A Signal-Based Model of Teleology in Tonal Music

by

Mark André Brand

Thesis presented in fulfilment of the requirements for the degree of

Master of Science in Engineering at Stellenbosch University

Supervisor: Prof. Thomas Niesler

Department of Electrical and Electronic Engineering

Co-supervisor: Mr Theo Herbst

South African College of Music

March 2016

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Mark Brand

March 2016

Summary

The operationalisation of music's affective agency is an open problem. This thesis notes a long tradition of associating that agency with goal-directed dynamics in the music, but also that those dynamical forces are consistently described in terms of a theoretical framework which privileges Western tonal music by its reliance on the constructs of pitch and rhythm. Against the backdrop of a revival in interest regarding the question of musical universals, an alternate, geometric intuition is then proposed to account for the privileged status of the perfect cadence in tonal music in terms of a least action principle. An extensive review of harmonic priming experiments underscores the robustness of that status, regardless of a subject's level of musical education or experience, and strongly suggests a degree of universality. A similarly extensive review of theoretical approaches to sub-Saharan African music reveals how such a geometric model might also account for teleology therein.

The proposed model is operationalised by building a signal processing pipeline containing a recurrent, complex-valued, gradient-frequency artificial neural network, where each neuron implements non-linear dynamics according to a fully expanded canonical form describing the behaviour of excitatory / inhibitory neural populations near a Hopf bifurcation. In the absence of any suitable measure of network energy in the literature, an energy balance relation is constructed whereby to monitor energy flows into and out of the network. In each of four experiments, the network is first primed with a one second stimulus containing the pitches of a dominant quartad in closed position on G4. Seven targets, being the available diatonic triadic continuations to the quartad, are then each presented to the network in turn, and the individual components of the energy balance relation are recorded as the network seeks out its equilibrium.

The first experiment employs sinusoidal stimuli, and wholly excludes inter-neural connectivity. This results in all targets taking the network to the same averaged rate of power supply in each case, being equal to the rate of dissipation at equilibrium, and provides no evidence of an advantage of any single target over the others. The second experiment implements typical "train-and-predict" semantics by allowing non-linear Hebbian learning to take place during the priming phase, holding learned connections fixed during the presentation of the target. This produces differentiation between the levels of supplied energy, favouring targets which contain the pitches of the prime, but does not thereby support our model. The third experiment allows learning to continue throughout both the priming and target phases, and thereby obtains a lower level of supplied energy in respect of the tonic target as compared to all others, a result which provides tenuous, but enticing support for our model. The fourth experiment employed piano tones instead of sinusoidal stimuli, and produced inconclusive results, thereby suggesting that a more complex network might be required to deal with real-world audio.

Opsomming

Dit bly 'n ope vraag hoe dat musiek se invloed oor die menslike gemoed bestuur mag word. In hierdie tesis word daar gelet op 'n lang-durende assosiasie met 'n doelgerigte dinamika, maar verder ook dat sulke kragte deurgaans beskryf word in terme van Westerse musiekteorie, by wyse van hul afhanklikheid van begrippe soos toonhoogte en ritme. Gegewe hernude belangstelling in musikale universaliteit, word daar dan hierin 'n alternatiewe, geometriese intuïsie voorgelê wat poog om die bevoorregte status van die volmaakte kadens in tonale musiek te bereken in terme van 'n algemene tendens tot maksimale doeltreffendheid. 'n Uitgebreide oorsig van sogenaamde “harmonic priming” eksperimente onderstreep die robuustheid van hierdie status, afgesien van 'n hoorder se musikale opvoeding of ondervinding, en dui sterk op universaliteit. 'n Verdere studie van teoretiese benaderings tot Afrikaanse musiek, suid van die Sahara, wys dat so 'n geometriese model ook hier soortgelyke doelgerigtheid mag ontbloot.

Die voorgestelde model word in werking gestel deur die bou van 'n seinverwerkingspyplyn wat 'n herhalende, kompleks-gewaardeerde, gradiënt-frekwensie neurale netwerk bevat, waar elke neuron nie-lineêre dinamika implementeer volgens 'n volledig uitgebreide kanoniese vorm wat die gedrag van stimulerende / inhiberende neurale populasie beskryf naby 'n Hopf bifurkasie. By gebrek aan 'n geskikte meetinstrument om die netwerk se energie te bepaal, word 'n energiebalansverhouding opgerig waarmee energievloei na en van die netwerk waargeneem kan word. In elk van vier eksperimente, word die netwerk eers voorberei deur middel van 'n een-sekonde stimulus wat bestaan uit die note van 'n dominantvierklank in geslote posisie op G4. Sewe teikens, wesenlik die beskikbare diatoniese drieklanke wat sou kon volg op die vierklank, word dan elk op sy beurt ingevoer, en die afsonderlike komponente van die energiebalansverhouding word aangeteken soos die netwerk sy ewewig uitsoek.

Die eerste eksperiment maak gebruik van sinusvormige stimuli, en sluit inter-neurale verbindings heeltemaal uit. Gevolglik neem elke teikenstimulus die netwerk na dieselfde gemiddelde drywing, dié gelyk aan die dissipasiekoers by ewewig, en verskaf dus geen bewys van enige voordeel wat een teiken mag geniet bo die ander nie. Die tweede eksperiment implementeer die tipiese “rig-af-en-voorspel” siklus deurdat nie-lineêre Hebbian leerprosesse mag plaasvind tydens die eerste fase, terwyl hierdie verbindings staties gehou word tydens die invoer van die teiken. Hierdeur word 'n onderskeid gemerk tussen vlakke van energietoevoer, met voorkeur aan teikens wat toonhoogtes bevat wat tydens die eerste fase gebruik is, en dus steun hierdie resultate nie ons model nie. Die derde eksperiment laat die leerproses toe om deurgaans voort te gaan, en verskaf sodoende 'n laer vlak van energietoevoer ten opsigte van die tonikum teiken, vergeleke met alle ander. Die resultaat verskaf skrale, dog aanloklike rugsteun vir ons model. Die vierde eksperiment maak gebruik van klaviernote, eerder as sinusgolwe, en verskaf onbeduidende resultate, 'n aanduiding dat 'n meer komplekse netwerk dalk benodig sal word om nie-kunsmatige klanke so te kan verwerk.

Acknowledgements

A number of people have, over the course of this project, and whether by argument, agreement, comment or tacit support, been prepared to contribute some of their time and energy to this project. I would like to express my sincere gratitude to each of the following:

- My supervisor, Prof. Thomas Niesler of Stellenbosch University, for boundless patience and trust;
- My co-supervisor, Mr Theo Herbst of the South African College of Music, for unrelenting support;
- Prof. Ed Large and Dr Ji Chul Kim of the Music Dynamics Laboratory, UCONN;
- Prof. Morné Bezuidenhout for professional and personal mentorship and inspiration;
- Mr Japie Coetzee, Mr Christiaan Pretorius and Dr M.C. du Plessis for selflessly and repeatedly availing themselves to discuss my flights of fancy. Many others have been engaged less formally, and I apologise for not mentioning each by name. My sincere thanks, regardless;
- My employer, Nelson Mandela Metropolitan University, for financial assistance in 2010, and continued logistical support thereafter;
- My colleagues, especially in the Department of Music, for enduring my frugal participation in institutional life;
- My family, both close and extended, for having been subjected to my extended absence in every imaginable sense except, perhaps, the physical.
- The community of Free/Open Source Software (FOSS), without which this endeavour would not have been possible for me. No proprietary software was used in conducting this research.

You have all made this possible, thank you.

Dedication

Hierdie tesis word opgedra aan Mia.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In conversation with a physics professor I was once asked: "What makes *major* happy and *minor* sad?" My answer clearly didn't satisfy him.[1] The fact is that, despite centuries of scrutiny by a dirge of the most pre-eminent minds,[2] little (if any) success has been had in bridging the divide between the physics and aesthetics of music. More particularly, the writings of so many music theoreticians are littered with metaphorical allusions to movement, force and space, yet our formal understanding of "what makes music tick" has essentially coalesced into a set of axioms divorced from any physical basis. Nonetheless, despite the post-modern mood (in music research, at least) which frowns upon their quest, some researchers (notably from disciplines other than music) have continued the search for empirically justified models to account for plainly observable, yet seemingly ephemeral musical phenomena.

1.1 Background to the Problem

Two themes underscore this research. Firstly, and more broadly, the development of Music Information Retrieval (MIR) (Herrera *et al.* 2009) has brought tremendous resources to bear on selected problems in the domain of music (MIREX 2011), in many cases confirming the complexity of "simple musical tasks". Conspicuously, the axioms of Western music theory seem to be taken largely for granted in such research, despite the inconsistencies alluded to above (these will be expanded upon later). Of course it may be argued that Western music theory defines these problems (e.g. key and tempo detection, transcription from audio to symbolic score, genre classification) and the data which is analysed (music recorded on Western instruments, or Western symbolic notation), but this is certainly not the case for all music of interest (Lartillot *et al.* 2008). The field of Computational Ethnomusicology has been proposed in apparent recognition of this (Tzanetakis *et al.* 2007), and has suggested various interventions, including the "design of domain-specific techniques". Nonetheless, a cursory survey of research in this field shows sustained reference to Western norms (e.g. by relating pitches to Western chromatic pitch classes, comparison of tuning systems to Western twelve tone equal temperament (12tet) or simply employing the "cent" scale in measurements), rather than attempting to explain observed phenomena in their own terms (Cornelis *et al.* 2010). Our interest here is in exploring to

---

[1] At this time I was able to offer an answer relating *major* and *minor* as formal inverses of each other, thus not as 'happy' and 'sad' *per se*, but rather as polar opposites (see (Mazzola 2002:146–148) and (Benson 2006:160–161) for two contrasting approaches in this matter – the former appealing to formal algebraic logic, the latter to psychoacoustic phenomena).

[2] Tonietti proposes "Euclid, Plato, Ptolomy [*sic*], Aristoxenos, … Maurolico, Benedetti, Stevin, Kepler, Galileo Galilei, Mersenne, Huygens, Descartes and Wallis. … Leibniz, Newton, a member (members?) of the Bernoulli family, D'Alembert and Euler" as a sample (Tonietti 2002). Many more could be added.

what extent MIR and/or Computational Ethnomusicology can be detached from the abstract formalisms of Western music theory, yet retain the ability to engage in a musically meaningful manner with audio signals.

The second, more specific theme is the notion of teleology, or “directed motion” in music, widely regarded as a significant component of music's affective influence (Meyer 1956). The association of music with motion is ubiquitous, and many have posited theories to account for this phenomenon. In the “Quadrivium” of ancient Greek scholars, the study of music (or rather, “harmonics”), was the study of the “internal motion” of discrete quantities, as opposed to the “external motion” of astronomy (Kline 1964). Eduard Hanslick highlighted the dynamism of music (Hanslick 1974), while Ernst Kurth proposed (somewhat metaphorically) “potential” and “kinetic” energies at work in music (Kurth 1922). This line of investigation has continued explicitly in the work of Geoffrey Chew (Chew 1991), and implicitly in other studies of “tonal tension” (Cohen 2001; Margulis 2005; Farbood 2006; Lerdahl & Krumhansl 2007). Perhaps the most contentious debate within these circles at present is whether there is any sense whatsoever in seeking a cause for these effects in the audio signal (Marmel *et al.* 2010). The present study will position itself within the context of Marc Leman's notion of “corporeal embodiment”, particularly the call for “... the development of techniques that extract action-relevant features from physical energy...” (Leman 2008:133).

This research will re-approach teleology in music from a purely physical point of view (a “*tabula rasa*” approach, if you will), specifically denying any appeal to the axiomatic formalisms of Western music theory. Its purpose will have been served if it produces a representation of perceived “movement” in the audio signal, particularly a representation which produces results which are somewhat congruent with extant computational methods of extracting comparable measures (e.g. (Margulis 2005), (Farbood 2006), or (Lerdahl & Krumhansl 2007)), but which is further able to extend its application to music that is inadequately represented by “Common Music Notation” (CMN) -based approaches.

1.2 Problem Statement

Conventional understanding of musical teleology is deeply embedded in the axioms of Western music theory. It is therefore at odds with the study of music falling outside of that specific tradition, and essentially absent from any understanding of music as a feature of every known human society throughout history. On the understanding that teleology lies at the heart of music's agency, we wish to develop a method to efficiently expose such dynamic tendencies in a cross-cultural way.

1.3 Purpose and Value of the Study

The operationalisation of music's effect on listeners is an open problem, with numerous research projects across the world examining its various facets. However, no published study could be found which spoke to musical teleology in a culture-neutral way. If this study succeeds in positing tentative evidence of such, it would immediately bolster attempts to understand music as a universal phenomenon, rather than as a socio-cultural anomaly. This, in turn, would speak as much to an enhanced understanding of the diversity of the world's musics, as it would to the creative desire to exploit music in ever more innovative ways. Profound implications for music education, music therapy and the business of music, for example (and no doubt applications too), would not lag far behind.

1.4 Research Question

Is there a correlation between the perception of musical teleology and any physical attribute of an audio signal? If so, then in terms of what understanding of musical teleology, and what physical attribute? What is the nature of such a correlation?

1.5 Research Objectives

This study will initially formulate a theoretical position, and then explore the implications of that position by way of *in silico* experimentation.[3]

1.5.1 Theoretical objectives

A culture-neutral theoretical position will be formulated to:

- account, in a universally generalisable way, for the musical tendency to proceed according to particular conventions, despite there being a virtually infinite set of possible continuations at every moment.

1.5.2 Experimental objectives

A set of experiments will seek to establish, within the limits of our resources:

- the manner in which such a “generalised” theory may be operationalised;
- the extent to which such operationalisation is able to predict teleological tendencies in Western music, given only a digital audio recording; and
- the extent to which one might make comparable predictions in respect of non-Western musics.

1.6 Methodology

By the prevailing scientific paradigm, we would ideally have a testable hypothesis about the relationship

---

[3] The term *in silico* first appears in the early work of Pedro Miramontes, followed by frequent appearances in publications of the Santa Fe Institute throughout the early 1990's. See, especially: Sieburg, H.B., 1990. Physiological Studies In Silico. *Studies in the Sciences of Complexity*, 12(2), pp.321–342. The term bears particularly strong associations with computer models of biological processes.

between a dependent, and one or more independent variables. The constructs at hand, though, are all deeply embedded in a theoretical framework which we will shortly show to be unsuitable for our purpose. Thus we cannot leverage the customary notes, chords and rhythms of Western music theory, but must first either find alternatives, or develop our own constructs which are able to address our questions solely with reference to the audio signal. Only then can we set about designing appropriate experiments to test the efficacy of those constructs, and ultimately to attempt to falsify our hypothesis.

1.6.1 Rationale of the research design

Mindful of Karl Popper's critique of induction, we nevertheless note his reluctance to level such criticism at "... the act of conceiving or inventing a theory." (Popper 2002:7)[4] Since the formulation and operationalisation of a novel construct is precisely what lies at the heart of our research, our axiological, ontological, epistemological and methodological choices have emanated instead from a pragmatist's, rather than scientific realist's theoretical perspective (Cherryholmes 1992), following the tradition most notably represented by John Dewey (1938), but also by W. V. O. Quine (1951) and C.I. Lewis (1970). This paradigm resonated particularly well with our interdisciplinary, exploratory enquiry by foregrounding the option to engage both qualitative and quantitative methods; a so-called "mixed methods" approach (Johnson & Onwuegbuzie 2004).

A purely quantitative study would have, in this particular case, been restricted to variables proposed by reigning conventions, which will, in due course, be shown to have been antithetical to the quasi-transformative, post-colonial subtext of this research. On the other hand, a purely qualitative study would have been unlikely to have provided the computational model desired here. Qualitative methods here addressed the need to explore and better understand the nature of the phenomenon under scrutiny, whilst quantitative methods sought to reveal proposed causality, or at least correlation between the variables identified in the qualitative study. No new phenomenological data was generated by the study, since adequate experimental support for the existence of the phenomena in question was found in reviewing extant studies. Rather, the emphasis fell on formulating and validating a model seeking to parsimoniously account for apparently disparate musical phenomena, specifically manifestations of musical teleology in different cultural practises, under a common principle.

The research design comprised, in its initial, qualitative component, an extensive, interdisciplinary literature review, wherein we pursued the identification and critique of themes which appeared to relate to the phenomenon under investigation. Specifically, we eclectically reviewed thematically-relevant evidence discovered in the fields of experimental psychology, Western tonal music theory, and theories of African music, and attempted to induce from such, by critically interrogating the methods and assumptions evident

---

[4] It is one of Popper's fundamental assertions that: "...there is no such thing as a logical method of having new ideas..." (Popper 2002:8)

therein, a tentative, predictive, universal theory of musical teleology. That contrary evidence often existed in the literature (and was presented when it seemed prudent to do so), was not at issue here, since pragmatism suggests that valuable insights may lie buried beneath the hegemonic discourses of the past. Rather, we were guided by a sustained awareness of the anticipated consequences of our findings. Specifically, we wished to decouple the study of musical teleology from a reliance on Common Music Notation (CMN), and thus the concomitant limitations imposed thereby. This, in turn, was driven by a desire to contribute generally to the development of Indigenous Knowledge Systems (IKS) in the South African context.

To clarify: while an initial literature review contextualised the rationale underpinning this research (Chapter 2), a substantial amount of further critical reading was required, in disparate disciplines, in order to adequately formulate the constructs employed in the quantitative component of the research. We therefore incorporated these extended engagements with the literature as a qualitative component within a mixed-method research design. We reported the findings of this biased and less systematic (albeit extensive) research component in section 3.6, and deferred the additional literature reviews to Appendix B and Appendix C.

The quantitative, experimental component of our research design, following on from the qualitative component, comprised the design and implementation of an *in silico* system (Ören & Yilmaz 2013:168) which sought to operationalise the theoretical model suggested by the qualitative study, and subsequent validation and exploration of that implementation by way of various numerical simulations, with critical consideration given to the results obtained thereby. To the extent that these results concurred with the predictions provided by extant theories of musics, we took such to strengthen our hypotheses proposing a universal mechanism of musical teleology. However, we have not attempted to definitively prove such (by providing and testing falsifiability criteria, for instance), nor to assert generalisability beyond the cases specifically examined, deferring such to a future, more comprehensive study. As regards the nominated test cases, we elected not to stray too far from the work of Large *et al*, given that we found, in due course, that our interests significantly overlap that research project, diverging only in the final instance.

1.6.2 Instrumentation

The experimentation of the quantitative component required a software tool able to perform novel analysis on a standardised, digital representation of an audio signal. In the course of discovering the precise nature of that analysis, strong consideration was given to MARSYAS (Tzanetakis & Cook 2000), as also to DANA (Rougier & Fix 2012), but it ultimately emerged that a built-for-purpose implementation was required. Discussion of that implementation features in Chapter 4. Our implementation ran its first simulations during the Spring of 2014, and Large informed us of the release of their MATLAB GrFNN 1.0 toolkit on 2 December 2014 (Large *et al.* 2014), an implementation which, at a glance, would have suited our purposes. However, since we did not have the use of MATLAB, we continued to employ our own tool, though we have enjoyed the subsequent benefit of Large *et al*'s source-code as *post hoc* validation of our own implementation.

1.7 Scope and Limitations

The objectives of this research are ambitious, given that every one of these is initially grounded, to a large extent, in speculation. We must therefore remain open to iterative re-scoping of the research project and refinement of its objectives as constructs emerge and feasibility becomes clearer. To begin with, we will restrict ourselves to a rather narrow reading of "teleology" in Western tonal music, as exemplified in the dominant-to-tonic harmonic progression which typically marks the strongest sense of finality in such music. We will then seek our generalisation by examining particular African musics, especially those associated with our own location in the Eastern Cape Province of South Africa.

The time and resources available to complete this project are limited, and so we do not expect to complete an exhaustive study. We will pursue the stated objectives regardless, producing whatever is attainable subject to these constraints. At a minimum, we hope to present preliminary, but promising evidence to support a data-driven view of musical teleology.

1.8 Overview of the Thesis

Further to the brief background sketched at the beginning of this chapter, Chapter 2 will contextualise and develop, with reference to extant literature, the two primary themes driving this study. The naïve intuition which orients our approach is presented in Chapter 3, together with the essential findings of our extensive, interdisciplinary readings, and the subsequent formulation thereof into appropriate experimental constructs. Chapter 4 will address the operationalisation of those constructs, while Chapter 5 will undertake the design of experiments seeking to validate the same, and implementation of test instruments. Chapter 6 will present the results of these experiments, followed immediately by our analysis and discussion thereof. Chapter 7 will make some concluding remarks and recommendations, and will point to future work.

Chapter 2

Framing the Problem

The brief introduction of the previous chapter will have raised a number of questions which will now be examined more carefully. Why should Western music theory not be assumed valid in the application of MIR to musics outside of its canon? What is axiomatic about Western music-theoretical approaches? What insights could one expect of a different theoretical approach? The problem is multi-faceted, and requires that we approach from a number of arguably overlapping perspectives. Our investigation of the "first theme" will take a fairly linear route in considering the limits imposed by a CMN-based approach to musical analysis, at first generally, then with particular emphasis on non-Western (especially African) music, and finally within the context of MIR. The "second theme" will be approached by first tracing historically the mainstream theoretical formalisms that have come to represent teleology in Western music, followed by a more eclectic survey of alternative, generally somewhat more metaphysical / metaphorical / aesthetic appeals to the same. Leading on directly from these, an examination of apparently divergent "post-tonal" approaches will reveal common principles at work, frequently unacknowledged by their own theories, and thus inform the development of our model.

