University of Huddersfield Repository

Jan, Steven

Computer Simulation of Musical Evolution: A Lesson from Whales

Original Citation


This version is available at http://eprints.hud.ac.uk/id/eprint/24946/

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

http://eprints.hud.ac.uk/
Computer Simulation of Musical Evolution: A Lesson from Whales

Steven Jan

June 29, 2015

Abstract

Simulating musical creativity using computers needs more than the ability to devise elegant computational implementations of sophisticated algorithms. It requires, firstly, an understanding of what phenomena might be regarded as music; and, secondly, an understanding of the nature of such phenomena — including their evolutionary history, their recursive-hierarchic structure, and the mechanisms by which they are transmitted within cultural groups. To understand these issues it is fruitful to compare human music, and indeed human language, with analogous phenomena in other areas of the animal kingdom. Whale song, specifically that of the humpback (*Megaptera novaeangeli*ae), possesses many structural and functional similarities to human music (as do certain types of birdsong). Using a memetic perspective, this paper compares the “musilanguage” of humpbacks with the music of humans, and aims to identify a number of shared characteristics. A consequence of nature and nurture, these commonalities appear to arise partly from certain constraints of perception and cognition (and thus they determine an aspect of the environment within which the “musemes” (musical memes) constituting whale vocalizations and human music is replicated), and partly from the social-emotive-embodied and sexual-selective nature of musemic transmission. The paper argues that Universal-Darwinian forces give rise to uniformities of structure in phenomena we might regard as “music”, irrespective of the animal group — certain primates, cetaceans or birds — within which it occurs. It considers the extent to which whale song might be regarded as creative, by invoking certain criteria used to assess this attribute in human music. On the basis of these various comparisons, the paper concludes by attempting to draw conclusions applicable to those engaged in designing evolutionary music simulation/generation algorithms.
Keywords
Memetics, meme, museme, proto/musilanguage, Universal Darwinism, recursive-hierarchic structure, whale song, creativity.

1 Introduction
Leaving aside the use of computers for the analysis of music, the computer synthesis of music – conducted variously under the rubrics of Evolutionary Computation (Miranda, 2004), Computer Simulation of Musical Evolution (CSME) (Gimenes & Miranda, 2011), and Musical Metacreation (MUME) (Eigenfeldt, Bown, Pasquier, & Martin, 2013), among others – has made rapid progress over the last fifteen years. Whatever the method underlying the synthesis – recombination (Cope, 2001), genetic algorithms (Özcan & Erçal, 2008), crowd-sourcing (MacCallum, Mauch, Burt, & Leroi, 2012) or biocomputation (Miranda, 2014) – music simulation systems have expanded the boundaries of computation and have produced music which is interesting and convincing.

Whilst not wishing to diminish its success, a significant portion of this work has arguably been conducted with too much focus on the mechanics of computation and on the aesthetic appeal of the end products, and with insufficient reflection on the evolutionary-functional context of what is being attempted. In particular, music is often taken as an immutable, Platonic “given”, with little consideration of its evolutionary history or function; how it relates to language, meaning and physicality; or how these aetiological factors impinge upon the methodologies chosen for its simulation. In saying this, I am referring to those simulations which aim to produce music which is recognizable and comprehensible to humans. While I am not denying the possibility of “non-anthropocentric” creativity (Velardo & Vallati, 2016) – the generation of music which is beyond humans’ perceptual-cognitive reach, and which might only be produced or appreciated by machines – it remains at present largely a theoretical possibility.

