Nelson Zagalo & Sandra Oliveira (2014)
Abordagens da Narrativa nos Media
Centro de Estudos de Comunicação e Sociedade, Universidade do Minho · ISBN 978-989-8600-31-8
pp. 123 -134

Beyond Music Information Retrieval: A Proposed Model for Automatic Generation of Dialogic Music

Para além da Recuperação de Informação Musical: Proposta de um Modelo de Geração Automática de Narrativas Musicais Dialógicas

Gilberto Bernardes & Matthew E. P. Davies

gba@inesctec.pt; mdavies@inesctec.pt
INESC TEC Sound and Music Computing Group

Abstract
In this article we propose a generative music model that recombines heterogeneous corpora of audio units on both horizontal and vertical dimensions of musical structure. In detail, we describe a system that relies on algorithmic strategies from the field of music information retrieval—in particular content-based audio processing strategies—to infer information from music recordings that, in turn, supervise generative music strategies. The model allows automatic remix and mashup creation and the manipulation of audio signals according to inferred stylistic features. In addition to widening the creative potential of non-experts, the model also expands knowledge in areas such as computational music analysis, generative music and music information retrieval.

Keywords: Generative music; remix; mashup; music information retrieval

Resumo
Neste artigo propomos um modelo de geração de música que combina conjuntos heterogéneos de unidades sonoras em ambas as dimensões horizontal e vertical da estrutura musical. Em específico, descrevemos um sistema que assenta em estratégias algorítmicas da área da recuperação de informação musical—em particular estratégias de processamento baseadas no conteúdo de sinais de áudio—para inferir informação de música pré-gravada que, por sua vez, supervisiona estratégias de geração de música. O modelo proposto permite criar automaticamente *remixes* e *mashups*, assim como manipular sinais sonoros de acordo com características estilísticas inferidas de uma base de músicas. Para além de possibilitar públicos não-especialistas de experienciar o domínio da criação musical, o modelo também expande conhecimento em áreas como análise computacional de música, música generativa e recuperação de informação musical.

Palavras-cheve: Música generativa; *remix*; *mashup*; recuperação de informação musical

Introduction

One of the most distinctive principles of postmodern art is the appropriation[1] of pre-existing "objects" with little or no transformation in the creation of new artworks (Chilvers & Glaves-Smith, 2009). Despite its extension through several artistic domains (e.g., visual arts: Sherrie Levine e Barbara Kruger; design: Philippe Starck; and literature: the Beat Generation movement), the practice of appropriation has its roots within music performance and composition (Brewster & Broughton, 2000). In music, appropriation techniques gained an increased popularity in the late-1960s and early-1970s due to their widespread adoption by pop artists (ibidem)—in particular through practices such as the remix, mashup,2 and cover versions. However, in music, the practice of appropriation is not exclusive to postmodern times, in fact, it dates back to 12th century polyphony (Griffiths, 1981; Burkholder, 1983).

More recently, the music appropriation phenomenon, in particular the approach used by pop artists, has captured the attention of researchers and software developers, who aim to design tools for automatic remix and mashup creation (Jehan, 2005; Collins, 2006; Bernardes, Guedes, & Pennycook, 2013; Davies *et al.*, 2013; Eigenfeldt, 2013).[3] The growing interest in the development of tools for automatic remix/ mashup creation is due to aesthetic and technological motivations. As Eigenfeldt (2013) claims, the use of a musical corpus as the basis of a postmodern aesthetic seems to have replaced the concept of randomness, which has a huge expression in modern music. This aesthetic shift was only possible due to recent technological advances, in particular research in MIR, such as beat and tempo detection, structural segmentation, and audio signal segregation, which offer a good basis for the creative manipulation of audio signals.

Our recent work has been at the intersection of the two above-mentioned fields—MIR and automatic music generation—and our most mature contribution is the software earGram (Bernardes *et al.*, 2013; Bernardes, 2014), whose architecture will be used to propose a new generative music model. The model we propose allows the "stylistic" manipulation of songs and/or automatic remix and mashup creation. Special attention will be given to the architecture and technical implementation of the model, without neglecting the motivation, context of emergence, catalysts, and pertinence of the proposal.