2.1 Common Music Notation-Based Approaches

While the term "music theory" has arguably accommodated various, somewhat distinct formal systems (e.g. Medieval neumatic writing, Baroque figured bass, Common-Practise tonality, Twentieth Century serialism), and has moreover been adapted to reflect various theoretical extensions (e.g. compositions by Harry Partch, Ben Johnston or James Tenney), it remains firmly rooted in a tradition of thinking about music in terms of a symbolic reduction. Barton McLean examines the causes and effects of this, proceeding from anthropologist Edward T. Hall's concept of "extension", whereby humans abstract their surroundings in order to increase efficiency (McLean 1981). Following on Hall's distinction between the complementary processes of "externalizing" and "internalizing", McLean contrasts composed and improvised music as examples, respectively, of these two ways of engaging with music. Particular difficulties arise in communicating internalised (e.g. improvised) ideas, and therein lies the advantage of an external representation (e.g. a composed score). However, the inevitable result of the externalisation is detachment, wherein the symbol takes on a life and meaning of its own. He continues: "... it seems to me that the classical serial technique as developed by Schoenberg and its further extensions thereafter are a prime example of the extension notion gone wild, to the point where the serial technique 'amputates' itself from the natural inclinations of the basic pitch materials and from vital human perceptual responses as well." (1981:335) McLean posits "several implicit and often erroneous assumptions [which] lie behind much music theory of the past several hundred years". Firstly, he questions the assumption that music can be meaningfully analysed and explained in terms of notes, rhythms and the like, "compartmentalization in a more or less arbitrary scheme", presumably assumed valid precisely because these are the raw materials with which the music is apparently constructed. His second (questioned) assumption is that music can be completely de-constructed using a "verbal-analytical system." He cites John Shepherd's contention that "anything which cannot be so reduced... immediately becomes non-knowledge". Third, he questions the suitability of assuming separation of form and content in music. Fourth, he highlights an implied assumption that "discursive modes of speech and the non-discursive modes of music are interchangeable..." Finally, and perhaps most pertinently within this study, he questions the assumption that "music is constituted of, and can be totally defined in terms of, its notation", pointing out the significance of "intangibles as performance-gestural modes, timbre, and texture". McLean observes a tendency amongst musicologists to "select works for study on the basis of their visual-conceptual properties alone", underscoring his contention that the written score (and the symbolism it represents) attains a higher status than the performed music from which it takes its *raison d'etre*. He cites McLuhan's assertion that "this separation of symbol from its meaning relates to the advent of movable type printing", contrasting the Western alphabet's "semantically meaningless letters... used to correspond to semantically meaningless sounds" against Chinese ideograms, wherein individual symbols are largely meaningful in and of themselves. Returning to the development of musical notation, McLean turns to Trevor Wishart: "... the neume did not attempt to mark out what we now have come to regard as individual pitches and units of rhythm, but only shapes and contours of melodic lines customary in current practice..." However, music notated in this way (typically called Medieval chant) is essentially monophonic (consisting of only one melody), and it is here that McLean seems to find the cause for much of his dilemma: "It was precisely this need to quantify pitch and rhythm, the two parameters necessary for scribing ensemble music with several vertical parts sung together, that produced our unique Western notational system." This, he says, turned out to be a "double-edged sword", because "the farther the basic sound material is extended intellectually, symbolically, or verbally, the farther one may be from the basic creative truth."

We might be reminded that McLean is here speaking of the suitability of CMN-based analytical approaches for understanding *Western* music. A similar argument is to be found in debates concerning the suitability of typical measures of audio quality (Harley 2004), arguing that "architects' specifications" are merely design tools, and not suitable as objective measures of quality, a view which clearly resonates with McLean's posited first erroneous assumption above. So then, how much more problematic would it be to apply such a system to the understanding of non-Western music? We will now consider selected instances in which typical Western-theoretical techniques have been applied to non-Western music.

David Temperley asked the question: “How well can African rhythm be reconciled with the prevailing music-theoretical view of rhythm?” (Temperley 2000). With an approach leaning heavily on Lerdahl and Jackendoff's seminal Generative Theory of Tonal Music (GTTM) (Lerdahl & Jackendoff 1983), he concludes that “African and Western rhythm are profoundly similar”, with the primary differences lying in the relative importance attached to the GTTM's “regularity rule” and “accent rules”, as well as differences of preferred alignment between meter and grouping (which are often confused or simply conflated, by his account). He adds, though, that motivic structure (“the network of rhythmic patterns in a piece that are heard as similar or related”) is not addressed by the GTTM model, and concedes that such may be of interest. Jonathan Stock evaluated the efficacy of standard Schenkerian analytical techniques as applied to selected examples from Africa and East Asia (Stock 1993), finding some measure of fit, but also some “departure from Schenkerian orthodoxy”. Both Temperley and Stock find it necessary to engage the question of the validity of Western analytical techniques to non-Western music, both arguing that it is unnecessary (and perhaps undesirable) to approach any particular music only from within its own purported theoretical framework. If the tendency to want to interpret non-Western music in terms of Western theoretical frameworks represents one side of the dynamic, then the other is surely evidenced in Gerhard Kubik's observations that “internalization of colonial norms in music has been the fate of a great number of Western-educated African musicians of the middle generation” (Kubik 1985:53). Kubik cites examples of African musicians having appropriated Western tuning systems, without any acknowledged awareness hereof, where he had explicitly observed different practises in earlier field trips. A similar sentiment has been expressed more recently: “...traditional music is being hybridized with music influences from around the world and frequently is only preserved in recorded archives.” (Tzanetakis *et al.* 2007:20) Thus is would seem that we should perhaps be less concerned with identifying overlaps (clearly there will be more and more as this process unfolds), and more concerned with identifying what has managed to survive outside of Western theoretical frameworks.

The application of computational analysis to music developed on two broad fronts: extracting symbols from audio signals (e.g. transcription-type tasks), and interpreting the relationships between such symbols within a given theoretical framework (e.g. notating functional harmonies, key following, identifying tone-row permutations, etc.). Both have carried forward the bias in favour of Western music theory. As Tzanetakis puts it: “The vast majority of existing work in MIR has focused on Western music. Computer technology itself has no intrinsic bias towards Western music; but since using advanced computer tools to study music tends to happen in universities, and since universities tend to be oriented towards European and European-derived art music traditions, these biases have crept into the field.” (Tzanetakis *et al.* 2007:6) The more recent appearance of a third broad approach, namely measurements of statistical similarity between signals, has paved the way for other approaches which do not invoke Western music theory as the ground truth. We will now review selected examples engaging with the application of MIR to non-Western music.

Gómez & Herrera investigated the extent to which audio data could automatically be classified as Western or non-Western (Gómez & Herrera 2008). They approached the problem by "focusing on tonal features", extracting descriptors related to "... the pitch class distribution of the piece, its pitch range or tessitura and the employed scale and tuning system, being the feature extraction process derived from mathematical models of Western musical scales and consonance." Their implicit assumption is that a weak fit (by Western standards) would be a marker of some music's non-Western character. Their various extracted Transposed Harmonic Pitch Class Profiles (THPCPs)[5] are analysed by two machine learning techniques (Decision Trees and Support Vector Machines, both implemented in the WEKA machine learning software (Witten 2011)) and applied to the task of automatically classifying various pieces of music in an independent test set (specifically, the Voyager Golden Record, cf. (Ferris 1978)). With a best result of 85.16% accuracy in the classification task, they concluded that their selected feature set does "... provide means of characterizing different traditions and styles." Their paper immediately attracts a response, published in the same journal issue (Lartillot *et al.* 2008), wherein their method is described as "audaciously situated on a challenging junction between computer science, cognitive science and ethnomusicology". The critique highlights latent "Eurocentrism" in the comparison to Western standards (e.g. reference to African music's "approximation of the Western tempered scale") and stated notions of "advanced musical systems". Most importantly, though, it is questioned whether an equal-tempered HPCP, albeit of increased resolution, can do justice to a study of non-Western music. Of course, the study is not of non-Western music, but rather seeks to establish markers of differentiation, which it apparently does. This exchange does, however, highlight traditional assumptions which would prove problematic for the application of MIR to ethnomusicology. These are: the assumption of octave equivalence, the use of logarithmically segmented pitch space as a reference, the assumption of a 440 Hz pitch reference, and traditional Western notions of pitch stability and distribution within the octave. The selected point of departure (a somewhat narrow reading of Western music-theoretical constructs) also opens the study up to criticism from Western musicologists (e.g. disregard for earlier European practises such as *musica ficta*, variable tuning systems and pitch standards).

Many of the problems above were confronted by the DEKKMMA project (DEKKMMA 2003), a large-scale digitisation of the audio collection of the Belgian Royal Museum of Central-Africa (RMCA). Early on, these researchers acknowledged that "music information retrieval research usually takes Western music and its musical characteristics and semantic descriptions as a standard, and develops tools following a series of assumptions based on Western cultural concepts." (Moelants *et al.* 2007) They also argue that "... the combination of digitisation and commercial large-scale distribution tends to push 'vulnerable' music even further into oblivion. Music information retrieval research should therefore take into account an ethical code that aims to develop tools for all types of music, not just for Western music." Their particular contribution in

---

[5] Essentially, a binned frequency histogram, here with a resolution of 10 cents, or 120 bins per octave.

that paper stems from the observation that "...in African music fixed tuning does not exist... relative pitch [higher-lower] is more important than absolute pitch." Their response is to move away from the quantised representation of pitch in terms of Western pitch classes, relying instead on a (quasi-) continuous representation in cents (relative to their nominal reference, A1 / 55 Hz). Such a profile is extracted from an entire piece of music by a monophonic melody extractor (which proved to have interesting applications in this polyphonic context) and then folded over into a single octave, thereby giving an "octave-reduced pitch analysis". Their discussion of these representations highlights the extent to which assumptions of octave-equivalence may be misplaced, as evidenced by multiple, closely spaced peaks near certain Western pitch classes.

Similar concerns are raised in MIR's engagement with Turkish music (Gedik & Bozkurt 2010). This study comprehensively reviews the "pitch histogram-based MIR literature on western and non-western musics", presents a "high dimensional pitch-frequency histogram representation without pre-assumptions about the tuning, tonality, pitch-classes, or a specific music theory", and ultimately applies this model algorithmically to the problems of automatic tonic detection and *makam* recognition (roughly analogues to key and mode detection, respectively). The representation chosen here is not reduced to a single octave, and is expressed in *Holderian comma* (*Hc*), rather than cents (the spectral resolution employed is 1/3 *Hc*, or 3x53=159 bins per octave). Fundamental frequency is extracted by means of the YIN algorithm (Cheveigné & Kawahara 2002) together with various post-filters. The extracted profiles are then matched (by "shift and compare") against various *makam* templates. Herein is raised the very interesting issue of what constitutes a suitable measure of similarity, to which we shall return later. Their results show high success on some counts, less so on others.

A broader evaluation of typical MIR classification techniques is undertaken in (Lidy *et al.* 2010). Specifically, this research undertakes the typical cycle of "train and predict" on three distinct data sets with the support of various feature sets. The data sets are a common (predominantly Western) training set typically used as a benchmark in MIR research, a set of Latin American dance music pieces, and a selection of African music from the Ethnomusicological Archive of the Belgian RMCA (mentioned earlier). The feature sets are firstly those provided by the MARSYAS framework, namely timbral texture (STFT and MFCC features), rhythmic (BEAT) and pitch (PITCH) content features[6], along with inter-onset interval histogram coefficients (IOIHC) by Gouyon *et al.*[7], rhythm patterns (RP) by Rauber *et al.*[8], and derived from

---

[6] See Tzanetakis, G. & Cook, P., 2002. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5), pp.293–302.

[7] See Gouyon, F. et al., 2004. Evaluating rhythmic descriptors for musical genre classification. In *Proceedings of the AES 25th International Conference*. London, UK, pp. 196–204.

[8] See Rauber, A., Pampalk, E. & Merkl, D., 2002. Using psycho-acoustic models and self-organizing maps to create a hierarchical structuring of music by sound similarity. In *Proceedings of the International Conference on Music Information Retrieval*. Paris, France, pp. 71–80.

these, statistical spectrum descriptors (SDD) and rhythm histograms (RH) by Lidy *et al*.[9] Two novel feature sets, temporal SDD (TSSD) and modulation variance descriptors (MVD), round out the arsenal. Experiments were conducted with each of the feature sets individually, as well as with the various possible combinations of STFT, MFCC and PITCH, plus one other, yielding a total of 17 feature sets. Classification was performed both by SVM, as well as by time decomposition. Some investigation was also done as regarding the effect of analysing the beginning, middle or end of the piece of music. The results of the analysis on African music are of interest here. Firstly, there was very little difference in the result whether analysing the beginning, middle or end of any given piece of music (contrary to the finding in the other two data sets). Secondly, in the absence of the classification label of 'genre' (as is typically employed in Western analyses), it was decided to substitute 'function'. The result: “... it seems that the concept of a 'function' is not captured very well by the audio analysis methods used.” (Lidy *et al.* 2010:11) We doubt many Africans would have been surprised by either of the above results. Far better results were reported for classification by country and by ethnic group, though the confusion matrices show interesting overlaps, specifically for geographically neighbouring categories. Again, not a particularly surprising result. Overall, particular combinations of the various feature sets do succeed quite well in capturing something of what distinguishes each of these different kinds of music.

What bears noting at this stage is that the first three examples remained closely bound to a music-theoretical framework of some sort. Gómez & Herrera extended the HPCP concept, Moelants *et al.* and Gedik & Bozkurt extended to the point of a (quasi-) continuous representation, but all had a template in mind and stayed quite close to their theoretical precepts. The Lidy *et al.* study, however, is a prime example of what can be achieved by letting go. Consider what is actually being “measured” here: mean and variance of the spectral centroid, rolloff, flux, zero-crossings per frame, as well as the first five Mel-frequency spectral coefficients, for example. These “features” have no correlates in music theory, and yet they clearly capture quantifiable differences between musics that we, too, hear as different. Further, it appears that research in computational ethnomusicology is being driven primarily by issues of accessibility, with deeper understanding of the music as a by-product, at best. Thus one should expect certain aspects to enjoy a privileged status, while others are left behind. What if we are able to establish a universally acceptable continuum of musical similarity measure, to the extent that all humans agree with the results? We would then know unambiguously what drawer to file our music in, but what would we know of what the music means, or does to us?

The issues that are being well attended to are reviewed in (Cornelis *et al.* 2010). Again, the emphasis is on preservation and access, with a call to “... reconsider the underlying musical assumptions of audio-based

---

[9] See Lidy, T. & Rauber, A., 2005. Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. In *Proceedings of the 6th International Conference on Music Information Retrieval*. London, UK, pp. 34–41.

content extraction, which are currently based on Western musical concepts." Further, the warning that "... the use of MIR techniques could further marginalize non-Western musical traditions." (2010:1009) The term "ethnomusicology" is here used to denote interest in the convergence of three distinct categories: music of cultures with no written tradition, non-Western classical music (e.g. Chinese, Indian, or Middle-Eastern) and folk music (both Western and non-Western). In other words, the word is used in its broadest possible sense to indicate the study of everything other than Western classical and commercial / popular music. Cornelis *et al*. proposes links between ethnomusicology and MIR research through "... the interest in using audio analysis tools... including very precise calculations of musical parameters not necessarily related to the Western musical framework." (2010:1010). A taxonomy of low- mid- and high-level descriptors is reviewed, indicative of the more recent willingness to find useful measures below the traditional music-theoretical surface of notes and rhythms. A cursory overview of papers presented at the annual conference of the International Society for Music Information Retrieval (ISMIR n.d.) indicates a slow but steady rise in interest in ethnomusicological matters, though clearly clustered around particular issues. This could indicate pockets of research expertise in some matters, or it may reflect the inability to find suitable MIR approaches to particular problems. Cornelis *et al*. does note in the survey of ISMIR papers a paradigm shift over time, both for ethnomusicological as well as for Western topics, from earlier reliance on annotation and description of single musical features toward the extraction of multiple signal-based features as a step toward higher-level processing. It is further noted that most studies focus on a particular musical style, instrument or geographic region, with few daring to suggest generalisability. Regarding the application of low-level signal descriptors, it is once again noted that "... global and statistical properties of these descriptors are useful for analysing audio that bears no relationship to Western musical concepts. However, the interpretation of these low-level features in terms of their musical meaning presents a main challenge in this case." (Cornelis *et al.* 2010:1012)

According to Cornelis *et al*. (2010), pitch analysis remains high on the agenda, despite the limiting Western view of pitch as a discrete category. Examples are cited wherein pitch is extracted from digital audio, only to be mapped onto equal-tempered pitch-classes. The practise presumably persists to enable similarity analysis and classification techniques. Assumptions about static underlying scales and equal octaves present further problems. As regards temporal organisation, it is pointed out that the application of metric grids (as would be appropriate for much Western music) typically obfuscates much of the real underlying rhythmic structure in non-Western music. Some domain-specific techniques have been developed, but these obviously do not generalise well. Particularly interesting approaches for the present study, based on self-similarity matrices, are found in (Pikrakis *et al.* 2004) and (Antonopoulos *et al.* 2007).

Cornelis *et al*. (2010) cites a number of examples of ethnomusicologists extending Western music notation in novel ways in order to better capture the characteristics of the music under scrutiny, but shows that (for lack of standardisation) this has led to vastly different transcriptions of the same music, and advises caution in relying on such. Clearly, some musics will be far more amenable to Western-style symbolic notation than others and where this is so, classification approaches based on self-organising maps or principal component analysis (applied directly to these symbolic representations) have yielded positive results. It does not appear as if such approaches have been attempted outside of European folk music, though. On the related issue of similarity matching, it is once again clear that symbolic representations attract the most attention, and particularly matters of meter and rhythm.

While pure signal-processing approaches continue to pose challenges, a large amount of effort is directed toward the provision of metadata. Problems relate primarily to the contradictory requirements of flexibility and standardisation, as well as to the laborious nature of the task. The incorporation of metadata provided by non-expert users (gathered by publicly accessible web-services) is being investigated as an alternative to the traditional route of expert description (2010:1020). Clearly, human description-based and signal-based approaches are complementary, though this research focuses on the latter.

The conclusions of Cornelis *et al*. (2010) which speak to this research are as follows: many of the most advanced tools remain reliant on the signal's "musical reduction to Western score notation", leading to "a loss of much of the rich and lively texture of this music"; "radical rethinking" is required to face the problems posed by ethnomusicological research, including the development of "more universal tools"; and better integration is required between "culturally neutral audio representations and features" and "symbolic and meaningful content descriptors". (2010:1025) We interpret the call for a "radical rethinking" by choosing to resist the temptation to merely supplant one system of symbolic extensions for another. We believe there is much more to be gleaned by examining the signal directly, informed by the tenets of music theory, but critically aware of its axiomatic foundations.

So, what is axiomatic about Western music-theoretical approaches? We will now summarise particular points which will have emerged from the preceding discussions. Firstly, notation-based theory sustains the myth that music functions in terms of its symbolic representation. It may even be pointed out that, given a certain level of training, a person may discern which notes sound "good" together, and which not, or whether the "pace" of the music is "frantic" or "relaxed", purely on the basis of the printed score, but this hardly scrapes the surface of musical experience. Those who have the ability to access a deeper appreciation from the score alone inevitably rely on a highly developed "auralisation" skill, the ability to evoke an imaginary (virtual) performance of the music at sight. In the process, much of the "meaning" attributed to the music does not originate on the page, but is largely projected from the reader's "aural memory". Secondly, the quantisation of pitch into discrete pitch-classes directs our attention to absolute pitch, while studies increasingly point to the more important role of the fluid relationship between pitches, particularly in African music. Thirdly, we tend to impose Western notions of "pitch stability" onto music which does not evidence the assumption of octave-equivalence, nor the desire to divide octaves into any particular number of equal parts. We do this without completely understanding the origins of such “stability”, even in Western music. We are prone to make similar impositions in the domain of rhythm. To these we should immediately add our commodification of music, as evidenced in our continued digital packaging of music as “songs”, or our preoccupation with classification according to “preference”. The notion of a “composition” (with a defined beginning, middle and end) clearly does not sit well with much non-Western music, at least not until Western contagion sets in. We do not, however, believe it necessary to dismiss Western music theory out of hand. By reflecting on particular historical developments we will highlight clues which assist us in reaching a different theoretical perspective. To what end? The discussion of our second theme will engage this.

2.2 Teleology in Music

The second major theme driving this research is the notion of motion directed towards a goal, or teleology, in music. The broader question from which this line of enquiry emerges is to what extent causal links can be established between music and emotion. The literature on the latter topic is comprehensively reviewed in (Konečni 2008), concluding that the support for a causal model remains weak, despite sustained interest. Studies in musical “tension” are seen to contribute to this broader field of enquiry. According to Farbood: “The perception of tension is an important intermediate step between the recognition of musical structures and the subjective, emotional response.” (Farbood 2006:23) Fred Lerdahl and Carol Krumhansl agree: “The ebb and flow of tension... appears to have a direct link to musical affect.” (Lerdahl & Krumhansl 2007) In some research the term “tension” is found to be inadequate. Margulis states that “[e]xpectancy has long been cited as a generator of musical affect.” (Margulis 2005) Ultimately these turn out to be two sides of the same coin, the former emphasising the “push”, the latter the “pull”. In either case, what is being addressed is the widely acknowledged experience of music as being imbued with motion. More specifically, Western tonal music characterises such motion in terms of tensions (or expectations) which may be resolved, deferred or denied.[10] This is the line we will take up here, reviewing particular contributions made in Western music theory, and attempting to generalise these into a culture-neutral theory of teleology in audio signals.