2 “Social-Emotive” Vocalizations as Music and Language
What we term “music” is related to what we term “language” not as discrete and distinct categories, but as opposite ends of a continuum of social-emotive vocalizations (Morley, 2012, p. 127) (which include “infant-directed speech”
(Dissanayake, 2008)). In early hominins (modern humans and our immediate ancestors) a “protolanguage”, or “musilanguage” (Brown, 2000), appears to have slowly diverged (beginning c. 200,000 years ago) into music and language, with the emotive, melodic and rhythmic aspects of the protolanguage becoming more concentrated in music, and the social, semantic and syntactic aspects becoming more concentrated in language (Mithen, 2006). Despite this divergence, both preserve traces of their protolinguistic ancestor: music retains certain lexical/grammatical elements (such as combinatorial rules of chord progression and phrase concatenation); and language retains certain melodic/emotional elements (such as the presence of intonation in a majority of languages, and pitch inflexion even in non-tonal languages).

Brown hypothesises the relationship between language and music as shown in Figure 1, which represents two levels at which music and language operate.

- **The phonological level** in language concerns phonemes and their assembly into words and phrases. In music this level concerns pitches and their assembly into motifs and phrases.

- **The meaning level** in language relates to “propositional syntax”, which “specifies temporal and behavioral relationships between subjects and objects in a phrase” and which “is based on relationships between actors and those acted upon” (Brown, 2000, pp. 292–293, 296). In music this level relates to “pitch-blending syntax” (Brown, 2000, p. 274), which correlates sound structures with their expressive effects.

At its far left-hand side, the continuum (top part of Figure 1) represents language as the use of sound for referential meaning and, at its far right-hand side, represents music – in its “acoustic mode” (Brown, 2000, p. 271) – as the use of sound for emotive meaning. Even though “pure” language and “pure” music are never wholly unmixed, there are a number of intermediate states between these two extremes, which afford evidence that the bifurcation of musilanguage is not (yet) total and that there are therefore modes of communication which, while primarily linguistic, retain significant traces of musicality, and vice versa.

### 3 Segmentation, Musemes/Lexemes, and Recursive-Hierarchical Structuring in Music and Language

Music and language divide a sound-stream into discrete particles owing to the segmentational pressures described by gestalt psychology (Deutsch, 1999),
Figure 1: The Music-Language Continuum (Brown, 2000, p. 275, Figure 16.1)
specifically the phenomenon of “chunking” (Snyder, 2000, pp. 53–56). One might term these particles “musemes” (musical memes) and “lexemes” (language memes), respectively (Jan, 2007, 2015a). Both are implicated in syntactic structure and referential function (stronger in language than in music); and both possess emotional-expressive content (stronger in music than in language). Given this segmentational tendency, and the consequent formation of particulate entities from a larger pool of sonic “raw materials”, it is likely that at some point in our prehistory the Variation-Replication-Selection (VRS) algorithm (Dennett, 1995, p. 343) “booted up”, a “Universal Darwinian” process (Dawkins, 1983) was initiated, and music/linguistic-cultural evolution followed.

4 Whale Song as Musilanguage

The acid test of Universal Darwinism is, obviously enough, its universality. While it is a hypothesis which is intrinsically difficult to falsify (Popper, 1959), one can buttress it by amassing – as, analogously, “traditional” evolutionary biologists have (Ridley, 2004, Chapter 3) – a large body of evidence in its favour. Some such evidence comes in the form of animal vocalizations – some zoomusicologists refer to them as “music(s)” – the study of which reveals a number of characteristics which are directly analogous to human music and language, and which are therefore pertinent to their computer simulation (Kirke, 2011).

Like human music and language, the song of the humpback whale (Megaptera novaeangeliæ) has a recursive-hierarchical structure: a series of distinct musilinguistic entities, in combination, engender a structure at the next higher level (Whitehead & Rendell, 2014). (Many of the following observa-
tions on whale song apply equally to the songs of certain bird species (Fitch, 2009; Earp & Maney, 2012.) In Payne’s formulation, shown in Figure 2, the basic building block is the unit, a single sound parcel (≡ note) with a specific pitch shape, timbre and duration. Units are combined into phrases (≡ longer “protemes”) of up to fifteen seconds’ duration. Phrases normally have at least six units and may be divided into subphrases (≡ shorter protemes) of up to seven seconds’ duration, each with a normal minimum of three units. Two or more phrases are combined, sometimes by repetition, into themes of up to two minutes’ duration. Themes assemble, in groups of up to ten, to form the songs themselves, lasting up to twelve minutes. If a song is repeated without a significant pause, it forms a song session, the longest recorded example of which is over twenty hours (Payne, 2000, pp. 136–137).