Besides this introduction, this article has three more sections in which we start by describing the state-of-the-art of generative music systems that embed learning capabilities (§ 2), followed by the socio-economical, cultural, and technological contexts that allow the emergence of a new generative music model (§ 3). Then, we

[1] In this article, the term appropriation refers to the practice of reusing an (existing) object in a different work. In music, this practice is also referred to as quotation, collage, borrowing, and sampling.

[2] Paradis (2007) claims that *mashup* is a manufactured buzzword with a vague meaning. I would add that Paradis' assertion might be extended to the concept of remix. Therefore, it is important to clarify that remix and mashup are understood here as musical pieces that result from the recombination of one or more songs on the horizontal or vertical dimensions, respectively.

[3] Automatic mashup creation was recently highlighted as one of the "grand challenges" of MIR (Goto, 2012).

present earGram, the system at the core of the architecture of the proposed system (§ 4), and, finally, we conclude by presenting a discussion about the possibilities and limits of the proposed model (§ 5).

1. Inferring Musical Structure as a Strategy to Supervise Generative Algorithms

Music has a long history of using generative (or algorithmic) strategies for composition. However, only in the mid-1950s the field started to raise the interest of the academic, scientific, and artistic communities by incorporating the processing power of modern computers in its practice (Hiller & Isaacson, 1957). In this article, we will limit our scope of action to computational algorithmic composition, and within this domain, we will give special attention to systems that learn from pre--existing musical examples to (re)create new musical structures.

Instead of developing generative systems that create music according to a set of previously encoded rules, today, we can make use of large-scale musical databases easily available on the Internet to infer models and rules to supervise generative music algorithms. That is, we start witnessing a predisposition to include listening, learning, and composition competences in generative music systems, as well as witnessing a gradual shift in the design paradigm of interactive music systems from an instrumental model (reactive) to a performative model (interactive).

Having constrained our object of study to generative music systems with learning capabilities, and before entering in a comprehensive description of related state-of-the-art work, we need to distinguish two possible approaches within this area. Their main difference relies on the type of representations used: either (i) symbolic representations of musical events, such as musical scores and MIDI information, or (ii) audio signals, which encode the auditory experience or performance. Even though our work focuses on the second category, the first is still an inspiration to us. Currently, we can extract different information from both representations, which demands the adoption of distinct processing strategies.[4]

Various generative music systems that embed listening and learning competences and deal with symbolic representations have been presented for the last two decades. Papadopoulos and Wiggins (1999) offer a comprehensive review of these systems and identify their main limitation as the lack of relevant artistic results, in particular due to the poor macrostructural organization of the generated music. A clear exception is the work of Cope (1996), namely his experiments in musical intelligence (EMI). Cope created computer programs that compose complete works in the style of various classical composers by recombining "stylistic signatures" (fingerprints) inferred from large-scale musical databases.

Computational generative music systems that recycle audio signals based upon learnt models of the signal's structure are a recent research topic with scant,

[4] For a comprehensive review of the various levels of music information that can be extracted and processed from both symbolic music representations and audio signals please refer to Vinet (2004).

but promising results. Within this approach, two of the most fertile topics are automatic remix and mashup creation (Jehan, 2005; Collins, 2006; Davies *et al*., 2013, Bernardes, 2014) and soundscape generation (Hoskinson & Pai, 2001; Strobl *et al*., 2006; Grill, 2010). Our work focuses mainly on the first task—automatic remix and mashup creation—and we expand current knowledge by studying an approach to sound unit recombination based on their rhythmic and harmonic compatibility on both the vertical and horizontal dimensions, rather than inspecting the presence of high similarity between sound events, thus offering a broader and richer range of musical possibilities. Additionally, despite recent efforts, so far, results focus almost exclusively on simple harmonic models, whose matching criteria happen in chroma space (i.e., 12 dimensions) that does not address spectral/timbral properties.