Cohen rigorously interrogates the concept of “directed motion” in music, uncovering its Aristotelian roots and restatement by Aquinas, its initial uptake with regard to music theory in Marchetto of Padua's *Lucidarium* (1317/18), and its transmission through Zarlino to Jean Philippe Rameau, the acknowledged father of modern tonal harmonic theory. (Cohen 2001) Rameau inherited the notion that “motion is the actuality of an imperfect thing tending toward perfection”, which had by now come to mean the progression from an “imperfect consonance” to a “perfect consonance”. Perfection was here found in the particular sensory integration experienced between simultaneous sounds each having fundamental frequencies that

---

[10] These expectations are formalised as rules for voice-leading and harmonic progression.

were nominally related by "simple fractions", particularly 1/1, 2/1, 3/2 and 4/3 (closely resembling the modern musical intervals of a unison, octave, perfect fifth and perfect fourth, respectively). Imperfection described other intervals, "consonant" by virtue of their expressibility by rational fractions (e.g. 5/4, 6/5, 5/3 and 8/5 – forebears of the modern major third, minor third, major sixth and minor sixth, respectively), yet "imperfect" because they lacked the particularly "smooth" and stable character of the perfect consonances.[11] For Marchetto, these had been "compatible dissonances" (as opposed to all other, ostensibly "incompatible" dissonances) by virtue of a very particular property: "... the fact that their constituent pitches occupy positions that are 'distant by the smallest distance' from those they will subsequently occupy in the following consonance, created when the two voices move on by contrary motion." (Cohen 2001:150) The core tenet here, taken up in Rameau's theory and still very much with us today, is that music is propelled by the sequencing in time of alternating, complementary sonic constructs (consonances and dissonances), with particular dissonances requiring resolution by equally particular consonances, the specific association determined by a principle of minimum deviation in opposite directions of the component fundamental frequencies. The emphasis, however, has shifted somewhat. Modern theory paradigmatically stresses the progression from dominant triad (the three-note chord built on the fifth degree of a scale) to tonic triad (the three-note chord built on the first degree) as prototypical of "forward movement" in tonal music, despite both sonorities being structurally identical (at least in the major key), and equally "consonant" at that. For Rameau, there could be no "forward movement" without at least the implication of actual dissonance, and thus he strongly recommended the inclusion in all dominant chords of the so-called seventh. (Cohen 2001:141)

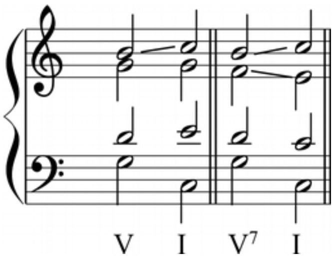


Figure 2.1: Dominant to Tonic Progression (CMN)

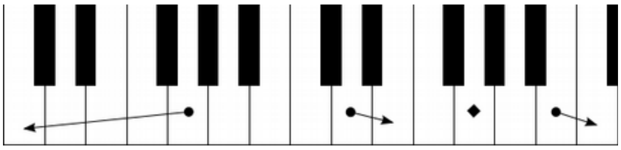


Figure 2.2: V – I at the keyboard

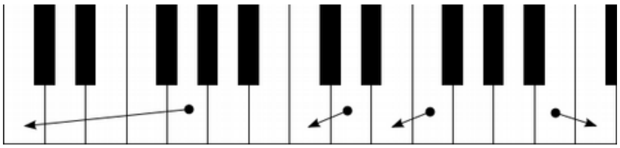


Figure 2.3: V^7 – I at the keyboard

In Figure 2.1, the difference is illustrated in modern Common Music Notation (CMN).[12] Each fragment shows the typical "voice-leading" employed in proceeding from a dominant chord (indicated by the Roman numeral "V", meaning the chord built on the fifth degree of the scale – in this case C major) to a tonic chord (indicated by the Roman numeral "I", indicating the chord built on the first degree of the scale). In the first

---

[11] The classification of consonances as "perfect" or "imperfect" is attributed to John of Garland in the thirteenth century, according to (Cohen 2001:148).

[12] The interested reader should note that the notation does not reflect the specific rational intervals characterising the theory of that time, though playing the fragment at the piano will convey most of what is being argued here.

fragment, the dominant chord is a triad (a three-note chord) consisting of the notes G (occurs twice), B and D, while the second fragment employs the dominant quartad (a four-note chord) containing G, B, D and F. Beside the lowest note (here the "root" of the chord) moving down a fifth to C in each case, all other notes map to the closest members of the tonic chord (either C, E or G). The diagonal lines indicate movement by semitone, the smallest interval in Western tonal music. In the first case, it is only the aptly named "leading tone" which moves by the smallest interval to its closest match in the tonic chord. In the second case, there are two such instances moving in opposite directions – precisely as required by Marchetto to effect directed motion.[13] One might now ask: was contrapuntal theory misguided? Did the principle of voice-leading proximity merely reflect a philosophical premise, a formal imposition which just happened to correlate somehow with how music was apparently heard to function at that time? Does the modern tonal view in terms of dominant-tonic relations supersede those principles? On what basis? Does "directed motion" still play a role in music today?

Rameau famously endeavoured to justify his theories in terms of the "Principe Sonore" or harmonic series, initially grappling with abstract notions of "undertones" to justify constructs such as the minor triad and subdominant (Ferris 1959:236). Clearly he was led by some intuition (or perhaps tradition) that the chord built on the fourth degree of the scale was of some particular significance: an "alternate" dominant, thus named subdominant.[14] Yet his prescriptions for the use of this chord (the chord of "*double emploi*") evidence a different interpretation (Thomson 1993). When preceding a dominant chord, it is advised that a "sixth" be added (yielding the notes F, A, C plus D in C major) which, in Rameau's new harmonic paradigm, is most logically interpreted as a regular quartad built on D, effectively resulting in a cascade of movements by consecutive fifths downward (or fourths upward).[15] It seems clear that Rameau was well aware of the "directed motion" effected by such a progression, a skeletal version of the same mechanism cited by Marchetto, but persisted nonetheless in forging a systemic framework defining dominant and subdominant as somehow opposite to each other. This, and other resultant inconsistencies are discussed in, for example, (Goldman 1968:68).

In modern Western tonal music theory, this triumvirate of tonic, subdominant and dominant reaches us through Hugo Riemann's theory of "harmonic function", a categorisation of every conceivable chord as representative of one of three basic functions within a particular key (Riemann 1896).[16] The propellant for perceived motion in music is now attributed less to inherent tendencies within particular dissonant sonorities,

---

[13] We must clarify that Marchetto belonged to the older tradition of "contrapuntal" music, and was not privy to modern notions of "chords" built as thirds stacked on a root. It is to Rameau that we typically attribute this more modern view.

[14] The subdominant lies the same distance from the tonic as does the dominant, but in the opposite direction.

[15] An alternative interpretation lies in the avoidance of "parallel fifths", a possible, though not unavoidable consequence of such a progression. Thus we lean toward the account given above.

[16] Though Riemann credited Heinrich Koch (1811) for this insight, it clearly originates with Rameau (Thomson 1993:389).

and more to "root movement" within the context of a key. In (Schoenberg 1969): "A certain order promotes such a *succession* of chords to the function of a *progression*. A *succession* is aimless; a *progression* aims for a definite goal." (1969:1) Here such root movements are classified as "strong/ascending" (fourth up, or third down), "descending"[17] (fourth down, or third up) and "superstrong" (one step up or down). Surreptitiously, the goal posts have shifted. Schoenberg is not classifying these progressions according to their perceived character, but rather according to their formal construction. A "superstrong" progression is so-called because the two chords involved share no notes (e.g. F, A, C and G, B, D), and thus hearing them in succession establishes, almost unambiguously, the "definite goal" which, according to Schoenberg, is the "...establishing or contradicting [of] a tonality." (1969:1) The "strong/ascending" and "descending" categories involve pairs of chord having significant overlap, and are thus less effective in this regard (e.g. G, B, D and C, E, G, or C, E, G and A, C, E). Yet the division into "strong/ascending" and "descending" does suggest some residual acknowledgement of different perceptual experiences in each case. The grand codification of this dynamic within mainstream Western music theory is found in Schenkerian analysis, a reductive technique which systematically schematicises the elements of a given musical score in order to expose the various levels at which the progression proceeds towards a goal (Schenker 1954; Schenker & American Musicological Society 1979). At the highest, most abstract level of the reduction lie the *Urlinie* and *Ursatz*, a simple stepwise melodic descent to the tonic note and a I-(IV)-V-I harmonic progression, respectively. Schenker went to great lengths to affirm these as the prototypical expression of "directed motion" in all tonal music, once again seeking justification in the harmonic series (as did Rameau) by positing that every musical tone carries the prototypical major triad in its own harmonic series (Schenker 1954:29), and that all tonal music amounted to a "composing out" or prolongation of this triad. Upon reflection we can see how the sense of local tension has now been supplanted somewhat by a greater concern for an overarching, global model of tension. Such a global view was carried forward by Fred Lerdahl and Ray Jackendoff's "Generative Theory of Tonal Music" (GTTM) which, though differing in many of the details, clearly perpetuates the spirit of a large scale reduction of any given music to some underlying ground truth (Lerdahl & Jackendoff 1983). The GTTM, inspired by Noam Chomsky's Generative Linguistics, posits a system of well-formedness rules, transformational rules and preference rules that govern listeners' perception and understanding of music. These rules are applied in four different contexts, yielding a grouping structure, a metrical structure, a time-span reduction and a prolongation reduction. Interestingly, the prolongation reduction (which most closely resembles the Schenkerian approach) has found its way (explicitly or implicitly) into recent studies of tonal tension (and by proxy, has been associated with theories of affect), yet the GTTM explicitly denies any attempt to engage with the interplay between music and emotional response, deferring instead to Meyer's (1956) assertion that: "to approach any of the subtleties of musical affect... requires a better understanding of musical structure." (Lerdahl & Jackendoff 1983:8)

---

[17] Schoenberg footnotes his aversion to the term "weak".

While we have traced increasingly formalised, systematic notions of "directed motion" above, we shall also consider some of the more "metaphorical" observations arising parallel to these. The volatile coexistence of idealist and materialist views in mid-nineteenth century Germany posed a particular challenge to the aesthetics of music. Eduard Hanslick responded by denouncing implied links to other arts (and their particular poetic idealisms), maintaining that music was wholly defined in terms of the structure of its "tonally moving forms". (Burford 2006) Hanslick rubbished the notion of music as being expressive or representative of emotions, saying: "... only the element of motion... this is the element which music has in common with our emotions..." (Hanslick 1974:38) He speaks of the kinds of ideas which can successfully be expressed in music: "This class comprises all ideas which, consistently with the organ to which they appeal, are associated with audible changes of strength, motion, and ratio... the ideas of intensity waxing and diminishing; of motion hastening and lingering; of ingeniously complex and simple progression; &c..." (1974:36) Hanslick looks quite optimistically toward the then recent work of Helmholtz (Helmholtz 1954), suggesting that "the latest results of physiological research" held the key to any kind of affective link (Hanslick 1974:115). He continues: "the two points at issue – namely, what the distinctive trait is of a feeling aroused by music, and whether this is of an essentially aesthetic nature – are settled by the recognition of one and the same fact: the intense action [of music] on our nervous system..." (Hanslick 1974:122) It appears that Hanslick had some disdain for the then emerging field of psychology. Edmund Gurney similarly emphasised an essential characteristic of music which he termed "Ideal Motion", but was further at pains to distinguish this from any kind of physical motion (Gurney 1880:165). Ernst Kurth's formulation of "potential" and "kinetic" psychic energies in music was posited in (Kurth 1922), and his views enjoyed considerable attention before being eclipsed by those of Schenker. Though discussing Kurth's ideas under a broad heading of "metaphorical appeals to motion", we must immediately note that Kurth himself did not intend the term "motion" figuratively. Kurth regarded the "totality of flowing motion" as the central characteristic of music, with form, the "control of energy through space and time", defining the tension between the energy and its manifestation (Hsu 1966). He contrasts "sound psychology" against "music psychology": "... in one, sound implies a penetration of the inner being; in the other, an outburst from the inner being", conceptually pre-empting more recent theories of musical perception (e.g. see the discussion of Jean Molino's Communication Stream in (Mazzola 2002:12–16), or of Brunswick's Lens Model in (Leman 2008:52)). Dolores Hsu cites the following translation from "Grundlagen": "It is mistaken to emphasize only the acoustic phenomena – the sounding and the sounds themselves with all their latent harmonic relationships – as the most essential and truly significant moments of melody, without paying attention to the connection with perceptions of a proceeding of forces between the tones." Further: "The musical phenomenon merely manifests itself in sounds, but is not based on them." (Hsu 1966:11) Kurth is far less interested in music's potentially mirroring a static natural principle such as the harmonic series (as many others have been) and more concerned with the perceived "energies", reflections of "psychic tensions".

“Sound is dead;” he says, “...what lives in it, is the will to sound.” Kurth's ideas have recently been characterised as a “Dionysian” counterpart to Schenker's “Apollonian” systematisation, though both theorists are seen to be “radically opposed to the 'verticalist' Riemannian orthodoxy of the late nineteenth century.” (Chew 1991:172,173) He is thus drawn to “disruptive, asymmetrical forces” and therein lies much of the attraction for modern theorists who “seek to account for phenomena that undermine order in music.” Geoffrey Chew has pointed to the possible integration of these disparate approaches, highlighting the lack of “intervallic theory”[18] in Schenkerian analysis,[19] and suggesting that Kurth's conceptualisations of “tension” and “energy” remedy this oversight (Chew 1983). Counter to Schenker's insistence that tension is released as the *Urlinie* descends to the tonic, Kurth directs our attention to the ascending leading-note. By extension, the latter view paves the way for interesting perspectives on other semitonal movements, both ascending and descending, diatonic or chromatic, in both major and minor modes. Unfortunately, Kurth's appropriation of the term 'kinetic energy' to denote melodic tension (particularly of 'leading-notes'), together with 'potential energy' to denote chordal tension, does not map well to the meanings of those terms in physics, and so one might find the tongue firmly in the cheek while attempting to follow at least some of his arguments. Nonetheless, Kurth represents another important marker in tracing the tradition of “directed motion” in music.

If a paradigmatic revolution (or rather return) is, in fact, afoot, then it is certainly reflected in William Thomson's more recent “attempt to outflank these empirical/phenomenal clashes” by positing a “scheme of pitch/time interactions, or *vectoral dynamics*.” (Thomson 1993) Thomson's “clashes”, simplistically stated, relate to inconsistencies observed in the naming of chords. More broadly, he interrogates the ubiquitous notion (in Western music theory, at least) that a given combination of simultaneously sounding notes might be meaningfully reduced to the name of one, presumably most important pitch (typically one of the sounding pitches). He points out: “A persistent question, since Rameau's marvelous revelation, is whether the root property can be understood as necessarily imposed by our processing of the sound signal, or whether it is merely a learned response whose characteristics depend on a particular ethnic conditioning.” (Thomson 1993:386) He briefly contrasts the use of the term “root” in various contexts, from mere taxonomic convenience to systemic imperative to conspicuous absence, and finally asks: “if there is basis for the concept, in what sense can it be aligned more persuasively with conditions of empirical verification and with what musicians seem to think they hear in music?” (Thomson 1993:388) Within the Riemannian functional perspective, the term “root” stresses a contextual interpretation: “a chord is what it does”. Thus a particular combination of pitches may be interpreted differently depending on what precedes and follows it, even to the point of postulating non-sounding pitches as “supposed” roots, should this make for a more satisfying contextual analysis. Thomson suggests that such tenuous foundations ultimately permit one to assign any

---

[18] The earlier discussion of Marchetto illustrates “intervallic theory”, as opposed to “harmonic theory” after Rameau.

[19] This, in itself, is considered somewhat surprising in hindsight, considering the widely acknowledged roots of Schenkerian theory in Baroque music (e.g. Bach and Fux).

conceivable pitch as the root of any conceivable chord. As he points out: “A decade rarely passes, for example, that we do not read yet another interpretation of the *Tristan* chord, some new slant on why it is a *this* chord or a *that* chord, why for acoustic reasons *this* is its root or for linear reasons *that*. Let us herald the conclusion that a root so difficult to agree on is really not much of a root at all.” (Thomson 1993:390) Thomson highlights two basic principles from which twentieth century “root theories” have been drawn: the attribution of “rootedness” to the pitch-class which carries the greatest reinforcement (according to some postulated quantitative measure of power); and appeals to various schemes of “pattern meshing of neural inputs” (template matching). The respective counter-examples of both of these approaches are well known: the former suggests that the root should ultimately follow the note played loudest or doubled at the octave (which is patently not the case), while the latter suggests that small deviations from template-prescribed ratios should render a chord “rootless” (which is invalidated by various tuning practises, most notably modern 12-tone equal temperament) (Thomson 1993:392). Reviewing important contributions made by Paul Hindemith (stressing the role of *difference tones*) and Ernst Terhardt (disambiguation by *virtual pitches*), including the extension of the latter's ideas by Richard Parncutt (Parncutt 1989), Thomson concludes that Terhardt's “virtual pitch” concept adroitly circumvents many of the difficulties plaguing earlier theories. In particular, the theory is significantly more accommodating of irrational intervals, while the appeal to a “learned template” (rather than a necessarily present physical interference pattern) makes it possible to talk meaningfully about non-simultaneous pitches as cognitive, rather than physical constructs. Even so, empirical validation remains elusive.[20] Thomson's research uncovers a small number of relevant studies, some of whose results might be summarised as follows: there is no evidence for 'natural categories' of pitch intervals, and in any case, musically untrained subjects do not discriminate to the degree that their musically trained counterparts do (Burns & Ward 1978); it is doubted whether “the perception of tonal centre is predicated... on a structural hierarchy of pitches” (Brown 1988:246); “root perception takes an unsuspected amount of time to accomplish” (suggesting a wholly cognitive, rather than automatic process), and “an interval's root potency is affected by contoural direction” (Smith 1967); and roots may exist, but not in every case, and are perhaps too unpredictable to be trusted (Hurwitz 1970). Yet Plomp observes that the 4th and 5th are more often confused with each other than either is confused with the tritone, despite each being closer to the tritone than to one another (Plomp 1965), and Hurwitz (1970:26) reports subjects, having been asked to select either note of a sixth dyad as root, singing instead a note between the two, completing the triad.

Thomson's response is classically Gestaltist. He proposes a contextualising view, a cognitive frame which makes sense of the sensory input by fitting that input to a pattern, for which the root merely serves as a “vectoral nucleus” (Thomson 1993:401). At this point, the parallels to Terhardt's “virtual pitch” are quite striking. However, Thomson distinguishes his own approach from Terhardt's by placing the primary

---

[20] See also, for example, Konečni's (2008) critique of various canonical methods to appreciate the difficulty of framing such experiments.

emphasis on the “purely pattern aspect”, being decidedly unconcerned with the quantitative comparison of virtual pitch salience (as in (Parncutt 1989), for example). The pattern in question remains the natural harmonic series, particularly its lowest nine partials. To this, he adds the concept of “circular neural configurations” (or “regenerative loop”), a concept underlying “memory trace”[21], in order that “tones separated in time can be weighted and comprehended in terms of part-whole schemata” (1993:404) (therein addressing oft-cited shortcomings in both Hindemith's and Terhardt's theories). On accounting for the origins of the “perceptual framing action of music”, Thomson is somewhat more guarded, but notes the recurrent use of “5/4th deployments of melodic pitches within the octave [providing] the primary distinction of authentic from plagal mode forms.” (Thomson 1993:405)

Thomson demonstrates his approach in a comparative analysis of the first 12 bars of Hugo Wolf's song “Das Verlassene Mägdlein” (1888). This particular passage is clearly heard to be in A (minor), yet neither a traditional “tertian” analysis, nor the application of Parncutt's method of extracting chroma salience (a specialisation of Terhardt's virtual pitch model, see (Parncutt 1989:146–150)) yields any unambiguous A's as roots.[22] Instead, Thomson directs our attention to the *ambitus* (or pitch range) of structurally significant sections of the extract (A_5-A_4 in the first two bars, A_5-E_4 in the four bars preceding the entry of the voice, E_4-E_3 over the first eight bars of the vocal melody, and A_5-E_2 over the full piano accompaniment), which contextualises the two tritones (F_5-B_4 and D_5-$G\#_4$ in measures 2 and 3) and projects unambiguous resolutions (C_4, E_4 and A_4, respectively). He further notes the expectation generated by a sequence of falling diatonic fifths which ultimately lead to the anticipated A. “In this texture,” he concludes, “contoural and rhythmic dynamics target the determining pitches A and E for whose archetypal pattern A is fundamental.” (Thomson 1993:409) Thomson stresses the significance of the *ambitus*, directly associating his pitch frame concept to the medieval distinction between authentic and plagal modes. Quoting philosopher Andrew Ushenko “...the bearing that sense data or images have upon one another in an aesthetic experience takes the form of a felt pressure or tension that forces the beholder's perception and imagination into a contextually prescribed course” and explicitly referring to “direction-bearing forces”, his “perceptual vectors” are nonetheless bound within that pitch frame, pointing to a particular pitch as tonic (a modern-day *finalis*, if you will). We find the interaction between the tritones and the pitch frame far more interesting for the present enquiry, and it is this aspect to which we shall return later. Finally, Thomson acknowledges the exceptional character of the selected musical fragment, “a phenomenally tonal passage in which chord roots are *not* a controlling cause”, reminding us that in most tonal music we should expect to find a high degree of alignment between these perceptual vectors and traditionally interpreted chord roots, precisely for the purpose of reinforcing the

---

[21] Thomson is here invoking Koffka’s notion, as elaborated in: Goldmeier, E., 2014. *The Memory Trace (PLE: Memory): Its Formation and Its Fate*, Psychology Press.