While the songs of different humpback groups are dissimilar in content, owing to variations in the configuration of their sub/phrases, the basic recursive-hierarchical structure remains constant across the species. This implies a genetic foundation for vocalization similar to the perceptual-cognitive constraints operative in humans (Lerdahl, 1992) and partly regulated by the “FOXP2” gene (Carroll, 2003). Perhaps the genetic basis for segmentation is common to certain primates and cetaceans, being an “ancestral homology” (Ridley, 2004, p. 431). Beyond inter-pod differences, there is also evidence of intra-pod variation. Comparing the same pod over successive years, their songs change, and this occurs during the active (summer) singing period and not as a result of inter-migration (winter) forgetting (Payne, 2000, p. 139).

Specifically, there is mutation of sub/phrases (adding/deleting/modifying units) and, at a higher hierarchical level, there is mutation of themes (by means of substitution of sub/phrases) (Payne, 2000, pp. 138–139). While the biological context for this mutability is sexual selection – songs are produced by males as part of mate-attraction/competition rituals (Parsons, Wright, & Gore, 2008) – the changes are cultural-evolutionary: there is a genetic foundation for all animal vocalizations but the kind of “creative” mutability observed in humans and whales is almost always driven by some form of memetic evolution.

Whale vocalizations are presently neither musical nor linguistic but, on account of having attributes of both forms of communication, proto/musilinguistic. Perhaps they will bifurcate at some point in the future, as is hypothesized to have happened in hominins, into even more complex and lyrical whale music and fully compositional – segmented, syntactic – whale language. There are several caveats to this, however. For one thing, music is an intentional object (Dennett, 1989): a dog is presumably oblivious to music as music because humans interpret its sound patterns in ways the
Figure 2: The Hierarchical Structure of Whale Song (Payne, 2000, p. 137, Figure 9.1)
dog cannot. More fundamentally, intentionality, not only in the engagement with music but also in the production and reception of language, presupposes consciousness.

A future whale language would therefore imply whale consciousness; or at least it would require – to reiterate Brown’s assertion – a developed form of “propositional syntax . . . based on relationships between actors and those acted upon”. Such syntax appears a likely consequence of those numerous systems of social organization where animals must negotiate relationships in the ultimate service of their individual selfish genetic advantage. Indeed, Carruthers (2002) hypothesizes that propositional syntax itself underpins and motivates consciousness, arguing that

[all] non-domain-specific [conscious and unconscious] thinking operates by accessing and manipulating the representations of the language faculty. More specifically, the claim is that [all] non-domain-specific [conscious and unconscious] thoughts implicate representations in what Chomsky . . . calls ‘logical form’ (LF). Where these representations are only in LF, the thoughts in question will be non-conscious ones. But where the LF representation is used to generate a full-blown phonological representation (an imagined sentence), the thought will generally be conscious. (Carruthers, 2002, pp. 658, 666; emphasis in the original)

As represented in Figure 3, domain-specific thought (e.g., a whale’s conceptions of geometry (the pod’s position in relation to other pods, for instance) and number (of conspecifics in the pod, for instance) could eventually be integrated by the kind of domain-general LF underpinning human language. If the protemes already constituting whale vocalizations come to “token” such a LF, it would then be potentially accessible to a whale’s ToM (Theory of Mind) module (if and when evolved) and thus made conscious. Given this, and as Blackmore would argue, a whale would then be able to answer in the affirmative the question “[a]m I conscious now?” (Blackmore, 2005, p. 27).