2. Historical Perspective and Emergent Context of a New Generative Music Model: From the *Hits* to Recommendation to Participation

The mass media culture established a global scale economic model built upon mechanical reproduction, blockbusters, and radio and television primetime. In sum, a culture of *hits*. Today, the ever-increasing storage and streaming capabilities of computers and networks are changing the established mass media culture. The emergent model retains economic exchange on a global scale, but focuses on rather smaller niches, with ideological and cultural affinities.

We shall examine the premises and effects of this new model in music. At the turn of the 21st century, portable digital players drastically increased the amount of music that we could own and easily listen to everywhere. What we hear has ceased to be dictated by the broadcast signal and even less by the limited shelf space of the record stores. Simultaneous to the possibility to carry with us the equivalent to a record store of the 1990s, another important change that has been happening since the late-1990s and early-2000s has been the new marketing strategies of audiovisual digital content on the Internet, made popular by platforms such as Napster,[5] Rhapsody,[6] Amazon MP3,[7] or the Apple iTunes.[8] These Internet services give access to huge amounts of musical content with reduced cost in comparison to traditional physical formats (e.g., CD or DVD).[9] One last phenomenon that cannot remain unmentioned here, due to its contribution to the emergence of a new economic model, was the advent of peer-to-peer (P2P) networks that not only allowed unlimited data sharing between users, but also instigated piracy.[10] Yet, in today's new economy, the hits have not disappeared, but they clearly "compete with an infinite number of niche markets, of any size" (Anderson, 2008: 5).

[5] Naspster retrived from http://www.napster.com, date accessed 12/05/14.
[6] Rhapsody retrived from http://www.rhapsody.com, date accessed 12/05/14.
[7] Amazon MP3 retrived from http://www.amazon.com/b?node=163856011, date accessed 12/05/14.
[8] iTunes retrived from https://itunes.apple.com, date accessed 12/05/14.
[9] According to Jehan (2005), in 2005, the number of worldwide digital music titles was estimated in 10 million.
[10] Chirs Anderson (2008) claims that in 2006 an average of 10 million Internet users shared information on P2P networks.

The easy access to huge amounts of music on the Internet created the demand for better retrieval mechanisms to navigate this vast search space, and effective recommendation strategies that could predict and anticipate user preferences in order to encourage people to discover and buy new music. Last.fm[11] and Pandora[12] are two Internet music services that make use of very refined algorithms to navigate large-scale music databases and learn user preferences in order to provide reliable recommendations. Recently, this same technology started to interest artists and music technology researchers, because its analytical power provides good data to support intelligent systems for generative music.

This new economic model also changed the roles of its agents, in particular the traditional notion of producer and user. In the mass media culture, both producers and users had very clear roles and were represented by different entities. Today, people have access to a large number of (free) multimedia production tools and channels to easily share their work on the Internet services like Youtube[13] or SoundCloud.[14] This, in turn, has exponentially increased the participative role of people that were consigned in the past to being consumers. Today, both amateur and professional musicians have almost equal access to the market, and the distinction between producers and users, so evident in mass media culture, has been radically shortened. There is even the need to create a new concept to designate this new agent that merges both the producer and the user in a single entity or person, which Bruns (2007) names *produsage*.

To summarize, (i) the large amount of data available in private and public databases, (ii) the easy access to production tools and dissemination channels for cultural goods, (iii) the recent efforts devoted to the development of creative tools for non-experts, (iv) the new phenomenon of *produsage*, and (v) the possibilities offered by recent technology to manipulate musical data enable us to envision a new model for automatic music generation, which we now describe.