[22] The assumption that the tonic chord must occur in the passage is, of course, an oversimplification. Many would be quite happy to observe the key signature and raised leading-note as sufficient grounds to justify what is heard to be A minor. Thomson's point, however, should be given some latitude, especially considering that neither key signature nor accidentals would necessarily be available to an analysis of the audio signal.

"vectoral focus".

Thomson's concerns remain rooted in tonality, but the interest in a return to earlier principles is already evident, severed from tonality, seventy years earlier: "For it is apparent, and will probably become increasingly clear, that we are turning to a new epoch of polyphonic style, and as in the earlier epochs, harmonies will be a product of the voice leading: justified solely by the melodic lines!" (Schoenberg 1983:389) Ben Johnston characterised Schoenberg as a "radical thinker motivated strongly by a claustrophobic sense of nearly exhausted resources." (Gilmore 1995:473) Of course Johnston was merely taking sides on a watershed moment in the development of Western compositional practise. In seeking to move forward, Schoenberg chose to forsake tonality (but retain equal temperament), while Harry Partch (a forebear of Johnston's tradition) chose to forsake equal temperament (retaining tonal principles). As it turns out, "voice-leading" represented the principle means of maintaining continuity for Schoenberg in atonal music, guided by three, apparently contradictory principles: "The chord progression seems to be regulated by the tendency to include in the second chord tones that were missing in the first, generally those a half step higher or lower. Nevertheless, the voices seldom move by half step. Then,... tone doublings, octaves, seldom appear. … There is perhaps also an instinctive (possible exaggerated) aversion to recalling even remotely the traditional chords." (Schoenberg 1983:420) Roeder weighs these principles against each other, examining the contradictions, and ultimately underscoring the central role played by semitonal voice-leading relations in effecting progression (though, in this case, between pitch-classes and not specific pitches) (Roeder & Schoenberg 1989). Partch would seem to have been motivated less by issues of "voice-leading" *per se*, presumably because these were implicit in his "Tonality Diamond", though he did coin the term "tonality flux" to describe "a sequence of chords, each of which resolves onto the next by narrow intervals." (Gilmore 1995:469) Rather, one might see voice-leading principles as incidental to gesture on an instrument such as the Diamond Marimba, for instance (Gilmore 1995:472). Thus it appears that there are particular "voice-leading" principles, largely implicit in traditional tonality, which become even more pronounced in pre- and post-tonal music. The most explicit formulation of these principles arises in Western music theory, so we will consider a particularly interesting review of these next.

David Huron accounts for the canonical "rules of voice-leading" by a somewhat more engaged review of perceptual principles, wherein he acknowledges the dynamics of melodic expectation, implication, tension and resolution, but immediately defers explanation thereof as "premature" (Huron 2001:3). His account is motivated by the widely professed principle that "voice-leading is the art of creating independent parts or voices." (2001:3) He therefore links, for example: Terhardt's "spectral pitch dominance region", via Parncutt's "tonalness" (Terhardt *et al.* 1982b, 1982a; Parncutt 1989), to range prescriptions; dissipation of Neisser's "echoic memory" (Neisser 1967) to the predominance of *legato* lines and short rests; Greenwood's tonotopic interpretation of Zwicker's "critical bandwidths" (Zwicker 1961; Greenwood 1961) to sensory dissonance and preferred "vertical" spacing between voices; DeWitt & Crowder's investigation of Stumpf's formulation of "tonal fusion" (DeWitt & Crowder 1987) to the avoidance of unisons, octaves and perfect fifths; Van Noorden's (1975) "fission boundary" and "temporal coherence boundary" to the preferences for step-wise motion (and also to account for pseudo-polyphony); and Chowning's (1980), Bregman & Doehring's (1984) and McAdams' (McAdams 1982, 1989) demonstrations of tonal fusion by co-modulation to the avoidance of parallel unisons, octaves and fifths. The preceding synopsis naturally glosses over Huron's meticulous application of each psychoacoustic principle to dozens of canonical (and some novel) rules within Western music theory, but not all are relevant here. Rather, despite his reluctance to directly engage the kind of dynamics we are interested in, we find particular moments in his arguments which betray a marked awareness of precisely what we are after.

Firstly, the overarching principle of "auditory stream segregation" (McAdams & Bregman 1979; Wright & Bregman 1987; van Noorden 1975) addresses the "Gestaltist" tendency of human listeners to group spectrally proximal sonic events into distinct "voices", and thus to perceive a succession of "notes" as separate, fluctuating lines of pitched sound (melodies). This should be seen in the light of at least four empirical phenomena, but particularly Bregman's conclusion for the "preeminence of pitch proximity over pitch trajectory." (Huron 2001:24) Streaming accordingly favours pitch-time traces formed by the association of a given pitch to the closest available subsequent pitch: "a *competition* between possible alternative organizations." (Huron 2001:35)[23] On the other hand, it is noted that timbral dissimilarity, as manifested for instance in divergent spectral centroids, exercises a significant counter-dynamic which competes against pitch proximity and event rate (Huron 2001:48). Second is the observation (in (Warren *et al.* 1972)) that "the frequency/intensity thresholds for auditory induction coincide closely with the thresholds for auditory masking." That is to say: if there is an expectation that a sound is likely to have been masked, listeners tend to hear that sound even if it is not (and perhaps never was) part of the signal (Huron 2001:11). Further to a postulated notion of the limbic system rewarding the brain for successfully parsing perceptual phenomena, where "the amount of pleasure evoked may be proportional to the perceptual difficulty posed by the scene", Huron characterises dissonance by a "negatively valenced phenomenal experience of listeners – akin to annoyance", and suggests that this might arise out of our (unacknowledged) recognition of masking, implying an inability to successfully parse the auditory scene (Huron 2001:56–58). Third is the relationship between critical bandwidth and sensory dissonance, particularly as a potential driver of musical motion, but also noting the converse risk of tonal fusion (2001:19).[24] Fourth is the significance afforded by Van Noorden's (1975) "fission boundary" to the interval of a semitone, suggesting that this is the largest interval

---

[23] Huron goes so far as to suggest "magnetic" attraction between pitches, but immediately notes some of the shortcomings of this metaphor (Huron 2001:36).

[24] Huron carefully distinguishes these terms by applying them as follows: "Perfect consonances typically exhibit low sensory dissonance and high tonal fusion. Imperfect consonances have low sensory dissonance and comparatively low tonal fusion. Dissonances exhibit high sensory dissonance and low tonal fusion." (Huron 2001:21–22)

(approximately) which guarantees the perception of a single line at common tempi (Huron 2001:27). More generally, Huron's invocation of "Fitts' law" (Fitts 1954), via an application of Körte's "third law of apparent motion in vision" (Körte 1915), can only be considered reasonable if one acknowledges motion of some sort in sound (see also (Gjerdingen 1994)). Fifth, the powerful effect of positively correlated logarithmic co-modulation on tonal fusion (Huron 2001:31) is of particular interest when we reconsider notions of distance measure later. It may be possible to reconcile this with the similarly effective co-evolution of amplitude envelopes (Huron 2001:39), perhaps a macro-scale manifestation of the same phenomenon. Sixth, we note with interest the various temporal thresholds proposed in the discussion of onset synchrony, suggesting shades of perception between 20 ms and 100 ms (the latter a recommended lower bound for notes to be perceived as temporally distinct) (2001:39–40). Seventh, the interpretation of Descoeudres' (1921) "*un, deux, trois, beaucoup*" phenomenon, as applied to the recognition of the number of concurrent voices in a given piece of music, suggests low-level cognitive constraints which might perhaps be innate, rather than learned (Huron 2001:46).[25] Finally, we note amongst Huron's conclusions that his view "... does not account for the sense of direction ("leading") that attends musical pitch successions... the above account implies that voice-leading should be equally effective when a work is played backward rather than forward." (Huron 2001:55) This is clearly a centrally important issue in this research: we are seeking to account not only for "motion", but for "directed motion".

If an exclusive focus on voice-leading has somehow obscured the teleological aspect of the dynamic we're investigating, then perhaps it is prudent to review the alternate paradigm. In Bob Gilmore's investigation into the legacy of Harry Partch, particularly as extended by Ben Johnston and James Tenney, Johnston is cited as characterising his own domain as closer to "the kind of complexity needed to understand the intricate symbiotic interdependence of organic life on earth" than to "the kind which clarifies the statistical behavior of inanimate multitudes." His assertion that "the extreme complexity of contemporary life [can] be reconciled with the simplifying and clarifying influences of systems of order based upon ratio scales" (Gilmore 1995:475) suggests an anachronistically cynical attitude toward a twentieth century world-view rooted in chaos and probability, compared to John Cage and Iannis Xenakis, for example. The waning of determinism and notions of observer neutrality, fuelling increased acknowledgement of the roles of subjectivity and pluralism, is evident in the natural sciences as much as in music (Delaere & Daly 1990), though residual positivist tendencies paradoxically persist in our marginalisation of certain views. Superficially, ratio-model based approaches (whether by appeal to a "natural" principle such as the harmonic series, or to the "divine" status of rational numbers) seem to have no place in the lair of Schrödinger's cat, but we might do well to reconsider such a view in the light of recent evidence suggesting that chaos and order are more closely related than we had hitherto thought. Stephen Wolfram, for example, has shown how apparent chaos emerges from the most ordered processes, suggesting the reducibility of seemingly complex

---

[25] Also consider limits discussed in (Harley 1991:12–13).

phenomena to simple, rational, discrete rules, given sufficient computational resources (Wolfram 2002). The pair correlation of the zeroes of the Riemann zeta function have been associated with numerous phenomena considered impervious to the Newtonian world-view, including the distribution of the primes, random matrices, and atomic energy levels (Borwein *et al.* 2006). Perhaps Johnston's reluctance to forsake rationality is not so misguided, after all.

At the heart of it, Johnston argued that "equal temperament – both as a tuning practise *and as a language* – confounded relationships that were distinct." (Gilmore 1995:475) The axiomatic nature of equal temperament has already shown itself to be problematic in the earlier review of computational ethnomusicology, and is here seen to be an undesirable imposition within some quarters of Western music too. Reviewing the multitude of tuning systems and temperaments which have emerged through the ages (Barbour 1972), and particularly attempts to solve problems by mechanical, rather than by theoretical means (Wolf 2003; Benson 2006), one might conclude that equal temperament eventually prevailed only when all other options had been exhausted. Given the technology of its time (particularly the rise of fretted and keyboard instruments), it best satisfied the competing demands for increased mobility between keys on the one hand, and playability on the other (Barbour 1972:6,8). Perhaps most tellingly, "... it is significant that the great music theorists, such as Zarlino, Mersenne, and Rameau, presented just intonation as the theoretical basis of the scale, but temperament as a practical necessity." (Barbour 1972:11) Attempts to reopen negotiations around that compromise solution today seem decidedly unwelcome.[26] What emerged is the twelve tone equal temperament (12tet) system, a discrete segmentation of audible frequencies into twelve logarithmically spaced increments for every doubling of frequency. This both posits and perpetuates a view of pitch as being one-dimensional, linear and discrete. Gilmore cites Mach: "A tonal series occurs in something which is an analogue of space, but is a space of one dimension limited in both directions and exhitbiting no symmetry... it more resembles a vertical right line."[27] Some have recognised the need to reintroduce higher-dimensional ordering, proposing spiral, conical and toroidal topologies instead (Shepard 1982; Krumhansl & Kessler 1982; Chew 2000), and typically acknowledging origins in Hugo Riemann's *Tonnetz* (a two dimensional lattice). An even older model, attributed to Leonard Euler, features prominently amongst those who have abandoned 12tet, but retained tonality. Guerino Mazzola dubs this "Euler space", with any given note occupying an "Euler point" $x=(p,s,r)\in\mathbb{Q}^3$, with frequency $f(x)=132\,Hz\cdot 2^p\cdot 3^s\cdot 5^r$ (Mazzola 2002:1031).[28] The bases could be any arbitrary set of primes, though musical practise has typically favoured those given above. Johnston's String Quartet No. 4 (1973), for example, employs the primes 3, 5,

---

[26] The truth (or not) of this, as well as the reasons, would make for interesting research. We state this here only as a considered opinion in the light of personal experience.

[27] Gilmore finds (Helmholtz, 1954), Mach, E., *Erkenntnis und Irrtum: Skizzen zur Psychologie der Forschung*, 2d ed. (Leipzig: J. A. Barth, 1906) and Koffka, K., *Principles of Gestalt Psychology* (New York: Harcourt, Brace and Company, 1935) as all evidencing this view.

[28] To clarify, Mazzola has here taken C_3 as his reference, rather than the traditional A_4 (440 Hz), and the exponents *p*, *s* and *r* refer to the octaves, fifths and thirds, respectively.

and 7 in mapping out his 22-note scale on a three-dimensional lattice. In such cases the need for a base 2 axis (and thus a four-dimensional representation) is typically averted by folding all pitches into a single octave. Nonetheless, aside from the impracticality of a visualisation, five-dimensional lattices (2, 3, 5, 7, 11) were used by Partch (Gilmore 1995:481) and he investigated the use of up to seven primes (Gilmore 1995:467). The primary benefit arising from the use of "Euler spaces" (or lattices, as they were known to Partch *et al.*) lies in their remoulding of a flat frequency space into a structure more closely aligned with harmonic perception.[29] Of course Shepard, Krumhansl and Chew undertook to do the same, but their approaches all tacitly accept the axiomatic formalisms of Western, equal-tempered, tertian harmony.[30] Though arguably exchanging one set of axioms for another, our interest in "Euler space" is motivated by the notion that a theory grounded in numbers might be significantly less culturally loaded. For Johnston, an important advantage of the lattice representation is its disclosure of "harmonic neighbours", and thereby the means to motion: "[p]atterns or 'paths' in the lattice may be used to give a sense of harmonic direction even to passages that lack entirely any conventional tonal or harmonic 'logic'" (Gilmore 1995:483), though it is not entirely clear to us how this was approached teleologically.

James Tenney used the term "harmonic space" to describe his multidimensional model of pitch perception, acknowledging Partch's contribution as "an indispensable *technical* point of departure, just as Cage's work has provided us with an essential *aesthetic* foundation." (Gilmore 1995:485) Gilmore also cites a reciprocal acknowledgement: "In a 1990 interview Cage describes listening to a performance of Tenney's *Critical Band* and realizing that 'there was no wall at all, that sound is by its own nature harmonious...'" Tenney explicitly called attention to the semantically problematic use of the term "*harmony*", bound as it was to the system of triadic tonality. Instead, he proposed the need for a theory of "*harmonic perception*", "... one component in a more general theory of musical perception." (Gilmore 1995:486) He envisaged a theory that was quantitative (and thus value-free), aesthetically neutral, culturally and stylistically general, "and in consequence relevant to music of different periods and cultures." His "harmonic space" extended earlier models in important ways.[31] Firstly, he dispensed with Partch's notion of "Utonalities", essentially inverse ratios derived from "Otonalities", which in turn derived from the harmonic series. Tenney apparently saw these as an all-too-convenient abstraction ("extension", see (McLean 1981)), generally invoked to account more elegantly for phenomena such as minor triads, yet having no basis in the real world.[32] The ratios in

---

[29] We have doubtless been somewhat reckless with our use of the term "harmonic perception", but the greater part of music theoretical history has been in accord with the notion of "harmonic" as somehow associated with simple ratios. We hope the reader will grant us this latitude here.

[30] Chew, for example, proposes an *a priori* spiral model built on fifths and major thirds, and then derives constraints according to the desired theoretical outcome (Chew 2000:61–98).

[31] Amongst these is his assertion that harmonic space "is not symmetrical. It clearly has an up and down." (Gilmore 1995:487) Though this would seem relevant for our purposes, we have not found anything further on the matter yet, and thus reduce this idea to a footnote for the time being. See (Belet & Tenney 1987).

[32] In this sense, "Utonalities" are somewhat reminiscent of other minor-mode justifications in terms of "undertones" (cf. (Rameau, 1971))

Tenney's harmonic space were "referential", rather than absolute, and he invoked the notion of "tolerance" to account for the listener's tendency to hear deviant pitches as approximations of simple, rational relationships. This tended to ameliorate the need for more than three or four primes (thus three or four dimensions), compared to Partch. It further allowed him to consider equal-tempered notes as suitable approximations in certain contexts. The particular relationships perceived were determined by overlapping constraints, most notably the interaction between the desire for simple, rational ratios and the desire to effect an optimally compact representation on his lattice of harmonic space. Tenney's measure of *harmonic distance* (Hd) is thus singled out as perhaps his most "crucial development", affording him the means towards "compactness". His is a Manhattan, rather than Euclidean metric, defined as $\mathrm{Hd}(a/b)=k\log(ab)$, with a/b the maximally reduced ratio representing the frequency difference, and $k=1$ indicating measure in octaves.[33] The resultant correlation between low "harmonic distance" and high perceived consonance was explicitly cited as one of the primary advantages of the model. Tenney's approach remained Gestaltist (and in this particular respect, contrary to Partch's) in that he viewed the harmonic series not as a causal factor, but rather as a physical manifestation of principles already evident in harmonic perception. He distinguished his approach from Helmholtz's "entitive referents,... [which] are generally dyads or other simultaneous aggregates *isolated from any musical context*... [and having] nothing to do with... motion, resolution, or chordal connections of any kind. It refers merely to the perceptual character of individual chords." Clearly, Tenney required of a model the ability to address the transitions between moments, to account for the "harmonic dynamic" in music, in this respect having more in common with earlier theorists like Rameau for whom this dynamic lay in "resolution" (Gilmore 1995:490). What remains unclear, however, is how Tenney's system informs "directed motion". His 1988 composition *Critical Band* (discussed in Gilmore's article) begins with a sustained 440 Hz tone, followed by a fanning out of other tones above and below according to the principle that preceding notes are the harmonic means of their successors. By our reckoning (as we shall later demonstrate), this is not progression, but regression, though it is of course entirely possible (or rather, likely) that this is precisely what the composer wanted.

Clarence Barlow frames Tenney's "tolerance" concept in a subtly different way. Rather than proposing an a priori set of rational ratios, his approach is to determine, given an arbitrary interval, the most likely *perceived* rational relationship (Barlow & Lohner 1987:47). He gives the following example: "... play the series of notes C, D, F, E (Mozart) and C-sharp, B-sharp, E, D-sharp (J. S. Bach) on a piano to compare the impression produced by the third C–E and the diminished fourth B-sharp–E, respectively. The degree of tension definitely differs between the two. This difference results from a more or less subconscious bending of the notes to an optimal tuning within the mind's ear of the listener." (Barlow & Lohner 1987:48) His candidate ratios are factored to primes, with an *indigestibility* function which weights the order of the primes

---

[33] Parameters a and b are initially read from two dimensions of Tenney's lattice, and the fraction a / b simplified until the smallest possible integer values of a and b become the terms of the product ab.

required against their maximum powers (smaller primes and lower powers are more *digestible*). Component indigestibility values are combined in a *harmonicity* function, which expresses "the degree of simplicity of a numerical relationship". Working backwards, Barlow now sets an arbitrary *minimum harmonicity* constraint and derives the "sequence of maximum absolute powers" (corresponding to the sequence of primes) required to formulate all intervals within a given pitch range (and satisfying the constraint), and ultimately generating the appropriate ratios. These are then weighted by a "Gaussian-like bell" according to a *tolerance* parameter around the reference pitch, finally yielding a "winner". He extends this principle to scales, admitting a nominal number of "winners" for each scale degree to the consideration of overall harmonicity between the intervals of that scale. This is established as follows: "... the sum of the harmonicities of all possible intervallic links between the degrees of the scale is determined for each constellation of tuning alternatives. The constellation that results in the largest harmonicity sum is chosen." (Barlow & Lohner 1987:48) Barlow summarises his result as the "*specific harmonicity* of the tuning constellation", the square of the number of pitches divided by the sum of the inverse harmonicities. However, after setting out to determine a perceptual framework, he paradoxically demonstrates his model with a real-time retuning implementation.[34] Having earlier posited that we perceive irrational relationships in simple, rational terms anyway, he develops a system which dynamically adjusts the tuning of pitches according to context, and seems disappointed at the "curious phenomenon" which causes pitches to "fluctuate microtonally" (Barlow & Lohner 1987:53). We contend that this is a positive result, indicative of the dynamism in musical perception. Certainly, such retuning would render subliminally shifting perceptions garishly overt, and perhaps precisely thereby undermine their effect, but Barlow's implementation should be seen as an experimental setting for explicating these relationships, not as a performance tool.

