8–23.

Collins, N. 2016. Towards Machine Musicians Who Have Listened to More Music Than Us. *Musical Metacreation*, ACM Computers In Entertainment, forthcoming.

Colton, S., Wiggins, G. 2012. Computational creativity: The final frontier? *European Conference on Artificial Intelligence*, Montpellier, 21–26.

Compton, K., Mateas, M., 2015. Casual Creators. *Proceedings of the Sixth International Conference on Computational Creativity,* Park City, 228.

Conklin, D., Witten, I. 1995. Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 24(1), 51–73.

Conklin, D. 2003. Music Generation from Statistical Models. *AISB 2003 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences*, Aberystwyth, 30–35.

Cope, D. 1996. *Experiments in Musical Intelligence.* A-R Editions.

Cope, D. 2000. *The Algorithmic Composer*. A-R Editions.

d'Escrivan, J. 2014. Videogames as digital audiovisual performance. *Conference on Computation, Communication, Aesthetics and X*, Porto, 248–259.

Dahlhaus, C. 1989. *Schoenberg and the New Music: Essays by Carl Dahlhaus*. Cambridge University Press.

Dahlstedt, P., McBurney, P. 2006. Musical agents: toward computer-aided music composition using autonomous software agents. *Leonardo*, 39(5), 469–470.

Deutsch, D. 2013. *Psychology of music*. Elsevier.

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

Eigenfeldt, A. 2006. Kinetic Engine: Toward an Intelligent Improvising Instrument. *Sound and Music Computing Conference (SMC)*, Marseille, 97–100.

Eigenfeldt, A. 2010. Coming Together - Negotiated Content by Multi-agents. *ACM Multimedia International Conference*, Firenze, 1433–1436.

Eigenfeldt, A. 2013. Generating Electronica - A Virtual Producer and Virtual DJ. *ACM Creativity and Cognition*, Sydney.

Eigenfeldt, A. 2014. Generating Structure – Towards Large-scale Formal Generation. *AIIDE*, Raleigh, 40-47.

Eigenfeldt, A., Bown, O., Carey, B. 2015. Collaborative Composition with Creative Systems: Reflections on the First Musebot Ensemble. *Proceedings of the Sixth International Conference on Computational Creativity,* 134–141.

Eigenfeldt, A., Bizzocchi, J., Thorogood, M., Bizzocchi, J. 2015. Applying Valence and Arousal Values to a Unified Video, Music, and Sound Generative Multimedia Work. *Generative Art Conference (GA)*, Venice.

Eno, B. 1996. Generative Music. http://www.inmotionmagazine.com/eno1.html, accessed Feb. 1, 2016.

Farbood, M. 2012. A parametric, temporal model of musical tension. *Music Perception*, 29(4), 387–428.

Galanter, P. 2003. What is Generative Art? Complexity theory as a context for art theory. *GA*, Milan.

Galanter, P. 2008. Complexism and the role of evolutionary art. *The Art of Artificial Evolution*, Springer, 311–332.

Galanter, P. 2012. Computational aesthetic evaluation: past and future. *Computers and Creativity*, Springer, 255–293.

Gifford, T., Brown, A. 2009. Do Androids Dream of Electric Chimera? *Improvise: The Australasian Computer Music Conference*, Brisbane, 56–63.

Gifford, T., Brown, A. 2011. Beyond Reflexivity: Mediating between Imitative and Intelligent Action in an Interactive Music System. *BCS Conference on Human-Computer Interaction*, Newcastle Upon Tyne.

Hamanaka, M., Hirata, K., Tojo, S. 2006. Implementing "A Generative Theory of Tonal Music". *Journal of New Music Research*, 35(4), 249–277.

Hedges, S. 1978. Dice music in the eighteenth century. *Music & Letters*, 180–187.

Hill, M. 2011. Beneath Clouds and The Boys: jazz artists making film music. *Screen Sound n2*, 27–47.

Hiller, L. 1970. Music Composed with Computers: An Historical Survey. H. B. Lincoln, ed. *The Computer and Music*. Cornell University Press, 42–96.

Hindemith, P. 1970. *The Craft of Musical Composition: Theoretical Part 1*. Schott.

Hoover, A., Szerlip, P., Stanley, K. 2014. Functional Scaffolding for Composing Additional Musical Voices. *Computer Music Journal*, 38(4), 80–99.

Huron, D. 2006. *Sweet anticipation: Music and the psychology of expectation*. MIT Press.

Keogh, E., Ratanamahatana, C. 2005. Exact indexing of dynamic time warping. *Knowledge and information systems*, 7(3), 358–386.

Kramer, J. 1978. Moment form in twentieth century music. *The Musical Quarterly*, 64(2), 177–194.

Kramer, J. 1988. *The Time of Music New Meanings, New Temporalities, New Listening Strategies*. Schirmer Books.

Kühl, O., Jensen, K. 2007. Retrieving and Recreating Musical Form. *Computer Music Modeling and Retrieval. Sense of Sounds.* Springer Berlin, 263–275.