3. A Proposed Model for the Automatic Generation of Dialogic Music

The system that we propose is a software that generates music automatically by recombining short segments of audio signals that have been previously analyzed and annotated. The generative algorithms are supervised by information inferred by machine learning algorithms from (i) a single audio track, (ii) several musical examples, or even (iii) meta-data and audio analysis data retrieved from large-scale Internet musical databases. Despite the focus on machine learning strategies, the system is not limited to the replication of musical style, but rather to the exploration of new compositional strategies that depart from models inferred from user-given

[11] Last.fm retrived from http://www.last.fm, date accessed 06/04/14.
[12] Pandora retrived from http://www.pandora.com, date accessed 06/04/14.
[13] YouTube retrived from https://www.youtube.com/, date accessed 12/08/14.
[14] SoundCloud retrived from https://soundcloud.com, date accessed 21/05/14.

musical examples. The inferred models not only automate several parameters of the generated music, but also allow for the easy and fast exploration of unknown musical spaces. The presented strategies rely on and extend the software earGram, which was previously developed by the authors and whose description follows.

3.1 EarGram: The Basis of a Generative Dialogic Musical Model

Eargram (Bernardes *et al.*, 2013) is an analysis-synthesis software for the creative exploration of large databases of audio snippets. Its architecture relies on concatenative sound synthesis (CSS),[15] a recurring research topic since the early 2000s (Schwarz, 2000; Zils & Pachet, 2001). In brief, CSS uses a large corpus of segmented and descriptor-analyzed sounds snippets, called units, and a unit selection algorithm that finds the best matching units from the corpus to assemble a target phrase according to a similarity measure in the descriptor space. The unit selection is done according to instructions given to the system, commonly as collection of content-based audio descriptions. Despite technological advances in the last decade, CSS still has limitations that prevent musicians with a traditional music education background to use the technique.[16]

The architecture of earGram relies on a CSS algorithm, but extends it by reducing the number of parameters that need to be defined by the user, and allows its specification in an intuitive manner through the adoption of musical concepts—thus increasing the usability for musicians. Of interest here is earGram's self-referential generative strategies, which create "variations" of a given song according to the audio source structure. In order to do that, earGram integrates listening and unsupervised learning capabilities—that infer musical structure from a user-assigned audio file—with composition algorithms, which generate sound mosaics by concatenating audio units. The results resemble, but are not limited to, practices like remix and mashup creation.

EarGram is composed of three modules: (i) analysis, (ii) composition, and (iii) music theory. The analytical module reveals information about several temporal scales of musical structure in an unsupervised, bottom-up (from the sample level to the macrostructure), and recursive manner. EarGram's analytical strategies adopt a reduced listening methodology (Schaeffer, 1966), which integrates content-based audio processing, data mining, machine learning, psychoacoustic dissonant models and music theory. The composition module manipulates the information inferred during analysis in an inverse fashion, i.e., in a top-down strategy, from the (manual) definition of sub-corpora that determine the material used in the various sub-sections of a piece (macrostructure) to the automatic selection of phrases and sound units. The third and last earGram module—music theory—feeds the system

[15] For more information, sound examples and to download the open-source and freely available software earGram please visit the following website: https://sites.google.com/site/eargram/.

[16] For a detailed description of the limits of CSC please refer to Bernardes (2014).

with a set of rules that support the generative music algorithms. From the possible applications of the system, we would like to highlight the possibility to change the underlying metric structure of a song, by infinitely extending an audio sample without repeating or looping, while retaining the structure of the original file, and the possibility to recombine audio units on both horizontal and vertical dimensions.

EarGram establishes a framework for music creation with the potential for recycling extensive databases; however, its current implementation has some limitations on the amount of data it can process, mainly due to its implementation in Pure Data. Additionally, adopting a more effective unit selection algorithm, in particular by mining the corpus prior to the retrieval task, would significantly reduce computational costs. We also believe that better and higher-level descriptions of the audio units would expand and improve the recombination strategies of the system.

3.2 Towards a New Model

In order to extend earGram's architecture towards the manipulation of larger musical databases, as well as reducing the number of user-specifications, we need to restructure the system's architecture. On the one hand we need expand earGram's architecture with two new modules—music critic and performer—and on the other hand we need to adapt the remaining reaming modules—analysis, composition, and music theory—to this new framework (see Figure 1).