In the "Second Essay" of his article, Barlow turns his attention to the perception of "meter", proposing a quantitative stratification of metrical positions or pulses within a bar according to a measure of *indispensability*. Once again, bars may be "factored" into an arbitrary sequence of primes, each yielding a set of fundamental indispensabilities. These, in turn, allow him to calculate the indispensability of any pulse within the bar. From this vantage point, Barlow launches a most interesting (and, as far as we can ascertain, unprecedented) inquiry into the establishment of a measure of *metrical affinity* between two meters of different stratification and speed (a quantitative investigation of poly-rhythm, if you will). His approach is to take the mean of the products of coincident *relative indispensability* values (the latter are indispensabilities normalised to 1), yielding his "coefficient of metrical affinity" (Barlow & Lohner 1987:58). Finally, having separately approached the ostensibly distinct perceptual domains of pitch and rhythm, Barlow compares metrical affinity values to harmonicity values for 32 different ratios, concluding from the high correlation observed that: "If one were to consider audible pitch as an extremely rapid series of pulses with frequencies

---

[34] We certainly do not mean to suggest that his experiment is not of interest, only that it might have contributed to a misreading of the greater significance of his results.

specified as tempo indications, intervallic harmonicity would be a kind of 'micrometrical affinity!'" (Barlow & Lohner 1987:60) This view is in line with our own intuition, and will be taken up in our model.

If our discussion has seemed to stray from the stated theme, then this is precisely symptomatic of how layers of abstraction have obscured the musical phenomenon which we are targeting here. Tuning theories which originally developed within the context of a more or less explicit theory of teleology (specifically as regards local tension and resolution) have come to prioritise more global concerns. While it is common to view earlier ratio-based tuning practises as "restricted", they arguably hold the advantage of containing their implied key centre within their structure to a greater or lesser extent, thus requiring less of the music to do so. Equal temperament, for all its flexibility and symmetry, clearly does not, which may account for the greater musical effort directed at this global goal. As a result, intervallic theory's musical propulsion by symmetric simplification (of intervallic ratios, that is) is supplanted by Schenkerian theory's explication of a key by a I-(IV)-V-I root progression, clearly an "extension" in McLean's terms and well detached from any cause, at that. Despite Rameau's plea to retain particular contrapuntal principles (by retaining the seventh in all dominants, for instance), it is his "tertian" view of harmonic organisation which has overshadowed Western thinking, and thereby a triad's dynamic tendencies have come to be determined by context, rather than by structure. However, if we have accepted Rameau's mitigation that the seventh may be implied (presumably by reference to the seventh partial of the harmonic series), then we might accept that it is similarly implied for all pitches, thus suggesting an "intervallic" dynamic to account for all "movements by fourth", not just dominant-tonic. Such an approach seems (for our purposes) vastly better equipped to account for certain pervasive harmonic practises than the axiomatic "cycle of fifths" typically invoked. Similarly, Chew's reconciliation of Kurthian and Schenkerian approaches to melodic tension reaffirms the central role played by small melodic intervals. This is further evidenced in our review of both "atonal" (explicit) and "post-tonal" (implicit) approaches. Thomson succeeds in balancing global and local tensions by interpreting tritones within the context of their surrounding pitch frames, and idea which resonates with Barlow's contextual rationalisation of intervals. Huron's review of explicit rules for voice-leading informs our view of "spectral streaming", particularly as regards the interactions of proximity, masking and co-modulation, as well as his invocation of "Fitts' Law" and limits of cognitive processing. While many have recognised the expedience of a multidimensional representation of pitch space, Tenney highlighted "Euler space" (he used the term "lattice") as a suitable component in an "aesthetically neutral, culturally and stylistically general" theory. In particular, his notions of "tolerance" (as taken up by Barlow) and "optimally compact representation" are of interest. Barlow's further contribution is his relating together of spectral and temporal domains by common principles. Most importantly though, Barlow closes out our review of directed motion by giving us the means (by rationalisation of irrational intervals) to return to Aristotelian principles in our consideration of teleology in music.

2.3 Concluding Remarks

We have surveyed the problem domain and motivated our line of enquiry by tracing our two primary themes in the extant literature. It should be clear that MIR needs to take a fundamentally different approach to music theory as it expands its interest to other musical traditions. To wit, a purely signal-based view of teleology in tonal music is conspicuously absent, and poses a promising alternative approach. Various pertinent principles have already emerged from the preceding review and will inform the design of our own model. Still further exploration will be incorporated into the research design, expanding on the principles identified thus far.

Chapter 3

Constructing Musical Teleology

We have set ourselves the task of developing a culturally neutral analytical tool for music, specifically a tool with which to identify the phenomenon of teleology in audio signals. Such an undertaking is inherently controversial, and requires that we carefully consider our methods in the light of prevailing views in various implicated disciplines. As we will show presently, much of the controversy arises precisely as a result of the interdisciplinary nature of such a project. We will begin our search for suitable design tools by eclectically reviewing various published critical engagements with the notion of "musical universals". The intention here is not to present an exhaustive study, but rather to sketch the great diversity of extant approaches, contrasting their respective attitudes towards endeavours such as ours.

3.1 On the Search for Musical Universals

A cursory survey of the world's diversity of musical practices falls woefully short of any kind of chance distribution, prompting Lerdahl & Jackendoff to declare that "...[t]he only answer we find defensible is that one does not have to learn the entire grammar from scratch... In other words, much of the complexity of musical intuition is not learned, but is given by the inherent organization of the mind, itself determined by the human genetic inheritance." (Lerdahl & Jackendoff 1983:281) Clearly, some believe that there are fundamental attractors, however abstract, upon which cultural embellishments are developed – a theme which, though frequently downplayed, does nonetheless reoccur throughout the literature. Still, the *nature versus nurture* debate in music has historically been characterised by some rather fiery polemical exchanges (Baily 1996; Blackman 1996; Huron 1996; Sorrell 1996; Walker 1996b, 1996a). Further, even if one does subscribe to the notion of "musical universals", then what remains contentious is their location, whether wholly cognitive or perceptual, or some combination of both.[35] If apparently "innate" aspects of musical processing are physiologically hard-wired, then such musical universals would be expected to be relatively absolute, subject only to slow-moving physiological adaptations. This would imply limits on what could feasibly pass as music, and some would argue that this is already in evidence. If the pertinent apparatus is predominantly cognitive, then one might expect far greater dynamism and variety (which once again, some would say is already in evidence), constrained only by the adaptability of the human mind. In the latter case, why should one expect to find traces of universality at all?

---

[35] In general, a sensory or perceptual account favours the existence of universal musical percepts, while a cognitive account leans towards greater emphasis on learned or conditioned response. By a simplistic reading, the latter rules out "musical universals". Upon further consideration, however, even learned responses, driven by "universal" environmental stimuli, could feasibly result in "musical universals".

Historically, the physical principle most frequently invoked to bolster claims to universality of particular music phenomena is the natural harmonic series. However, Helmholtz rebutted by arguing that "... the construction of scales and of harmonic tissue is a product of artistic invention, and by no means furnished by the natural formation or natural function of the ear, as it has been hitherto most generally asserted." (Helmholtz 1954:365) Nonetheless, Lerdahl and Jackendoff review some of the impediments to a simple dismissal of such universality: widespread adoption of the principle of octave equivalence, including evidence of its effect on animals (Deutsch 1982:241–269), the frequent mistaking of perfect fifths for octaves, or simply the fact that overtones of a common fundamental are consistently judged consonant. On the other hand (a notable exception to the principle just cited), the perfect fourth was at one time judged as consonant within Western music, later as dissonant, though its position in the harmonic series remains unchanged. Further, the dissonance attributed to the perfect fourth (despite its position in the harmonic series) stands in contrast to the acknowledged consonance of the minor triad (which does not occur in the harmonic series). Various examples are cited (Lerdahl & Jackendoff 1983:291–292) of musical idioms employing intervals which cannot be reasonably derived from the harmonic series (which, for that matter, includes modern 12tet), though even these continue to employ the unison and octave. However, the notion of a tonal centre, of some pitch representing a point of maximum stability and repose, remains. "Thus, beyond the octave, the fifth, and perhaps the major third, it is difficult to make any useful connection between the overtone series and the universality of tonality." (Lerdahl & Jackendoff 1983:292–293) Instead, they find the approaches of Mursell, Partch, Johnston and others, stressing a mental rather than physical account, to be more satisfactory: "Perception of relative intervallic stability is a correlate of frequency ratios: the simpler the ratio, the easier to comprehend, or the more stable, an interval is." (1983:342, fn. 7) Johnston notably extended the same idea to the rhythmic domain.

Lerdahl and Jackendoff propose three properties which appear to be applied consistently across most (though not all) musical traditions. First amongst these is the selection of a particular configuration of pitches, most usually constrained by octave equivalence (e.g. Western scales or Indian thãts). Second is the designation of a "basis tone", and likely a secondary point of stability set off against this (as in the Western dominant, a fifth above the tonic, or the Indian samvãdi, a fourth or fifth above the vãdi). Third is a measure of relative stability amongst the various pitch choices. As an interesting example of the latter property, they highlight Western tonal music's parallel application of melodic stability (wherein the interval of a diatonic second is stable), harmonic stability (wherein the diatonic second is unstable), and harmonic progression via the cycle of fifths (which essentially balances the other two). They conclude that the pervasiveness of these three general characteristics outside of Western tonal music "... suggests that they form an important part of the cognitive organization with which listeners structure musical experience." (1983:296) Furthermore: "... the fact that certain trends appear among simpler tonal systems (for example, the frequent use of small-ratio intervals as points of harmonic stability, and the prevalence of some kinds of pentatonic scales over others) suggests the possibility of an innate system of preferences among tonal systems...” (1983:296)

Eugene Narmour's influential Implication-Realisation (I-R) model (Narmour 1990) explicitly partitions those factors presumed innate (termed bottom-up processes) from those presumed learned (top-down processes), originally framed by overt statements characterising the universality of the former. The I-R model has nonetheless contributed significantly and successfully to many computational models of musical tension or expectation, though predominantly within the idiom of Western tonal music (e.g. (Farbood 2006), (Margulis 2005)). A notable, relevant exception is the cross-cultural study of melodic expectancy as perceived by Pedi traditional healers listening to North Sami Yoiks (Eerola *et al.* 2009),[36] which employed the simpler two-factor version of the I-R model developed by Schellenberg (Schellenberg 1997). Eerola *et al* found that the I-R model modelled “...Gestalt-based expectancies with considerable success”, though not as robustly as statistical models did. They conclude: “As several data-driven models were successful in explaining the expectancies of listeners from various cultures, claims of cultural relativism … for musical processing are probably overemphasized.”[37]

Comparative studies of perceptual and cognitive temporal processes in individuals varying in age or musical training have been employed in a bid to “... tease apart the processes that appear to be 'innate' or 'hard-wired' (functioning at birth, determined by genes, independent of environmental influence and experience) and those that develop with maturation, acculturation (learning by immersion in the auditory world around us), or explicit training.” (Drake & Bertrand 2001:18)[38] Under the overarching theme that “temporal processing is limited by memory space and processing time”, Drake and Bertrand propose a tentative list of five temporal processes which address these constraints, and may thus merit consideration as musical universals: “[w]e tend to group into perceptual units events that have similar physical characteristics or that occur close in time” (2001:20); “[p]rocessing is better for regular than irregular sequences... [w]e tend to hear as regular sequences that are not really regular” (2001:21); “[w]e spontaneously search for temporal regularities and organize events around this perceived regularity” (2001:23); “[w]e process information best if it arrives at an intermediate rate”(2001:23); and “[w]e tend to hear a time interval as twice as long or short as previous intervals” (2001:24). Theirs is a speculative paper, though, an invitation to all researchers to engage on these and other similar issues in the light of emerging evidence.

A similarly propositional paper, albeit somewhat more extensively argued, is “In Time With The Music: The concept of entrainment and its significance for ethnomusicology” (Clayton *et al.* 2005). Entrainment is here defined as a process whereby two or more autonomous rhythmic processes (which might be oscillators) are seen to interact, each adjusting to characteristics of the other and ultimately “locking in” to a common

---

[36] This is an extension of (Krumhansl *et al.* 2000).

[37] The term “cultural relativism” implies an absence of “absolutes” or “universals”.

[38] We footnote the proposed distinction between “acculturation” and “enculturation” elsewhere, though it is clear from the context that the term “acculturation” is here used in a general sense, encompassing both meanings.

phase and/or periodicity. The theory goes beyond simple resonant induction, generally seeking its cause in weak, non-linear couplings. Though examples are given of physical and biological processes, ranging from Huygen's observations of pendulum clocks through the circadian rhythms of cyanobacteria, the authors also apply the concept at a more abstract level to inter- and intra-personal relationships in human societies (*social entrainment*). Rather than a wholly cognitive account of human motor theory though, theirs is a leaning toward a far more pre-attentive process, controlled by a neural gating mechanism. In other words, what we do in response to perceptual stimuli is less a function of positive will to action, and more a consequence of our ability (or inability) to selectively inhibit or control such action. This is a controversial stand, to be sure, but one that has been echoed elsewhere (Bargh & Chartrand 1999; Leman 2008). By their review of the ubiquity of entrainment in various fields, at least implicitly if not explicitly, and most notably in the work of ethnologists from Erich von Hornbostel and Curt Sachs through Alan Lomax, John Blacking, Charles Keil and Stephen Feld, Clayton *et al* argue that ethnomusicology might already be far better positioned to incorporate such empirical methods than has hitherto been admitted, and has much to gain from such an approach.

> "*Entrainment in musicking implies a profound association between different humans at a physiological level and a shared propensity at a biological level. The implications of this view for studies of socialisation and identification are obvious, as are the implications for questions of enculturation: someone's ability to respond appropriately to a given musical stimulus can, since it is a learned application of a basic biological tendency, be a marker of the degree to which an individual 'belongs' in a particular social group*" (Clayton *et al.* 2005:21).

The appeal to the ethnologist's imperative for social contextualisation is clearly evident. Perhaps less welcome is the baggage of "basic biological tendency" and the concomitant implications of universality. As Bruno Nettl reflectively puts it: "When I was a student I was taught that any attempt to generalize about the music of the world should be countered by an example falsifying that generalization." (Nettl 2001:463) The various responses to Clayton *et al*'s proposal (in the same volume) attest to a mixed reception. In some cases, the authors contend, they have been misunderstood. Yet other responses are clearly enthusiastic. John Bispham summarises succinctly:

> "*Put bluntly... ethnomusicologists have in recent times sought to explore music as a culturally constructed and cross-culturally diverse phenomenon, while cognitive musicologists or music psychologists, by investigating the biological determinants of musical abilities, have been, at least implicitly, concerned with the search for musical universals. This should not, however present us with an insurmountable barrier for cooperation and integration. It simply requires members on both sides to take on the intuitively sensible view that music is a culturally constructed phenomenon built upon universal biologically determined foundations.*" (Clayton *et al.* 2005:43)

Yet, even Bispham oversteps the boundaries (in the authors' view) in proposing a distinction between self-entrainment and interpersonal entrainment (Clayton *et al.* 2005:44). Jonathan Pearl similarly appeals for a distinction between physical and cognitive entrainment (Clayton *et al.* 2005:59–60). This, the authors contend, is contrary to essence of what they are proposing: a unitary phenomenon which plays out identically at various levels (Clayton *et al.* 2005:68). Mari Riess Jones contributes by warning of the methodological difficulties inherent in non-linear models, and reiterates the pivotal role of perturbation in identifying entrainment (Clayton *et al.* 2005:47–48). But it is Edward Large who arguably holds the lynch-pin.

Edward Large has advanced a compelling account of musical universals rooted in properties of non-linear resonance (Large & Tretakis 2005; Large 2006, 2010b, 2011, 2010a; Velasco & Large 2011), having demonstrated the utility of such a theory in both temporal and spectral organisation of music. Setting out a decade prior to Clayton *et al*'s paper from the somewhat more modest, but nonetheless provocative proposition that human neural rhythms resonate with musical rhythms, thereby giving rise to the perception of *meter* in *Western* tonal music (Large & Kolen 1994), Large's commentary on the entrainment proposal recognises the implication that "... nonlinear resonance underlies musical rhythm *in all cultures*... [and thus]... represents a substantive musical universal" (Clayton *et al.* 2005:52), a conjecture which remains central to much of his work since. An *Arnol'd tongues* bifurcation diagram illustrates the effect of driving a non-linear oscillator with a fixed, periodic stimulus while varying the natural frequency of that driven oscillator (see (Large 2010b:202, Fig. 4)). The results indicate an essentially infinite number of resonance regions corresponding to ratios of relatively prime integers. What distinguishes this account from typical small-integer-ratio accounts, though, is that the width of each region is a function of coupling strength. In other words, a plausible mechanism emerges to explain how nearby, more complex ratios might be perceived as functionally equivalent to simpler ratios. Large and other's mathematical analyses of resonator networks have revealed various "... generic properties of nonlinear frequency transformations, including extreme sensitivity to weak stimuli, sharp frequency tuning, amplitude compression, frequency detuning, natural phase differences, and nonlinear distortions... consistent with psychoacoustic phenomena, such as hearing thresholds, frequency discriminations, loudness scaling, Steven's rule, harmonic distortion and combination tones." (Large & Tretakis 2005:55) Both Large and Jones warn that Clayton *et al*, though arguing for a non-linear interpretation of entrainment, nevertheless draw on analytical methods which make strong assumptions of linearity (Clayton *et al.* 2005:47–48,53). Despite such criticism, and especially considering the more consistent systematic bases subsequently proposed by Large, entrainment does suggest precisely the kind of culture-free theoretical framework sought by this research.[39]

While Large has extended his "universal" hypothesis to simultaneously address matters of both rhythm

---

[39] By this position, we stop short of engaging on the more thorny issue of whether a mathematical basis can be construed as "culture-free". We acknowledge the importance of such a debate, but we must, of necessity, defer that matter to another forum.

and harmony, others have been decidedly more cautious in suggesting such parallels. Jeff Pressing is a notable exception (Pressing 1983), his group theory-derived views on the apparently "diatonic" template underlying the "standard pattern" in Ewe drumming, for instance, having attracted significant contemplation (Agawu 2006; London 2002). Further, Pressing explicitly links the ubiquity of polyrhythmic organisation in world music to "... perceptual entrainment effects in ritual, sometimes resulting in trance induction...", ostensibly as a result of "multiplicity of perception" (Pressing *et al.* 1996:1127). The parallels between harmonic dissonance and rhythmic dissonance, both defined in terms of the relative complexity of the ratios involved, are alluring, at the very least.

Studies under the banner of "temporal processing" do appear to be generally more amenable to the idea of "universals", and these are by no means restricted to matters of rhythm. In fact, the quest to fully understand human pitch perception must reconcile two competing theories, each apparently incapable of accounting for the entire phenomenon. These are the so-called "place theory", whereby pitch discrimination is a function of the point along the basilar membrane which resonates in response to a given stimulus, and the "temporal theory", according to which the pitch percept is extracted from the pattern of neural firings in the auditory nerve. Gerald Langner, for instance, has argued extensively in support of the latter (Langner 1997), without ruling out the contribution of the former (Langner *et al.* 1997). More recently, and in the light of emerging neurophysiological results, Langner has come out even more strongly in support of a neurobiologically grounded temporal theory, going further in asserting that: "... fundamental aspects of harmonic perception are by no means just culturally inherited. Instead the recognition of musical harmony, or consonance, is an intrinsic property of our brains." (Langner 2005:25)

A key impediment to the place theory is the phenomenon of the "missing fundamental", whereby the pitch percept remains when the fundamental is removed from a complex tone (Schouten *et al.* 1962), giving rise to various spectral domain based "template fitting" theories (Terhardt *et al.* 1982a, 1982b). Timothy Griffiths applied iterated rippled noise (IRN) in order to investigate the analysis of complex audio features and, in the course thereof, to weigh up these alternative hypotheses (Griffiths 2001). Focussing on "complex features likely to be important in musical analysis, below the level of semantic and affective processing" (2001:133), Griffiths considers evidence from functional imaging studies (PET, fMRI and MEG) as indicators of those cortical areas "normally involved" in temporal processing, while studies of patients with particular brain lesions suggested the "necessary" areas. The tenuous nature of the findings are largely down to the limits of the available technology, since non-invasive imaging techniques are subject to interference of neural activity in the various layers of the brain. This prompts Griffiths to "... declare an open mind about the existence of ascending processing systems for sound sequences such as music", since "... the ascending auditory pathway … affords an extensive mechanism for the processing of complex signals before the cortex is reached." (2001:134) If we understand Griffiths correctly, he appears to be suggesting a "bottom-up" model of musical processing, in contrast to the findings of others, on the basis of auditory brainstem response, for a "top-down" model (Marmel *et al.* 2010).[40] These conclusions are not necessarily contradictory, however, since Griffiths focussed on the ostensibly more elementary percept of pitch while Marmel *el al* dealt with the perception of harmony. Griffiths argues that the ascending pathway provides a form of temporal integration, the results of which manifest in the images produced by available imaging techniques. He further suggests that autocorrelation provides the most parsimonious mechanism for such integration (Griffiths 2001:139). In investigating cortical activation due to various pitch- and rhythm-based stimuli, he finds that "... direct comparison of the two conditions showed no significant differences in activation... consistent with a common mechanism for the two tasks at some level..." (2001:139)

Tramo *et al* ascribe the ubiquity of musical practises resembling Western tonal constructs as evidence of the requisite "universal competence" in auditory and cognitive functions, citing experimental results suggesting that "... similar perceptual attributes can be associated with similar emotions and social contexts across different cultures." (Tramo *et al.* 2001:93–95) Contrary to the prevailing view of "sensory consonance" (or the absence of "roughness") as a prime driver of musical perception (Terhardt 1984), they argue for the reconsideration of pitch relationships. This is informed by their own analyses of "...the responses of over 100 cat auditory nerve fibres to the minor second, perfect fourth, tritone, and perfect fifth." Briefly, their method is to examine the series of action potentials fired by each auditory nerve (its *spike train*), compiling an *all-order interspike interval (ISI) histogram* in respect of each. Combining these shows "... the ISI distribution in the entire population of auditory nerve fibers." (Tramo *et al.* 2001:101) The emergence of major peaks on such a graph (see (2001:102, Fig. 4)), matching the fundamental period of complex tones (periodicity pitch), is regarded as "... essentially the time-domain equivalent of Terhardt's spectrally based subharmonic sieve for virtual pitch extraction." (2001:101) Minor peaks are seen to form simple periodic patterns in the case of the fifth and fourth, while appearing quite aperiodic in the case of the tritone and minor second. A particularly interesting perspective on this view of the signal is that it appears to lend support to the notion of a "subharmonic series" of "undertones", an idea notably invoked in various derivative forms by Rameau (1971) and Hindemith (1961), amongst others, but since dismissed as fantasy. Further, Tramo *et al* are able to formulate a "... physiological measure of the strength of the fundamental pitch..." which is found to be highly correlated with "... previous psychoacoustic measures of the 'clearness' of musical intervals..." (Tramo *et al.* 2001:103) In sum, they make a provocative case in respect of "... the neurobiological foundation for the theory of harmony in its vertical dimension." (2001:103) Along the way, they show that such all-order ISI histograms closely resemble the autocorrelation of the waveform (see (2001:94, Fig. 1)), thus providing a non-invasive route to comparable signal characteristics. It bears noting that their study tacitly acknowledges the role of pitch relationships in the "horizontal" dimension of harmony (the sequencing of chords in time), seeking only to reaffirm the role played by such relationships in the

---

[40] Read *bottom-up* versus *top-down* as *innate* versus *cognitive*/*learned*, or in the limit, *nature* versus *nurture*.