Lerdahl, F., Jackendoff, R. 1983. *A Generative Theory of Tonal Music*. MIT Press.

Lester, J. 1986. Notated and Heard Meter. *Perspectives of New Music*, 24(2), 116–128.

Lewis, G. 1999. Interacting with Latter-Day Musical Automata. *Contemporary Music Review* 18(3), 99–112.

Livingstone, S., Muhlberger, R., Brown, A., Thompson, W. 2010. Changing Musical Emotion: A Computational Rule System for Modifying Score and Performance. *Computer Music Journal*, 34(1), 41–64.

Maconie, R. 1976. *The Works of Karlheinz Stockhausen*. Oxford University Press.

Meyer, L. 1956. *Emotion and Meaning In Music.* University of Chicago Press.

Miranda, E. 2001. *Composing music with computers*. CRC Press.

Narmour, E. 1990. *The Analysis and Cognition of Basic Melodic Structures*. University of Chicago.

Narmour, E. 1991. The Top-down and Bottom-up Systems of Musical Implication: Building on Meyer's Theory of Emotional Syntax. *Music Perception: An Interdisciplinary Journal*, 9(1), 1–26.

Nattiez, J. 1990. *Music and discourse: Toward a semiology of music*. Princeton University Press.

Negretto, E. 2012. *Expectation and anticipation as key elements for the constitution of meaning in music.* Teorema 31(3), 149–163.

Nierhaus, G. 2009. *Algorithmic composition: paradigms of automated music generation*. Springer Science & Business Media.

Nyman, M. 1999. *Experimental music: Cage and beyond*. Cambridge University Press.

Pachet, F. 2004. Beyond the cybernetic jam fantasy: The continuator. *Computer Graphics and Applications*, IEEE, 24(1), 31–35.

Parks, R. 1989. *The Music of Claude Debussy*. Yale University Press.

Pasquier, P., Eigenfeldt, A., Bown, O., Dubnov, S. 2016. An Introduction to Musical Metacreation. *Special Issue: Musical Metacreation,* ACM Computers In Entertainment, forthcoming.

Pearce, M., Wiggins, G. 2001. Towards a framework for the evaluation of machine compositions. *Artificial Intelligence and Creativity in the Arts and Sciences,* 22–32.

Pearce, M., Wiggins, G. 2007. Evaluating cognitive models of musical composition. *International Joint Workshop on Computational Creativity.* University of London, 73–80.

Pérez, R., Sharples, M. 2001. MEXICA: A computer model of a cognitive account of creative writing. *Journal of Experimental & Theoretical Artificial Intelligence*, 13(2), 119–139.

Prokopenko, M. (2009). Guided Self-Organization. *HFSP Journal,* 3(5), 287–289.

Rowe, R. 1992. *Interactive music systems: machine listening and composing*. MIT Press.

Rowe, R. 2004. *Machine musicianship*. MIT press.

Salzer, F. 1962. *Structural hearing: Tonal Coherence in Music*. Courier Corporation.

Schedel, M., Rootberg, A. 2009. Generative Techniques in Hypermedia Performance. *Contemporary Music Review*, 28(1), 57–73.

Schenker, H. 1972. *Harmony*. University of Chicago Press.

Schoenberg, A., Stein, L. 1970. *Fundamentals of musical composition*. Faber & Faber.

Shannon, C., Weaver, W. 1949. *The mathematical theory of communication.* University of Illinois Press.

Sloboda, J. 1991. Music structure and emotional response. *Psychology of Music*, 19, 110–120.

Smith, B., Garnett, G. 2012. Improvising Musical Structure with Hierarchical Neural Nets. *AIIDE*, Palo Alto, 63–67.

Sorensen, A., Brown, A. 2008. A Computational Model For The Generation Of Orchestral Music In The Germanic Symphonic Tradition: A progress report. *Australasian Computer Music Conference*, Sydney, 78–84.

Spector, L., Alpern, A. 1994. Criticism, culture, and the automatic generation of artworks. *National Conference on Artificial Intelligence*, Seattle, 3–8.

Stockhausen, K. 1963. Momentform: Neue Beziehungen zwischen Aufführungsdauer, Werkdauer und Moment. *Texte zur Musik 1*, 189–210.

Suurpää, L. 2006. Form, structure, and musical drama in two Mozart expositions. *Journal of Music Theory* 50(2), 181–210.

Tzanetakis, G., Cook, P. 2000. Marsyas: A framework for audio analysis. *Organised Sound*, 4(03), 169–175.

Veale, T. 2012. *Exploding the creativity myth: The computational foundations of linguistic creativity*. A&C Black.

Veale, T. 2015. Scoffing at Mere Generation. http://prosecco-network.eu/blog/scoffing-mere-generation, accessed Feb. 1, 2016.

Whittall, A. 2016. Form. *Grove Music Online. Oxford Music Online.* Oxford University Press, accessed Feb. 1, 2016.

Xenakis, I. 1992. *Formalized music: thought and mathematics in composition* (No. 6). Pendragon Press.