Figure 1: Algorithmic chain of the proposed generative music model

We will start by detailing the changes to undertake in the existing modules. The overall structure of the analysis module remains unchanged, with the exception of the extended possibility to process meta-data and audio analysis data retrieved from Internet databases. The analysis module is responsible for inferring models (e.g., histograms and *n*-grams), patterns (e.g., typical harmonic progressions and common rhythmic patterns) and additional mid-level information of the audio source(s) (e.g., macrostructure, tempo, metrical structure, and key). The inferred models encode structural trends of the corpus—a collection of audio units gathered from the segmentation of a song or a set of songs with some common denominator, such as: style, composer, epoch, etc. The main difference in relation to the current version of earGram is the possibility to process musical data retrieved from Internet databases, such as Magnatagatune (Law & Ahn, 2009) and the Million Song Dataset (Bertin-Mahieux, Ellis, Whitman & Lamere, 2011), which provide a good set of audio features for music protected by copyright. The resulting data is then formatted and sent to the music theory module.

In earGram, the set of rules that support the generative music algorithms are models inferred from the audio source in combination with music theory principles that were embedded in the original source code. In the new model, the music theory module is radically different because it does not have any knowledge a priori; instead, it only consists of a storage pool that processes and saves the data it receives form the analytical module, i.e. the music theory module is seen as a blank repository of musical models. The more data it gets the better it refines and improves its models.

The composition module uses the generative algorithm previously developed in earGram to process the information stored and manipulated in the music theory module, interpreting it as ranges of possibilities for particular parameters. These ranges can be seen as constraints that are applied to the search space (i.e., corpus) during generation. A strategy to minimize the lack of structure in the meso and macro times scales of musical structure will be further researched, because, contrary to earGram, the proposed system does not need to work or select the units for concatenation in real-time and can plan the structure of the generated audio on a larger scale. The proposed system will focus on the generation of remixes and mashups and the manipulation of audio signals by imposing stylistic features inferred from one corpus on a different corpus, allowing the possibility to work with style hybridization. We would like to highlight that it is not our main objective to emulate the style of particular musical examples; instead, we want to use inferred information from a coherent corpus to explore new and unknown spaces by applying it to different corpora.

The proposed system not only expands knowledge in machine learning strategies applied to musical information, but also establishes a computational music cycle with almost infinite possibilities for generating new music. Even if the current music production methods would cease, the possibility to recycle the actual corpus of digital music would offer endless possibilities, because the system would be able to feed itself with new musical examples. In order to regulate the quality of the generated music of this cyclic system, it is nonetheless necessary to introduce a critic element in the processing chain, which is responsible for evaluating the generated music and consequently validate the generative model(s) applied. Given the complexity and subjectivity of the task, this procedure is done manually by the user. The information given by the user is then used by the music theory module to weight (reinforce or exclude) data from the system.

The last module, performer, receives and processes all the information sent by the composition module in order to play the selected audio units. This module not only deals with the concatenation quality between audio units, but it also responsible for applying several audio effects (e.g., reverb, filters, compression, or morphing) at the end of the processing chain in order to enhance creativity. A final issue that we want to address in this new system in relation to earGram is the rhythmic alignment of overlapping sound events, mainly for mashup creation, which will be consider as a performance issue, and make use of a time-stretching algorithm to regulate the synchronization and/or complementarity between overlapping rhythmic events.

4. A Temporary Discussion: From Large-scale Music Processing Strategies to the Emergence of a Culture of Amateurism

In this article we proposed a generative music computational model for remix and mashup creation or stylistic manipulation of high-level structural features of a song by models inferred from musical analysis data available on the Internet. The technical basis of the system relies on the architecture of earGram (Bernardes *et al.*, 2013), i.e., it is based on an algorithmic chain that recombines content-based audio processing, psychoacoustic and music theory models, unsupervised machine learning techniques and generative strategies to recycle musical data.