"vertical" dimension (simultaneously sounding notes) (2001:113). Their findings provide a further stepping stone toward greater emphasis on such relationships in effecting directed motion in both domains.

Cook presented 72 three-note chords to 30 musically untrained undergraduates for subjective evaluation of "harmoniousness" on a six point scale, and subsequently extracted three significant factors from the data. (Cook 2001). The greatest of these, accounting for roughly half of the observed variance, was the presence of a minor or major second in the chord (an interval of one or two semitones, respectively). The second most significant factor, though, proved more interesting. Cook established measures of total theoretical dissonance by three methods: by employing only the three fundamentals (Sethares 1993); by including the five lowest harmonics (Kameoka & Kuriyagawa 1969); and by employing data obtained empirically from another of his own experiments. Despite his stimuli in this experiment containing no upper partials whatsoever, all three methods were shown to be necessary to best account for the perception of harmoniousness in the presented chords (Cook 2001:383). Taken alone, any one of these methods was able to account for only a small part of the observed variance. The third emergent factor was a new insight, what Cook termed "chordal tension". All of his three-note chords having symmetric structure (same interval between both inner pairs of notes), including their inversions, were found to be significantly less harmonious, though this effect was overshadowed by the first factor in the case of the two smallest intervals. He argues for a distinction between an "unresolved" chord and a "tense" chord (2001:383). Cook stresses the detection of dissonance and tension as potentially universal, deferring the particular nature of the resolution to cultural and individual differences (2001:384).

Though the broader program of scientific research into music, for all its apparent paradoxes, has commanded significant and sustained interest and effort, such has not taken place without some quite severe criticism of the futility of the task. Ian Cross surveys various extant views on the matter, finding that "[t]he consensual view from within the humanities appears to be that music is cultural rather than natural", and that "... a generalizable – and hence scientific – account is neither relevant nor possible." (Cross 2001:29) He goes on to ask: "If the very concept of music is so variable and inextricable from its cultural context, how can we expect to seek, far less find, its biological foundations?" (2001:30) A popular springboard for such debates is the difficulty of translating the term "music" (Nettl 2001:466), to which Cross adds observations regarding music in which listeners from different cultures disagree on the placement of the *tactus* or "beat". As he puts it, describing an encounter with a particular piece of music from Northern Potosí in Bolivia: "It always feels as though one is tapping on an offbeat." (Cross 2001:31)[41] Nonetheless, ethnologist John Blacking maintains that "every known human society has... music" (Blacking 1995:224), while Alan Merriam asserts that music "is a universal behaviour." (Merriam 1964:227; Cross 2001:32) Resolving such

---

[41] Cross's assertion is interesting insofar as he leads the reader to believe that such an incongruent perception is anomalous in music. In fact, setting up and obliterating such expectations has become something of a cliché in modern popular music production. Could Cross's example not have originated in much the same way?

ambiguity, Cross argues, requires a reformulated definition of music. He notes three initial requirements that should be incorporated in such a definition: indivisibility of movement and sound; multiplicity of reference and meaning; and the lack of immediate and evident efficacy. Still, taken together with the obligatory reference to temporal patterning of sound, such a definition fails to impress. (Cross 2001:33) Cross notes research by Sandra Trehub, Hanus and Mechthild Papousek, and Colwyn Trevarthen on the musicality of infants and caregiver-infant interactions, particularly the emerging notion of musicality as innate and domain-specific, though such claims typically require corroborating evidence of adaptive value in evolution, something that no-one has thus far been able to produce irrefutably. Thus, Dan Sperber describes music as an evolutionary parasite (Sperber 1996:142), and Steven Pinker calls it "auditory cheesecake" (Pinker 1999:534), stating that "... music could vanish from our species and the rest of our lifestyle would be virtually unchanged." (Pinker 1999:528) Still, Cross argues, such views are constrained by their own ethnocentric origins, having little significance outside of the commodified manifestations of music in the Western world. Further, "... music appears in these theories largely as disembodied sound oriented towards individual hedonism, a notion quite untenable before the advent of recording technology." (Cross 2001:36) A more compelling theme emerges from research focussing on human interaction, particularly within groups. Brown coins the term "groupish-ness", a "suite of traits that favor the formation of coalitions, promote cooperative behavior towards group members and create the potential for hostility towards those outside the group." (Cross 2001:37) In this context, Cross points out that "... music is consequence free in that it is not *directly* functional... the singularity of the collective musical activity is not threatened by the existence of multiple simultaneous and potentially conflicting meanings... a risk-free space for the exploration of social behavior..." (2001:38)

In arriving at the conclusion for music's "floating intentionality" or "transposable aboutness", Cross comes closest to our own agenda, particularly in the following comment: "Hence one and the same musical activity might, at one and the same time, be about the *trajectory* of a body in space, the *dynamic emergence* or signification of an affective state, the achievement of a *goal* and the *unfolding* of an embodied perspective." (Cross 2001:38) (italics added) Thus he formulates the following definition:

"*Musics are cultural particularizations of the human capacity to form multiply-intentional representations through integrating information across different functional domains of temporally extended or sequenced human experience and behaviour, generally expressed in sound.*" (Cross 2001:38)

With this, Cross characterises music as uniquely human, having likely "played a significant role in facilitating the acquisition and maintenance of the skill of being a member of a culture..." (2001:38) In particular, he highlights the sustained role played by music in "... individual and group encounters with the numinous and in the modulation of affective state" (2001:39), suggesting that it is precisely music's deictic intentionality, resonating with the simultaneous relevance and paradox so characteristic of both the metaphysical and the affective, which underpins this association.

Elsewhere, Cross distils his definition as follows: "music embodies, entrains and transposably intentionalises time in sound and action." (Cross 2003:108) Here he hastens to distinguish the capacity to *entrain* from the capacity to *mean*, though acknowledging the patent overlaps. Given the emerging view of music as intrinsically bound to human interaction, he questions the extent to which the exploration of neurophysiological correlates of the acoustic signal in individuals can be expected to give more than a severely limited, culture-specific account. Rather he advises a "triangulation" of music as both a biological and a cultural phenomenon (2003:109). What role then, if any, could be played by biology? Cross expects that cultural context conditions neurological correlates, likely shared in part with speech, but that the entire relevant neurophysiology would extend beyond the auditory pathways into the limbic system and centres of motor behaviour.

By taking the view that music is uniquely human, Cross avoids an emerging line of enquiry which precisely seeks to account for a possible adaptive function. Marc Hauser and Josh McDermott have argued for the value of investigations into animal behaviour in seeking answers to such questions (Hauser & McDermott 2003). Their approach postulates, for humans at least, an initial innate state of musical knowledge, transformed by experience to a mature state of musical knowledge. By bringing together evidence of protomusical behaviour in various species, they argue that these, at least, constitute an innate basis upon which human musicality has been constructed.[42] Rhesus monkeys, for instance, have demonstrated the capacity for octave generalisation of diatonic melodies. They did not respond to transpositions by anything other than a number of octaves, nor to octave transpositions of atonal melodies, further suggesting a privileged status for diatonicism, perhaps by way of a particularly parsimonious coding scheme (2003:665). Still, there are at least three foils contradicting such implications: human infants do not seem to extract musical keys automatically; humans generally do not discriminate between keys, or the effect is weak, at best; and behavioural similarities do not necessarily imply similarities in neural mechanisms (2003:666). The perceptions of consonant and dissonant chords have also been found to modulate neural responses in the primary auditory cortex of awake Rhesus monkeys, with "... the magnitude of the oscillatory phase-locked activity highly correlated with the extent of dissonance" (2003:666) In fact, the particular responses obtained were found to reflect the degree of "beating", the characteristic amplitude modulation resulting from interference between partials of the component pitches. The preference shown by 16-week old human infants for consonant, rather than dissonant chords then appears at some level consistent with the attribution of innateness for such percepts. To the claim that musicality is an adaptation affording the modulation of emotions, Hauser and McDermott mention converging observations from the realm of animal vocalisations: "... many submissive or affiliative calls tend to be harmonically structured...; attention-getting

---

[42] As the argument goes, analogue behaviours in different species may be attributed to the likely existence of prototypical behaviour in a common ancestor.

signals often have rising frequency contours; aggressive calls are often short, staccato bursts...” (2003:666) They report cross-cultural comparisons of human emotion perception in world musics having lately served up commonalities (primarily the effect of tempo, though other factors have been implicated), fuelling renewed speculation as to the existence of innate processing mechanisms. Another reported study compared the abilities of newborn human infants and cotton-top tamarin monkeys to distinguish between two languages. Both groups succeeded when the sentences were presented in the normal forward direction, while both groups failed when the recordings were played in reverse. Presuming that the responses relate to the salience of rhythmic cues in the respective languages (obfuscated by the reversals), this once again suggests similar innate perceptual abilities in both species, ostensibly evolved for purposes other than either language or music.

By extracting and comparing Hockett's “design features” of language with those of instrumental and vocal music and innate human vocalisations, it is possible to identify substantial overlaps as well as important distinctions (Fitch 2006:177). Fitch notes the scepticism of ethnomusicologists towards musical universals, in contrast to the attitude of linguists, but nevertheless proposes nine additional, ethnomusicologically informed design features which distinguish music from both language and innate vocalisations. Within this framework, he critically engages various instances of apparently musical behaviour in animals. A definition of “song” as “complex, learned vocalization”, for example, allows a distinction between bird, whale and human song, on the one hand, and that of frogs and crickets, on the other. Humans are thus the only singing primates, but not the only singers. Adopting a definition of instrumental music as “... the use of the limbs or other body parts to produce structured, communicative sound, possibly using additional objects”, Fitch finds such to be rare amongst vertebrates, though the best examples are found amongst our closest relatives (chimpanzees and gorillas) (2006:183). Hauser and McDermott reject such analogues, citing the extremely limited behavioural context of animal song, the sole purpose being communication, and that animal singing is predominantly a male behaviour (Hauser & McDermott 2003). Fitch contends that such arguments are based on inference about adaptive function, rather than empirical observation *per se*, that separate levels of biological explanation are here being conflated, and that the notion of singing as male behaviour arises from skewed emphasis on results obtained in temperate, rather than tropical species (Fitch 2006:184). Thus he underscores the notion of inter-species analogues in song as posited by Aristotle, Charles Darwin, Peter Marler and others. Of course, neither songs nor the organs required to perform them fossilise, but archaeological research has produced flutes dating back at least 36,000 years, along with various other artefacts which are plausibly regarded as instruments. Moreover, there is the somewhat more polemical find of a purportedly Neanderthal flute at the Divje babe I cave site in Slovenia, dated to at least 40,000 years ago (Kunej & Turk 2001). By implication, the origins of instrumental music are thus pushed back to the common ancestor of Neanderthals and modern humans, presumably the so-called Heidelberg Man, around 500,000 years ago. This may pre-date full spoken language (Fitch 2006:197).

Yet another approach to the question of musical universals finds its evidence in the statistical distribution of frequency deviations in inflected speech (Schwartz *et al.* 2003). A corpus of English language utterances, reflecting eight major dialect regions of the United States as spoken by 441 male and 189 female speakers, all of the same ten sentences (TIMIT), was randomly sampled in respect of each speaker to obtain a large number of 100 ms segments. Each segment was further normalised with respect to both the maximum amplitude in the segment, and the frequency at that amplitude. This particular method of normalisation sought to eliminate bias resulting from assumptions of harmonicity in the sampled segment (2003:7161). Amplitude-frequency combinations for each segment were then binned into histograms representing probability distributions and averaged per speaker. These were combined to obtain a normalised spectrum for the corpus as a whole. The same procedure, minus the grouping by individual speaker, was also applied to a second corpus comprising 1000 utterances in Farsi, French, German, Hindi, Japanese, Korean, Mandarin Chinese, Spanish, Tamil and Vietnamese (OGI Multi-language Telephone Speech Corpus). A measure of relative power concentration at each of the spectral maxima was obtained by regression analysis, with the residual mean normalised amplitude values being well fitted to a logarithmic function ($r^2 = 0.97$). A secondary measure was derived from the slope around each local maximum (2003:7161). Musical ratings of consonance were taken from seven of the empirical studies reviewed in (Malmberg 1918), and the robustness of these ratings assessed by reliability analysis treating each interval as an item and each study as a measure, yielding Cronbach's α of 0.97. Agreement between the studies was generally higher in respect of relatively consonant intervals, and lower in respect more dissonant intervals.

A key insight reached in this study is that power maxima generally lie at the third or fourth harmonic of the fundamental produced by the human vocal apparatus. This is significant because it explicitly accounts for the presence of frequency ratios corresponding to the *reciprocals* of 2, 3, 4, and 5, rather than simple integer *multiples* of the fundamental. Moreover, these relationships have been determined wholly in terms of the empirical data, and not by any *a priori* analysis of an idealised harmonic series (Schwartz *et al.* 2003:7163). Of course the cause of such is a matter of physiology, so it is not entirely surprising that the overall results show a marked consistency with respect to the location of maxima, prompting the view that "... the statistical structure of speech sounds... is a universal feature of the human acoustic environment." (2003:7163) The authors now set out to investigate three "widely shared phenomena in musical perception that require explanation in terms of the probabilistic relationship of auditory stimuli and their sources", these being the chromatic scale (as opposed to any other nominal subdivision of the octave), preferred use of certain intervals from that scale, and the broad consensus found on consonance ratings (2003:7164). Noting exceptions such as the Indonesian *pélog* scale (used in Gamelan orchestras), there is nonetheless a marked tendency amongst all the world's known musics to have used pitches apparently found in 12-fold subdivision of the octave now known as the chromatic scale.[43] The power spectra obtained previously are shown to contain maxima at precisely those ratios typically invoked to account for the intervals of the chromatic scale, with the exception of the minor second. The five greatest statistical concentrations are shown to correspond to the intervals of the pentatonic scale. Schwartz *et al* characterise the consonance/dissonance continuum as follows: "The more compatible of these combinations are typically used to convey 'resolution' at the end of a musical phrase or piece, whereas less compatible combinations are used to indicate a transition, a lack of resolution, or to introduce a sense of tension in a chord or melodic sequence." (2003:7165)[44] Using two different measures of power concentration, being the residual amplitude at each local maximum, as well as the slopes at these maxima, consonance is shown to be highly correlated with the relative concentration of power in human speech at the associated ratio (2003:7165).

Though one might all too easily dismiss Schwartz *et al*'s approach as merely a revamped take on the canonical harmonic series explanation, their approach stands apart from that tradition by its emphasis on learned pattern. This is of course immediately reminiscent of Terhardt's harmonic template (Terhardt *et al.* 1982a, 1982b), there also proposed with reference to the spectral properties of speech, and partial precedents are acknowledged in (Kameoka & Kuriyagawa 1969) and (Sethares 1998), though all of these remain closer to a psychoacoustic account of consonance in terms of their reference to physical interaction between the partials (Schwartz *et al.* 2003:7166). Learning has also been central to other theories (Tillmann *et al.* 2000; Jonaitis & Saffran 2009), but then without explicit reference to the spectral properties of speech. Speaking to a possible adaptive function of musical percepts, Schwartz *et al* engage the inherent ambiguity in audio stimuli, the physical characteristics present at the ear being incapable of specifying physical characteristics of the source, suggesting that a probabilistic processing strategy best serves the biological imperative to respond appropriately to such ambiguity.

An extension of the implicit learning hypothesis manifests in a study aiming to establish to what extent musical percepts depend on formal training (Bigand & Poulin-Charronnat 2006). Upon reviewing extant research, it is shown that apparent distinctions between musicians and non-musicians have been largely overstated, and may be accounted for in terms of methodological flaws (2006:102–104). Chief amongst these flaws are ambiguous formulations of experimental tasks, tasks requiring explicit judgements, the use of technical terms, and the use of protomusical stimuli rather than real musical material. The first experiment in this study compared the ability of musicians and non-musicians to identify musical variations on a theme, yielding better than chance results in the case of non-musicians, though lower than those of musicians. This was cited as evidence of non-musicians' ability to parse the implied harmonic structure. Though the second

---

[43] Of course, there is the nagging question of precisely whose chromatic scale is being referred to here, but we will not allow that issue to detract from the thrust of the argument.

[44] The authors use the term "compatible" without further qualification, but we may infer that they are referring to the subjective sensory quality of certain "combinations" of pitches "belonging together", and others less so.

experiment required participants to make explicit judgements about musical stability (specifically, an evaluation on a seven-point scale of stability in respect of each note of a melody), the researchers felt that the underlying percept of "completion" was sufficiently general to not bias either group. Once again, results for both groups were significantly correlated. Furthermore: "Multiple regression analyses further revealed that pitch factors (tonal weights, melodic interval sizes) and rhythmic factors (durations and metrical weights) significantly contributed to perceived tensions in both groups, with the place of the notes in Western tonal hierarchy (tonal weights) as the most important contributor." (2006:107) A follow-up task sought to quantify sensitivity to changes in induced harmony and/or rhythm, and demonstrated once more how slight the differences between the musical perception of musicians and non-musicians actually are. In particular, both groups were required to judge "how many notes had been changed" in response to a manipulation of mode (thus a change in implied harmony),[45] rhythm or both (2006:107). Again, the results of both groups were highly correlated, with changes to both mode and rhythm ranking as most effective. More interesting, though, is that the ratings given (as high as 83.87% for musicians and 69.87% for non-musicians) had little bearing on the actual number of changes, which likely would have been less than 20%. This simply points to an illusory percept, which Bigand and Poulin-Charonnat offer as evidence that "... Western tonal melodies are mentally represented by an abstract structure, which typifies the principal dynamic trajectories developed during listening." (2006:108) They suggest that "... these trajectories seem to be based on the relationships of musical tensions and relaxations that span different levels of musical time." (2006:108) Noting similar correlations between the results obtained in respect of musician and non-musician participants in studies focussing on chord sequences, they conclude that musical tension perceived by Western listeners, whether musically trained or not, may reliably be predicted by the same factors.

Trained musicians also seemed no better off than non-musicians in a task designed to test comprehension of large-scale musical structures (2006:115). Interestingly, non-musicians were found to be equally adept to musicians in detecting that two atonal canons had been composed on the basis of the same dodecaphonic tone row, to which Bigand & Poulin-Charronnat reply: "... if both groups have the same ability to acquire knowledge about these new idioms, this observation would reinforce the claim that a sophisticated musical competence exists in human beings, and that it is rather independent of intensive explicit training." (2006:115)

Affective response to music remains an important line of enquiry. It is widely presumed that musicians are better able to discriminate subtle difference of expression, and thus experience a wider variety of affective responses, but that the strength of those responses found in both musicians and nonmusicians is

---

[45] Bigand & Poulin-Charronnat termed this a change of key, though the pitches of the melody were minimally modified only to satisfy the new key signature. Thus a change of "mode" seems a more apt description. In any case, their method implies a strong bias to perceive the altered melody in terms of the new key, rather than in terms of the mode. We are not entirely convinced that this is necessarily so, and the matter invites further research.

comparable (2006:117). In seeking to avoid the linguistic confounds inherent in explicit tasks, various methods have been proposed, including multi-dimensional scaling (MDS) techniques (Bigand *et al.* 2005).[46] Participants were asked to group together musical excerpts which induced similar affect, thus being encouraged to focus on their own emotional responses without the need to verbalise such (Bigand & Poulin-Charronnat 2006:118). Given 27 30-second excerpts, both musicians and non-musicians distinguished eight groups on average, contrary to the expectation of finer discrimination in musicians. Intergroup comparison, by way of 27 x 27 matrices of co-occurrence, showed high correlation which, taken together with the results of MDS, once again suggested no significant difference in emotional response. Reducing the length of the excerpt to only one second had a weak effect on emotional response, but none on either the number of groups or the content of those groups.