5 Creativity in Whale Song

As cultural evolution finds high-level novelty – and thus an expansion of society’s conceptual and expressive vocabulary – by means of processes operating via memetically based Darwinism, so biological evolution finds low-level novelty – and thus an expansion of its replicators’/vehicles’/species’ survival
Figure 3: Thought, Modularity and Language (Jan, 2015a, p. 16, Figure 1)
“vocabulary” – by means of processes operating via genetically based Darwinism. In this sense, the problem-space exploration characteristic of evolution is a form of creativity.

It is tempting to term this Darwinian creativity but because, in an Universal-Darwinian view, all creativity – natural and cultural – is driven by the VRS algorithm, the term is tautological. If creativity is a process of generating novelty such that (i) new ways of connecting existing elements within a problem space are found; or (ii) new areas of a problem space are investigated; or (iii) a problem space is redesigned in order to locate more radical solutions, then these three categories – Boden’s combinational, exploratory and transformational creativity, respectively (Boden, 2004, pp. 3–6) – readily describe processes intrinsic to (Universal) Darwinism.

Whale song arguably involves all three forms of creativity: its sub/phrases are assortatively recombined; the “hypervolume” (Dawkins, 2006, p. 65) of all possible songs is progressively explored; and the systemic level of organization may, in principle, be transformed by the power of the VRS algorithm to expand the range of replicable combinations of units and sub/phrases – thus, a whale analogue to tonality might, in principle, become an analogue to atonality (Jan, 2015b).

6 Conclusion: Some Lessons for the Computer Simulation of Musical Evolution

This discussion has said little on the computer simulation of musical evolution itself, but the observations made are, I would suggest, directly relevant to all who seek to use computers to emulate human creativity. In summary:

- Universal Darwinism allows us to regard seemingly dissimilar phenomena using the same conceptual framework. The sounds of music, the sounds of words, and certain vocalization of animals arguably devolve to the same thing: they are replicated cultural patterns, or pro/mus/lexemes.

- Most animal vocalizations are social-emotive: they communicate visceral feelings, often apropos relationships with conspecifics. This attribute is burned into music and language as well – we can usually tell from somebody’s tone of voice whether they are happy or sad, even if the words they use say the opposite; and the character of a piece of music is usually readily discernible.
• A special feature of language is that social-emotive sound-parcels become associated (either “iconically” or “indexically” (Tolbert, 2001)) with objects in the real world and concepts arising from interaction with these objects, giving complex semantics – and a necessarily richer syntax – on top of the existing phonology.

• To social and emotive, one should add embodied (Shapiro, 2011): music and physical movement are inseparable and – as further evidenced by primate vocalizations and their associated movements (Fitch, 2006) – are likely to have been so for early hominins.

• It is likely that whale song is at the stage in its evolution reached by early hominins around the point of the bifurcation of musilanguage into their current trajectories. Whales await the “cognitive revolution” (Harari, 2014, p. 22), itself partly driven by (musi)language, which bootstraps this process.

The implications of these points for computer simulations of music are that:

• Simulations should ensure that the fundamental musical units manipulated by the system make perceptual-cognitive sense, certainly if the emulation of anthropocentric/zoo-centric creativity is intended.

• Simulations should attempt to incorporate not just the phonetic and syntactic dimensions, but also the semantic. In this way the resulting music makes contact with the social-emotive-embodied (musilinguistic) dimension central in the early evolution of music.

• Simulations should also recognize the element of sexual selection (Miranda, Kirby, & Todd, 2003). While certainly not the whole story, it is likely that human music – and almost certainly whale song – evolved partly as a result of male displays of reproductive fitness (Miller, 2000).

These “shoulds” are not intended to sound authoritarian and peremptory; but if recreating a similar, albeit “counterfactual”, path to that taken by human music is the aim of computer simulations (and other aims and other strategies are possible and valuable), then it makes sense to go with the grain of Universal Darwinism; to try to (re)connect music with sociality, emotion and physicality; and to see music and language, and their common social-emotive-embodied driving forces, as two sides of the same evolutionary coin.
References