The proposed system fits into the historic relationship between music and technology, and takes it further by computationally modeling the life cycle of music creation, i.e., implementing in the same system listening, learning, and composition competences. This framework shows great potential for the development of new tools for algorithmic composition with a high level of sophistication and embedded logic. Modern computers—the principal catalysts of this new framework—are understood here as active agents of the creative process and extend it to levels of sophistication of a performer, shifting the creative paradigm of generative music systems from a reactive (instrumental) model to an interactive (performative) model. Even though computers are far from emulating the skills of an expert musician—we should note that encoding the behaviour and logic of an expert musicians in computer programs is as difficult as understanding the inherent mechanisms of human cognition—computers, better than any human being, can read, learn, store, and manipulate large amounts of data in order to infer higher-level information (patterns) from it. In summary, recent technology in combination with extensive repositories of musical data drive the emergence of new creative models, as well as new means of production, distribution, and reception of musical content and open new paths for more intelligent generative music services on the emergent Web 3.0.

In addition to the contribution of the proposed model to generative music, its machine learning methods expand knowledge in areas such as music information retrieval by extracting information from large-scale musical databases for style-specific manipulation and understanding.

The proposed model primarily targets non-experts by providing them ways to creatively explore new musical spaces along with a substantial increase in their participative role. Booth (1999) notes that by increasing the level of participation, users are encouraged to engage in a more enriching experience. Additionally, our model is on the borderline between a certain degree of familiarity and novelty. As Navas (2012) claims, "when people hear their favorite songs mashed up, it is very likely that they will get excited and find pleasure in recognizing the compositions" (p. 104). In addition to the model's advantages to non-experts, it also promotes the creativity of experts, namely the exploration of non-linear music composition structures. The jump from a fixed (documental) to a non-linear (performative) format by using tools that can manipulate musical data, puts into question a large part of the conceptual and technical basis of music creation.[17]

As Manovich (2013) claims in his most recent thesis, software became the communication interface *par excellence*. In recent years, the vast majority of the communication tools were adopted and encoded as software made available for large-scale Internet users. However, the breakthrough innovations that we witnessed in photography during the last decade are far from being reached in music. In sum, we anticipate that the proposed model may have a series of impacts at the technological, social, and human behavior levels.

Finally, there is also a series of questions and limits to take into account. There are legal issues involved in the (re)use of copyrighted music, which might prevent the implementation of the model, or impose severe constraints to it. This last problem also raises an interesting aesthetic discussion about the ownership, authorship, and authenticity of the generated music. Additionally, it would be interesting to understand how non-linear music formats like the results produced by the detailed model affects the listeners' reception, meaning, and emotion. All these questions will only have a valid answer once the system has been implemented and usability tests have been run.

Acknowledgments

This research has been made possible by the funds of the Media Arts and Technologies project (MAT), NORTE-07-0124-FEDER-000061, financed by the North Portugal Regional Operational Programme (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF), and by national funds, through the Portuguese funding agency, Fundação para a Ciência e a Tecnologia (FCT).

References

Anderson, C. (2008) *The Longer Long Tail: Why the Future of Business is Selling Less of More*, New York: Hyperion.

[17] For a comprehensive review of the technical limitations of non-linear music please refer to Brosbøl (2006).

Bernardes, G., Guedes, C., & Pennycook, B. (2013) "EarGram: An Application for Interactive Exploration of Concatenative Sound Synthesis in Pure Data" in Aramaki, M., Barthet, M., Kronland-Martinet, R., & Ystad, S. (eds.) (2013) *From Sounds to Music and Emotions*, Berlin-Heidelberg: Springer-Verlag, 110-129.

Bernardes, G. (2014) *Creating Music by Selection: Content-based Algorithmic-assisted Audio Composition*, PhD dissertation, University of Porto.

Bertin-Mahieux, T., Ellis, D. P., Whitman, B., & Lamere, P. (2011) "The Million Song Dataset" in *Proceedings of the International Society for Music Information Retrieval Conference*.

Booth, W. C. (1999) *For the Love of It: Amateuring and its Rivals*, Chicago: University of Chicago Press.