Contradictory evidence might be found in a line of enquiry which has pointed out neurophysiological differences between the brains of musicians and non-musicians (e.g. (Schlaug 2001)), though Bigand & Poulin-Charronnat argue that “[s]ome of these differences are obviously associated to the learning of motor skills involved in the playing of a musical instrument.” (Bigand & Poulin-Charronnat 2006:123) Overall, such differences are regarded as slight considering the extensive musical training involved. They conclude that, while intensive training affords the acquisition of skills required by professional musicians, possibly including a degree of neurophysiological adaptation, “... it is not what determines the musical ability of human beings.” (2006:126)

Catherine Stevens advances cross-cultural studies as essential to testing the generality of contemporary musical theories, critically engaging different traditions of musical thought, and increasing cultural understanding (Stevens 2004), further suggesting that “[c]omplementary trans-disciplinary approaches may also minimize bias from a particular ethnocentric view.” Of course, the notion of science as “value-free” has long been challenged, but Stevens elects to defer that debate. Her definition of “cross-cultural” explicitly invokes the “...search for general mechanisms or universal phenomena.” (2004:433) Against the backdrop of a presumed universal deployment of discrete pitch levels, between five and seven, unequally spaced, to the octave, she first reviewed studies probing the ability of listeners to engage meaningfully with unfamiliar intervals, scales and tunings, finding evidence that musical training might have facilitated tuning judgements in unfamiliar material (2004:434). Studies of Western infants, employing the “operant-head-turn-procedure”, suggested that a culture-specific perceptual reorganisation, relevant to musical tuning, takes place between the ages of six and twelve months, and that the performance of the youngest infants might suggest processing based on the semitone. Modal perception was examined in the context of Arabic improvised music (*taqsim*), specifically by way of “... tasks such as identification of musical elements, segmentation of the work, verbal descriptions and performed melodic 'reductions' of the segments.” (2004:434) While both European and

---

[46] Review (McAdams *et al.* 1995) for application of MDS in the auditory domain, as well as (Plomp 1976) and (Wessel 1979).

Arab listeners segmented in response to salient surface features such as pauses and register changes, Arab listeners were furthermore found to be sensitive to emblematic “melodico-rhythmic configurations” and subtle modal changes that went unnoticed by European listeners.

The interplay of schematic and veridical expectations, shaped by metric and rhythmic pattern, as well as by tonal and melodic structure, is widely understood to afford music its emotional significance. There is thus great interest in the extent to which canonical schematic models, such as Narmour's I-R model or Krumhansl's tonal hierarchy, generalise. The predictions of the I-R model have been shown to hold up in the context of diverse musical materials, “... including British folk songs, Webern lieder, and Chinese pentatonic songs” (Stevens 2004:434). North Sami yoiks were found to not comply with the principles of Western tonal music, and so, to the extent that “outsiders” were able to assimilate certain of those characteristics, Krumhansl concluded that “... there is a core set of psychological principles that underlie melody formation” (Krumhansl *et al.* 2000:13–14), augmented by listeners' sensitivity to, and assimilation of the “... statistical distribution of tones and higher order statistics such as two- and three-tone transitions.” (Stevens 2004:435) Though the bottom-up principles of the IR model were initially proposed as potentially innate, self-organising artificial neural network models have succeeded in extracting the same principles from the music itself.[47] The emerging view of human cognition as an on-line statistical process, with particular sensitivity to tonal hierarchies,[48] has found support in the context of Indian, Balinese and Korean Court Music. Stevens reminds us, though, that “universality does not imply innateness.” (2004:435)[49] Infant and developmental studies have proposed various candidate musical universals. Five-year-old children, for instance, have been shown to be affected by key membership, but not by implied harmony. The latter becomes evident in the performance of seven-year-old children as in adults. Thus key membership is proposed as at least potentially universal, and implied harmony as most likely not. In fact, says Stevens, implied harmony is cross-culturally rare (2004:435).

The study of rhythmic organisation in African music has frequently fallen prey to the misguided application of Western music theory and cognitive models, as illustrated in Vijay Iyer's example of a West African standard bell pattern and its likely interpretation by three different cultural groups. (Stevens 2004:435; Iyer 1998) The Western bar-line (whether explicit or by implication) seems to be the most obtrusive imposition, obscuring, by Iyer's account, the “... unambiguous but culturally specific... suggestion and complexity” whereby “[t]he metre is encoded in the rhythm itself.” (Stevens 2004:435) Jeff Pressing's cross-domain linkage of structural principles in both pitch and rhythm, which he terms “cognitive

---

[47] As the argument goes, being able to extract purportedly innate principles from the stimuli by a process of learning invalidates their innateness. Though this reasoning has merit, it is far from conclusive.

[48] Ideally, these hierarchies a not specified *a priori*, but rather emerge from the learning process. In the case of the MUSACT model, important structural features are, in fact, specified *a priori*.

[49] And for that matter, innateness does not imply universality. Yet, in both cases, we are inclined regard a single counter-example as sufficient grounds to dismiss such a proposition.

isomorphisms", are common cyclic structures constrained by octave equivalence and "... the perceptual equality of smallest intervals" on the one hand (the domain of pitch), and "... repeating isorhythms based on a uniform fastest unit" on the other (the domain of rhythm) (Pressing 1983). In terms of these, "... all of the Western church mode isomorphisms of scale and time lines occur in West Africa..." (Stevens 2004:436). On the other hand, Clayton looked at the concept of *tal* in North Indian (Hindustani) music, concluding that "... while metre is not a factor in all music, neither is it restricted to the West." (Stevens 2004:436) Stevens goes on to observe Western music's systematic avoidance of the perceptual ambiguities inherent in polyrhythms, a significant feature of much African music and likely valued for its ability to exploit entrainment effects. The effect of a given culture's spoken language on its music has also been found to be potentially significant. Patel & Daniele reported that normalised Pairwise Variability Indices (nVPI) showed parallel differences between the rhythms of English and French classical music themes, compared to the rhythms of spoken English and French (Patel & Daniele 2003).

Stevens, too, senses the threat of Western cultural encroachment. She, like Huron, believes that massive ethnomusicological archiving initiatives are simply not enough, since "... relatively few focus on cognitive processes and employ the necessary tools afforded by rigorous experimental design and analysis." She proposes that "[r]esearchers will need to be creative in applying a range of interdisciplinary methods including ethology and ethnography, analysis of musical materials, through to rigorous and ingenious design of experiments that test particular hypotheses developed in the field." (Stevens 2004:437) To the question of universals she proposes the following candidates: "...elementary auditory grouping strategies, a stable reference pitch, the division of an octave into scale steps, the use of reference pulses, and the induction of rhythmic patterns by an asymmetrical subdivision of time pulses." (2004:437) All musical cultures are thus, by implication, presumed to have arisen from the elaboration of a few such universals.

Bruno Nettl traces the waxing and waning of interest in musical universals within ethnomusicological circles, particularly with reference to the rise of what has been termed "paleomusicology", the speculative study of the origins of music (Nettl 2001). Recalling two special ethnomusicological journal issues from the 1970's devoted to the subject of musical universals, he reports that virtually all contributors remained decidedly sceptical toward the notion. More recent trends in linguistics and social anthropology, however, seem to have rejuvenated interest herein. Nettl revisits the obligatory starting point of attempting to find a common term for music, reviewing the preponderance of bifurcations such as *Musik* and *Tonkunst* in German, *Muzika* and *Hudba* in Czech, and *Musiqi* and *Khandan* in Persian culture. At the other extreme are single terms which encompass music in a greater scheme of activities (singing, movement, ceremony, etc.) include *saapup* amongst the Niitsítapi or *nkwa* amongst the Igbo. Nettl considers a definition of music as a form of sound communication distinguished from ordinary speech, but even this definition is challenged by the existence of some intermediate forms (Nettl 2001:466). He must ultimately settle for a vague notion of "musicness", as distinct from "speechness", along with the peripheral notion that one is always singing or playing "something" (a song, a work, something having a distinct identity),[50] as opposed to, say, just dancing.

Nettl is considerably more reluctant to engage widespread musical phenomena, to which exceptions nonetheless exist, as universals. "We are of course reduced to playing games. Accepting the idea of statistical universals means abandoning the principle that there is a significant difference between universality and popularity." (Nettl 2001:467) However, he continues: "... we have had to admit that some things... are enormously widespread." (2001:468) He summarises such characteristics as follows:

"*All societies have vocal music. Virtually all have instruments of some sort, although a few tribal societies may not, but even they have some kind of percussion. Vocal music is carried out by both men and women, although singing together in octaves is not a cultural universal, perhaps for social reasons. All societies have at least some music that conforms to a meter or contains a pulse. The intervallic structure of almost all musics involves, as the principal interval, something close to the major second but to be sure, not with precision; I am talking about anything, say, from a three-quarter tone to five quarters. All societies have some music that uses only three or four pitches, usually combining major seconds and minor thirds.*" (Nettl 2001:468)

To these, Nettl adds the "... importance of music in ritual, and... in addressing the supernatural", shared by all known societies, as is "... the use of music to provide some kind of fundamental change in an individual's consciousness or the ambience of a gathering." Also, he notes the use of music to mark important events, and to accompany virtually all forms of dance. (Nettl 2001:468) Considering emerging findings in respect of animal musicality, Nettl acknowledges another onslaught on the assumptions traditionally held within ethnomusicology:

"*Is music a characteristic of Homo sapiens alone? Most ethnomusicologists probably think so, I have to confess; but... the taxonomy that we Western observers are hesitant to impose on non-Western cultures is possibly valid for other species. Once established, such a theory might require ethnomusicologists to change their definitions and approaches.*" (Nettl 2001:470)

Still, there remains the theory that music might have developed as any number of independent strands, and that such patent similarities are merely incidental. Curt Sachs, for instance, held up a distinction between "logogenic" and "pathogenic" (likened to Apollonian and Dionysian) origins (Sachs 1943:41,52). However, it is not hard to see how such a distinction might have been employed to monopolise aesthetic discourse at some point in time, and so, considering some of the paradigmatic shifts in world view over recent decades, it is hardly surprising that we seldom hear such talk from post-modern commentators. Nettl

---

[50] Presumably, Nettl means this to include "free" improvisation, on the basis that even such is driven by some conceptual entity (such as particular thematic or harmonic material), at least if it is to be received as music.

recognises the potential feelings of hopelessness that might be evoked by this debate, but is nonetheless positive about the value of greater exchange between biomusicology and ethnomusicology. Most significantly, he concedes that "... universals do exist in musical sound and in musical conceptualization and behavior", adding that "[t]hose that involve musical style are at best statistical" (Nettl 2001:472). The casual reader might miss the innuendo in the latter remark.[51]

Clearly, the matter of the "musical universals" is far from being resolved. On the one hand, there is flat-out denial of the very possibility (Walker 1996b). On the other, considerable effort is being applied to establish physiological and ethological bases upon which to build precisely such a view (Tramo *et al.* 2001; Hauser & McDermott 2003; Schwartz *et al.* 2003; Fitch 2006). In between lie various extensive programmes of theoretical and empirical research which grapple with the more enigmatic percepts of musical cognition (Deutsch 1982; Krumhansl & Kessler 1982; Lerdahl & Jackendoff 1983; Narmour 1990). On the weight of the evidence we have encountered, the prevailing view seems to suggest that music has something of the nature of the mythical Ouroboros: it perpetuates by devouring its own tail. In other words, music is created to engage our conditioned responses, but those same responses are conditioned by the music in the first place. This interpretation allows us to neatly dispense with any kind of comparative study of the world's musics. The notion of music as a "universal language" was neither "a lie", nor even "a damned lie"... it was merely "a statistic" (Velleman 2008). Yet, there remains a recurrent, if downplayed tendency to acknowledge the possible influence of innate predispositions. Might such be revealed by Fitch or Hauser & McDermott's studies of the musical behaviours of other species? Could Trehub's and others' studies of infant musicality hold the key? Will Large's non-linear resonances be found operating in the auditory cortex, arguably a physiological validation of Clayton *et al.*'s entrainment proposal, or do the answers lie in the ascending auditory pathways as postulated by Griffiths? We have omitted, in the foregoing, to review one particular experimental paradigm which has come to our attention, specifically because it holds the potential to produce some of the insights we require. An extensive review of "musical priming" is included as Appendix B, and key results are summarised in section 3.6 of this chapter.

3.2 Re-Modelling the Perception of Directed Motion

In reviewing extant engagements with directed motion in Western music, albeit in terms of expectation (Meyer), tension (Farbood) or implication (Narmour), we find that there is a predominant tendency to view such as a spectral phenomenon. In other words, particular combinations of pitches, either by virtue of their own internal "dissonant" structure, or by their relationship to a larger implied "key" structure, evoke expectations of what should follow, and these expectations might then be satisfied, deferred or denied. A broad review of theoretical approaches to African music (Appendix C, see summary in section 3.6) has

---

[51] It is likely that two readers, possibly one from the humanities and one from the sciences, might interpret the label "statistical" in quite opposite terms.

turned up little that is immediately comparable. However, notable accounts have emphasised the functional importance, in African music and its derivatives, of polymetric structures. That is to say, such music may be found to simultaneously group, temporally speaking, in two or more distinct ways. The relationship between these groupings, particularly the listener's experience of shifting phase toward some future point of alignment, is then believed by us to be an important constituent of the sense of forward motion in such music.

On the face of it, there might seem to be little to link these accounts, particularly in view of a Western conceptualisation of harmony and rhythm as separate. However, both of these "domains" are defined with respect to time, and the same simple superparticular ratios found to be operating in polymetric music have clear parallels in the construction of harmonic intervals, as well as to historical notions underlying harmonic function in Western music. An important distinction nevertheless remains herein, that harmonic intervals find their resolution in other, simpler harmonic intervals, while polymetric phrases apparently resolve at the point of phase coincidence. However, the absence of greater similarity on this count, as reflected in the literature, does not exclude the possibility that we have hitherto overlooked more comprehensive parallels operating at other scales in the music. It is precisely such speculation which fuels this particular investigation.

Western conceptual baggage is evident in much MIR research, there being widespread, implicit acceptance of an axiomatic separation between melody, harmony and rhythm, privileging of 12tet pitches and divisive mono-metrical structure. A conventional approach to the problem of identifying directed motion might thus proceed by extracting symbolic pitches, analysing such in terms of their implied ternary harmonies, and interpreting each harmonic structure, locally and globally, in terms of overall progression towards a tonal centre. Pitch detection in polyphonic music, however, remains a challenge. Thus many such endeavours simply dispense with the problem by employing higher level symbolic data, such as MIDI-encoded events, and thereby tacitly incorporating all of its concomitant axioms. The approach to be taken here finds its provenance in periodicity, and particularly in the transaction between multiple periodic phenomena. As such, no distinction will initially be drawn between periodicity resulting from timbre, harmony or rhythm, beyond noting the timescale of that periodicity. Instead, greater emphasis will be placed on recognising harmonic relationships between such simultaneous periodicities, and in formulating a general, cross-cultural interpretation of the perceptual implication thereof.

3.3 Harmonic Mapping

Taking stock, we currently have a vague notion of the kind of patterning we hope to expose in music. Fundamentally, we expect it is characterised by the principle of *harmonic series*, being that series obtained by extending some reference value by an arbitrary number of integer multiples of itself. However, various natural or cultural constraints might have necessitated deviations from the ideal, and so we require a mechanism with sufficient "tolerance" to be able to rationalise apparently irrational relationships in the signal. We also intend, as far as is practicable, to apply uniform analytical techniques in both spectral and temporal domains. More important still, we will appeal to our own novel interpretation of the harmonic series as a prototype of "directed motion".

As discussed earlier, "directed motion" has been strongly associated with the "resolution" of a given ratio to the closest "simpler" ratio by "contrary motion". Consider then a prototypical sawtooth wave having an arbitrary fundamental frequency f and harmonics $2f$, $3f$, … nf. Shifting the fundamental by some factor x would linearly expand or contract the same series to become xf, $2xf$, $3xf$, … nxf. Seen simply as an injective mapping, nothing comes to light. However, considering the importance attributed to pitch proximity in our earlier discussions (of (Huron 2001), cf. our pp. 23 - 25), we might rather view this as a mapping between "closest" components of the respective series (by Dirichlet's drawer principle, a surjective mapping). The figures below indicate these two different perspectives. In Figure 3.1, we simply plot adjacent fundamentals, together with their harmonics, as points along a frequency axis. As might be expected, any particular harmonic may be traced out as a ray extending diagonally away from its origin. However, one can hardly miss the parabolic paths which also appear.

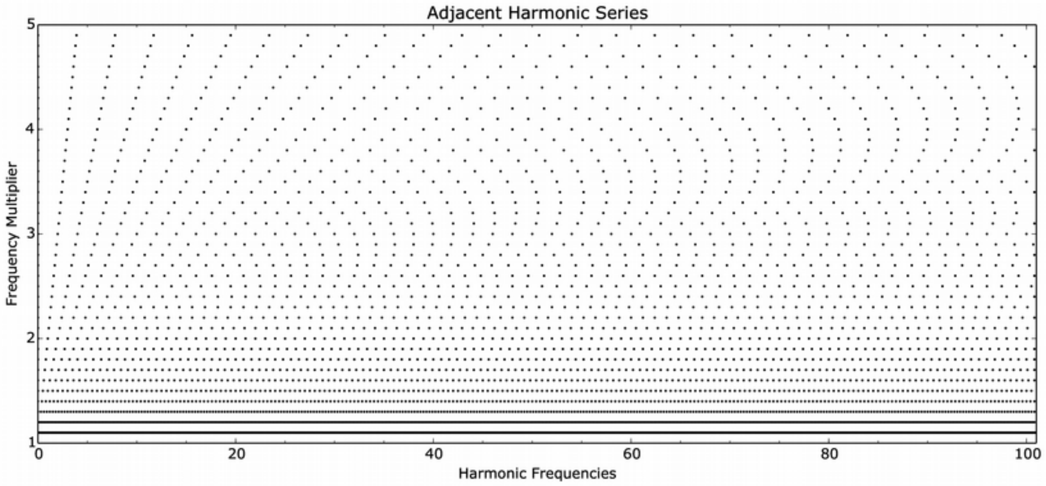


Figure 3.1: Comparison of adjacent harmonic series

Analysis of these patterns immerses one in number theory, and we have scoured a range of sources in search of an appropriate means of expressing and exploiting these patterns, but to no avail. Clearly, the turning point of each parabola marks the point at which harmonic n of one fundamental finds itself suddenly closer to harmonic n-1 of the next, adjacent fundamental. Our intuition is that such transitions might mark important musical intervals. In particular, we hope to find a reasonable ordering whereby to account for the particular topology of tonal space, being that sounds that are adjacent on the piano keyboard are most distantly related, whilst sounds that are separated by a much greater distance are perceived to be closely

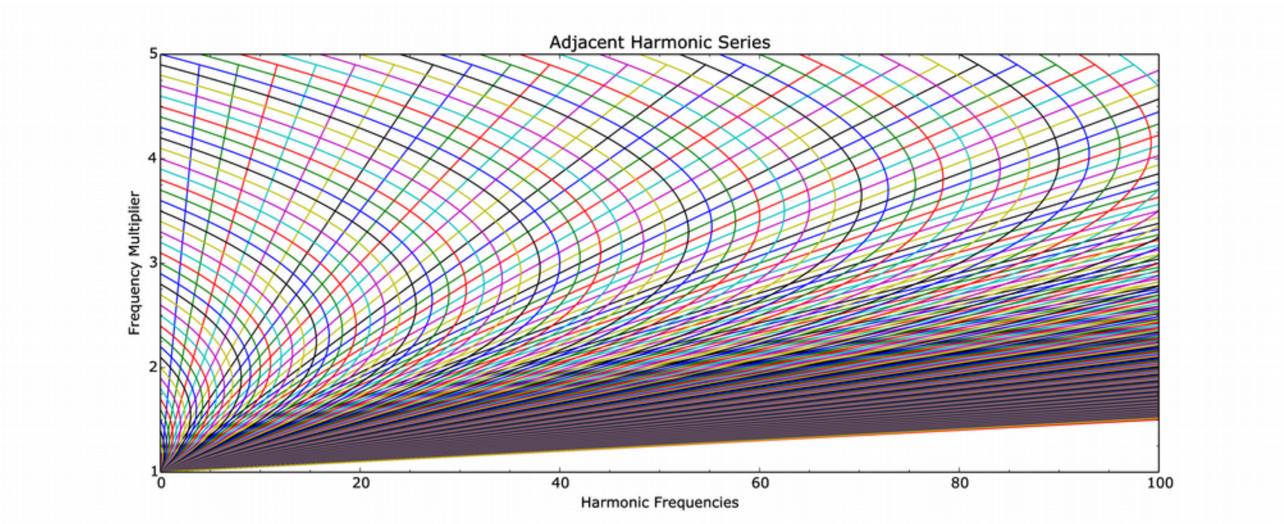


Figure 3.2: Analytic representation of harmonic mapping

related. Both rays and parabolas may be expressed analytically, as shown in Figure 3.2. Closer comparison, of course, will reveal that these are not the only patterns occurring in the first plot (two more sets of parabolas are evident, at least). What is more, the patterns which do appear here are dependent on the particular ordering of the various component series along the *y*-axis. Still, the parabolic expressions, in particular, suggest that there may be an alternative way of expressing distance between musical sonorities that does not rely on the Western music theory's axiomatic "cycle of fifths". Consider, for instance, the smoothest mapping between various diatonic chords (these in C major), together with nearest-neighbour mapping between all harmonics which might emerge from each of the component fundamentals, as shown in Figure 3.3. The chord progressions shown are quite arbitrarily selected to illustrate that, while Western music theory concerns itself with "voice-leading" in terms of the frequency proximity of the component