Brewster, B. & Broughton, F. (2000) *Last Night a DJ Save My Life*, New York: Grover Press.

Brosbøl, J. (2006) *Non-linear Composition in Popular Music*, unpublished research report, University of Sydney.

Bruns, A. (2007) "Produsage: Towards a Broader Framework for User-led Content Creation" *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition*.

Burkholder, J. P. (1983) *The Evolution of Charles Ives's Music: Aesthetics, Quotation, Technique*, PhD dissertation, University of Chicago.

Chilvers, I. & Glaves-Smith, J. (eds.) (2009) *Dictionary of Modern and Contemporary Art*, Oxford: Oxford University Press.

Collins, N. (2006) "BBCut2: Incorporating Beat Tracking and On-the-fly Event Analysis", *Journal of New Music Research*, *35*(1): 63-70.

Cope, D. (1996) *Experiments in Musical Intelligence*, Madison, WI: AR Editions.

Davies, M. E. P., Hamel, P., Yoshii, K., & Goto, M. (2013) "AutoMashUpper: An Automatic Multi-song Mashup System" *Proceedings of the International Society for Music Information Retrieval Conference*.

Eigenfeldt, A. (2013) "The Human Fingerprint in Machine Generated Music" *Proceedings of xCoAx2013: Computation, Communication, Aesthetics, and X*.

Griffiths, P. (1981) *Modern Music: The Avant-garde Since 1945*. New York: George Braziller.

Grill, T. (2010) "Re-texturing the Sonic Environment" *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound*.

Goto, M. (2012) "Grand Challenges in Music Information Research" in Muller, M., Goto, M. & Schedl, M. (eds.) (2012) *Multimodal Music Processing*, Dagstuhl Publishing, 217-25.

Hiller, L. & Isaacson, L. (1957) "Musical Composition with a High Speed Digital Computer" *Audio Engineering Society Convention 9*.

Hoskinson, R. & Pai, D. (2001) "Manipulation and Resynthesis with Natural Grains" *Proceedings of the International Computer Music Conference*.

Jehan, T. (2005) *Creating Music by Listening*, PhD dissertation, Massachusetts Institute of Technology.

Paradis, O. (2007) "Bashing the Mashup" *Wired*, *15*(4), retrived from http://www.wired.com/wired/archive/15.04/rants.html, date accessed 25/05/14.

Law, E. & Ahn, L. (2009) “Input-agreement: A New Mechanism for Collecting Data Using Human Computation Games” *Proceedings of the 27th International Conference on Human Factors in Computing Systems*.

Manovich, L. (2013) *Software Takes Command*, Bloomsburry Academic.

Navas, E. (2012) *Remix Theory: The Aesthetics of Sampling*, Viena: Springer-Verlag.

Pascal, R. (2001) “Style” in Sadie, S. & Tyrell, J. (eds.) (2004) *The New Grove Dictionary o Music and Musicians,* 2nd ed., *18*:316-321.

Papadopoulos, G. & Wiggins, G. (1999) “AI Methods for Algorithmic Composition: A Survey, a Critical View and Future Prospects” *AISB Symposium on Musical Creativity*.

Roads, C. (2004) *Microsound*, Cambridge, MA: The Massachusetts Institute of Technology Press.

Schaeffer, P. (1966) *Traité des Objets Musicaux*, Paris: Le Seuil.

Schwarz, D. (2000) “A System for Data-driven Concatenative Sound Synthesis” *Proceedings of the International Conference on Digital Audio Effects*.

Strobl, G., Eckel, G., Rocchesso, D., & le Grazie, S. (2006) “Sound Texture Modeling: A Survey” *Proceedings of the Sound and Music Computing Conference*.

Vinet, H. (2004) “The Representation Levels of Music Information” *Computer Music Modeling and Retrieval*, Springer Berlin Heidelberg.

Zils, A., & Pachet, F. (2001) “Musical Mosaicing” *Proceedings of the International Conference on Digital Audio Effects*.