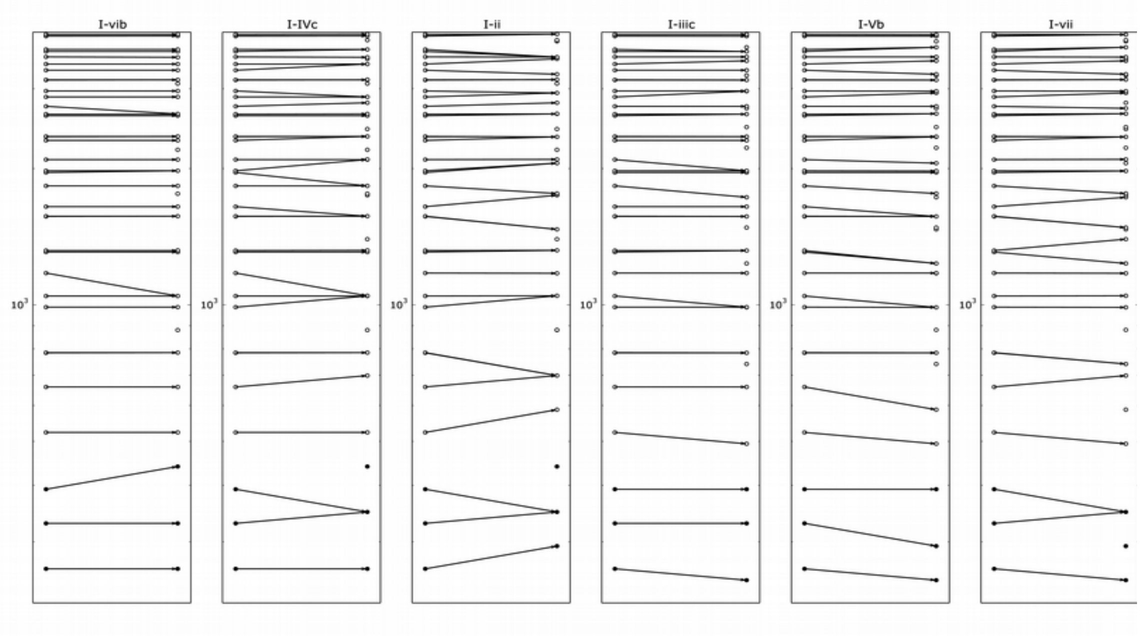


Figure 3.3: Nearest-neighbour mapping between components of selected chords

fundamentals (here indicated with filled circles), we intend to extend that concern to also incorporate harmonics (which we have indicated with unfilled circles). Clearly, the nature of such a mapping does not appear to mirror the conventional voice-leading mapping in any simple way.

3.4 A Preliminary Distance Measure

What is then required is to formulate a distance function which expresses the "smoothness" of the mapping presented above, and to compare this for various values of *x*, the interval shift. The selected measure, for the moment, is a simple sum of the quotients of each element in the first sequence and its closest element in the second sequence. Thus:

$$d(A,B)=\sum_{n=1}^{\infty}\left[\frac{n}{x}\right]\frac{x}{n}-1 \tag{3.1}$$

where the square brackets represent the "nearest integer" function, $n\in\mathbb{Z}$ denotes the sequence of integer-multiple harmonics, and $x\in\mathbb{R}$, the interval shift, is here varied from ½ to 2 (down an octave, to up an octave). Since the values obtained vary about 1, the subtraction normalises the result. Consider Figure 3.4.

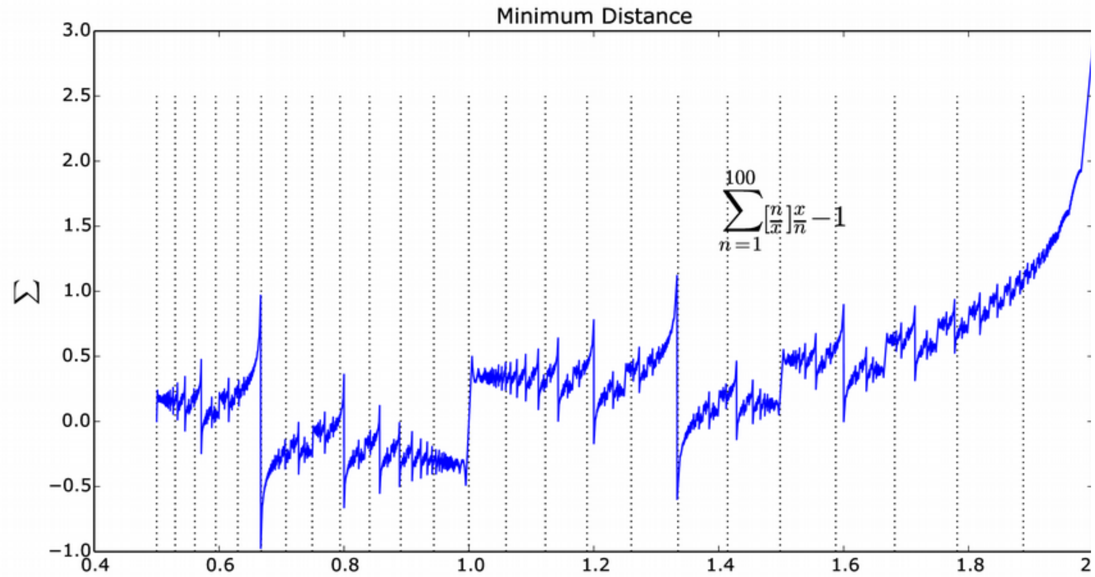


Figure 3.4: Distance between harmonic series, offset by interval x

Dashed vertical lines indicate the intervals of equal temperament, each line being separated by a 12tet semitone from the next. As *x* varies, there are clear minima in terms of this distance function. In the lower octave, 0.66 and 0.8 stand out (2/3 and 4/5, respectively). These would appear to correspond in Western music theory to a downward movement of a perfect fifth and a major third, respectively. However, while the equal-tempered approximation of the downward fifth manages to capture something of its idealised minimum, the downward third completely misses the mark, instead coinciding with a relative maximum. Should such a progression impart a sense of "directed motion" (as Tenney contends, see (Belet & Tenney

1987)), then it would presumably not be apparent in equal temperament. To our knowledge, Western music theory does not view such a progression as instrumental in effecting "directed motion".

The minima in the upper octave again indicate the significance of the ascending perfect fourth (complement of the descending fifth), minor third and minor sixth (the latter the complement of the major third indicated above). Once again it can be seen that 12tet intervals miss the minima, except in the case of the ascending perfect fifth.

Ideally, we would now proceed to an analysis of the selected distance function. However, the rounding or nearest integer function ([ ]) renders it undifferentiable.[52] In the particular case chosen (a simple harmonic series), the deviations do form a clear linear pattern. However, it would be highly desirable to be able to generalise the method to compare other real-world structures, such as the stretched harmonic series of many instruments, or perhaps those which result from the interaction between transversal and longitudinal modes of vibration (e.g. vibrating bars). The method must also expand to accommodate an arbitrary number of simultaneous fundamentals. Ultimately, it is hoped that this technique will extend to the time domain, perhaps bringing a parallel insight into matters of musical rhythm.

3.5 Comparable Approaches

The observations above support the adoption of a "Euler space" -based approach to the musical analysis. The values of x found significant above are all reducible to simple integer ratios, or products of small primes raised to small integer powers. Decomposition of arbitrary intervals to points on a discrete lattice are expected to better express perceived musical "relatedness". Consider a three-dimensional space with the axes representing the primes 2, 3 and 5. All spectral components related by octaves would lie in a line along the "2" axis ($1:2^n$). Those related by a number of "Pythagorean fifths" would lie on a line cutting the plane between the "2" axis and the "3" axis ($3^n:2^n$). Those related by "Just major thirds" would similarly lie on a line cutting the plane between the "2" axis and the "5" axis ($5^n:4^n$). By way of comparison, the "Pythagorean major third", which differs from the "Just major third" by approximately 22 cents, would lie on the same plane as the "Pythagorean fifth" ($81:64 = 3^4:2^6$).

If we accede to the contentions of Tenney, Barlow and others that listeners routinely abstract irrational relationships between musical elements to rational ones, then we need to find a suitable method for doing so here. Barlow's approach is as follows (Barlow & Lohner 1987):

- He proposes an *indigestibility* function, expressing the "simplicity" of a numerical relationship:[53]

---

[52] In fact, the problem is Diophantine, has no closed solution and is recognised as NP-hard. It is a problem of integer linear programming, for which only iterative solutions exist.

[53] Barlow's *indigestibility* function would appear to have been inspired by Leonard Euler's *Gradus Suavitatis*.

$$\xi(N)=2\sum_{r=1}^{\infty}\left\{\frac{n_r(p_r-1)^2}{p_r}\right\} \tag{3.2}$$

where

$$N=\prod_{r=1}^{\infty}p_r^{n_r};$$

p is prime; and

n is a natural number.

In other words, N is factored to primes, and those primes and their exponents all contribute to the calculation of the indigestibility value. Barlow notes that $\xi(ab)=\xi(a)+\xi(b)$.

- He proposes that, for any interval P:Q, indigestibility values be combined to determine the *harmonicity* of the interval:

  $$h(P,Q)=\frac{\operatorname{sgn}[\xi(P)-\xi(Q)]}{\xi(P)+\xi(Q)-2\xi(\text{hcf}_{P,Q})} \tag{3.3}$$

  where

  $\operatorname{sgn}(x)=-1$ when x is negative, otherwise

  $\operatorname{sgn}(x)=+1$;

  $\text{hcf}_{a,b}$ is the highest common factor of a and b; and

  $\xi(x)$ is the indigestibility of x.

  Thus higher component indigestibility values produce lower harmonicity. Barlow regards the sign of the result as indicating which pitch functions as root. Positive values indicate the lower pitch, negative values the upper pitch.

- From his formulation of harmonicity, he now imposes a nominal minimum desired harmonicity and asks what *sequence of maximum absolute powers* satisfies the constraint. In other words, given an arbitrary sequence of primes, what is the maximum exponent to which each may be raised without causing harmonicity to fall below the selected value? This he obtains as follows:

$$N(p)=\left[\frac{\omega+1/h}{1+(\log(256)/\log_e(27))}\right], \text{ where } p=2, \text{ otherwise} \tag{3.4}$$

$$N(p)=\left[\frac{\omega+1/h}{\xi(p)+(\log_e(p)/\log_e(2))}\right] \tag{3.5}$$

where

p is the prime number, the maximum power of which is desired;

h is the minimum harmonicity;

ω is the pitch range in octaves;

$\xi(x)$ is the indigestibility of x; and

$[x]$ is the integer part of x.

- The *maximum powers sequence* obtained in this way drastically reduces the number of intervals to be considered, being guaranteed to span all intervals more harmonic than the stated minimum. The converse, however, is not usually true. It is possible to use the sequence to construct intervals which are less harmonic than required. Thus the values obtained are used to combinatorially generate intervals, each of which is evaluated in terms of the harmonicity function above.
- Finally, Barlow considers the linear distance between the interval to be rationalised and the candidate rationalisations. He applies a “Gaussian-like bell” over the candidates, centred on the value to be rationalised with variance set so that his nominal *tolerance* equals that value at which the damping factor is 20. The resulting weighting balances the harmonicity of the candidates against their degree of approximation, and thus establishes a “winner”.

Barlow's method is interesting in that it speaks directly to the domain of music, but perhaps more importantly in that it proposes a method of reducing the “search space” in which to seek solutions.

The problem of finding rational (possibly integer) constrained solutions to a given problem is the subject of Diophantine analysis. Diophantine approximation addresses the kinds of concerns we have here: the finding of a “best” rational approximation to a given irrational value. These are typically approached by way of Dirichlet's box principle, continued fraction expansions, Farey series or Minkowski's Theorem (Waldschmidt 2009). Within the context of electronic warfare, Vaughan Clarkson seeks an optimal method of detecting radar pulses of unknown periodicity (Clarkson 1997). He reviews various methods, ultimately settling on a refinement of Cassels' Algorithm. In our preliminary investigation, we have confined ourselves to the more naïve Euclid's Algorithm. Clarkson slightly improves Euclid's Algorithm as follows:

```
begin
    $\eta_{-1} := -1; \eta_{-2} := \alpha;$
    $p_{-1} := 1; p_{-2} := 0;$
    $q_{-1} := 0; q_{-2} := 1;$
    $n := 0;$
    while $\eta_{n-1} \neq 0$ do
        $a_n := \left\lfloor \frac{-\eta_{n-2}}{\eta_{n-1}} \right\rfloor;$
        $p_n := p_{n-2} + a_n p_{n-1};$
        $q_n := q_{n-2} + a_n q_{n-1};$
        $\eta_n := \eta_{n-2} + a_n \eta_{n-1};$
        $n := n+1;$
    od;
end.
```

The output of this algorithm yields steadily improving approximations of α in *p*/*q*, and could quite conceivably be modified to consider Barlow's harmonicity constraint. Clarkson reviews various other methods which are yet to be considered.

Before we can apply our hunch to the problem domain, we might want to consider just how much of the signal will be relayed by the auditory mechanism. As a point of departure, we will review the model employed by Richard Parncutt in his psychoacoustic approach to harmony (Parncutt 1989). Parncutt's model is largely based on the work of Ernst Terhardt, but is most notably distinguished in the manner in which it is discretised to deal directly with the pitches of equal temperament. Nonetheless, his input (integers representing *pitch categories*, rather than frequencies, with C_4=48) is processed as follows:

- Each input *pitch category* P_1 is expanded to a series of sixteen harmonic "partials" according to the formula $P_n = P_1 + [12 \log_2(n) + 0.5]$, where [x] denotes the integer part of x and n is the ordinal of the partial. Note again that these are pitch categories, not frequencies.

- Component *auditory levels* (*yl*, in dB, representing the amount by which the component exceeds the auditory threshold) are set for each partial as follows:

$$yl(n) = \frac{P_n(120 - P_n)}{60}\left(1 - \frac{(P_n - P_1)}{120}\right) \qquad (3.6)$$

The inverted parabola of the first half of this function approximates Terhardt's "spectral dominance phenomenon" (Terhardt *et al.* 1982a:683), while the second half serves to increasingly attenuate higher order partials, the slope being shallower for lower fundamentals (ranging from -8 dB/8^{ve}) than for higher fundamentals (up to -17 dB/8^{ve}). Where partials coincide, their intensities are summed:

$$YL(P)=10\log_{10}(\sum 10^{yl(P)/10}) \tag{3.7}$$

- Masking effects are approximated by first converting *pitch categories* to a formulation of *pure tone height* (derived from Parncutt's own semitone-based representation of critical bandwidth):

$$H_p(P)=\sqrt{x^2+44}+x-2 \tag{3.8}$$

where $x=P/5-10$ (P is the pitch category)

This *pure tone height* is used to calculate the distance between every two pitch categories, multiplied by a slope parameter k_M (typically 25 dB per critical band), and thus obtain a nominal attentuation (*masking level*) to be applied to the *auditory levels* of the component *pitch categories*:

$$ml(P,P')=YL(P')-k_M|H_p(P')-H_p(P)| \tag{3.9}$$

The combined effect of multiple maskers on a given *pitch category P* is obtained by summing equivalent pressure amplitudes of all contributing maskers as follows:

$$ML(P)=max\left\{20\log_{10}\left(\sum_{P'\neq P} 10^{ml(P,P')/20}\right),0\right\}, \tag{3.10}$$

where the *max* function prevents a negative result.

- The *audible level* is now obtained for each partial by subtracting the effective *masking level* from the calculated *auditory level:*

$$AL(P)=max\{YL(P)-ML(P),0\}, \tag{3.11}$$

where the *max* function again keeps the result positive.

- Finally, partials are caused to "saturate with increasing audible level" as follows:

$$A_p(P)=1-\exp(-AL(P)/15) \tag{3.12}$$

Parncutt's model continues by applying a stylised harmonic template (rooted in 12tet) to each pitch category, testing for the existence of expected overtones and thereby establishing a quantitative prediction of the *audibility* of that pitch category as the fundamental of a complex tone. However, it serves us better to first turn our attention to an extension of this model, specifically one accepting input in terms of frequency, rather than pitch categories (Parncutt & Strasburger 1994). This model does not assume octave equivalence, and is proposed as having a "considerable degree of generality." The following points outline the approach taken here:

- Pitches may be input as pure-tone components of arbitrary frequency in Hz (though a formula for obtaining equal temperament pitch categories is provided).
- *Auditory level* is specified relative to the "free-field threshold of hearing in quiet" (in dB), which is taken from (Terhardt *et al.* 1982a:682) as:

  $$L_{TH}=3.64\,f^{-0.8}-6.5\exp\{-0.6\,(f-3.3)^2\}+10^{-3}\,f^4 \tag{3.13}$$

  Thus the *auditory level* is the number of decibels by which the pure tone component of frequency *f* exceeds the value calculated by the above formula. This is essentially an inverse approach to *spectral dominance*.
- The determination of masking effects once again invokes the concept of *pure tone height* (or *critical band rate*), but this time employs the formula provided by (Moore & Glasberg 1983:752):

  $$H_p(f)=11.17\log_e\left(\frac{f+0.312}{f+14.675}\right)+43.0 \tag{3.14}$$

  Individual masking contributions and cumulative masking are obtained in precisely the same way as before, except that it is frequencies, rather than pitch categories which are plugged into the equations, and the slope parameter k_M is set between 12 and 18 dB per critical band.
- The *audible level* AL and *pure tone audibility* A_p are derived precisely as above.

The latter model also proceeds to apply a harmonic template, once again in terms of equal temperament pitches, and thus ceases to be of immediate interest here. However, both models suggest methods of accounting for perceived *auditory level*, particularly in terms of compound masking effects. The Glasberg & Moore equation has since been updated (Glasberg & Moore 1990), and the masking equation might be improved by incorporating realistic asymmetries in auditory filter response. Impressive results, for instance, are reported for the one-zero gammatone filter (OZGF) (Lyon *et al.* 2010).

3.6 Findings of a Qualitative Study

In Appendix B we present our review of "priming" research in psychology, specifically as it pertains to musical stimuli. Briefly restated, priming experiments present some kind of context, or prime, to a subject before following with a target. Given a simple perceptual judgement aimed at the target, subjects provide measures of reaction time and error rate as experimental data. In harmonic priming, both prime and target are chords. In general, priming serves up faster reaction times and lower error rates in respect of expected targets, relative to a given prime, than unexpected targets.

We extensively surveyed the multitudinous ways in which priming has been leveraged in order to reveal the nature of music processing by humans. As has repeatedly been shown, priming provides a rather robust indicator of musical expectations, and has thereby served up sufficient evidence, for our purposes, to support the special status of particular harmonic progressions relative to others. We therefore dispense with the need to provide empirical evidence of our own to support this: it seems quite clear that, within Western music, root movements by fourth up / fifth down, and especially the dominant-to-tonic perfect cadence, are highly expected progressions. It is less clear whether that special status arises purely by enculturation, by the fostering of a mental schema, or by statistical learning, having greatest expectation for whatever has been most frequently encountered before. These two interpretations have given rise to so-called “top-down” and “bottom-up” accounts of musical processing, respectively, with various models proposing what particular contribution is made by each. Our own interpretation is “bottom-up” in the sense that we accept that stimulus-driven learning reinforces and enhances many musical percepts. We contend, though, that a particular schema (loosely, tonality) is favoured over all others by virtue of non-linear auditory processing, thereby introducing a strong “top-down” contribution. This view leads us to expect a high degree of universality for musical aspects which can be understood as the artefacts of such non-linear processing.

We now seek to understand musical teleology in terms of non-linear processing, but let us first establish: what constitutes a “music-theoretic understanding of musical teleology”? To our reading, it is the Schenkerian view of a “biology of tones” (Schenker 1954), and particularly its extension into the GTTM, which continues to dominate this space, at least as regards Western tonal music. It is quite plain, though, that the seeds of Schenker's *Bassbrechung* lie in the *lebenskraft* (Bent & Pople 2001:541–2) of Hugo Riemann's fundamental progression (I – IV – V – I). And Riemann is, in turn, deeply indebted to Rameau, whose notion of a fundamental progression, according to Thomson (1993), is clearly grounded in root movement by fourth up / fifth down (I – ii_6 – V^7 – I), and who furthermore insisted that the dominant seventh, like a Galilean/Keplerian *vis motrix*, was a indispensable functional requirement, and not a mere ornamental embellishment (as is typically taught today). Furthermore, though Rameau framed harmonic principles in a bold new tertian way, his resolution of dissonance merely systematised extant contrapuntal practise which may be traced back via Gioseffo Zarlino to at least Marchetto da Padova (fl. 1305 – 1319) (Cohen 2001). What reaches us 700 years later is then seen to have ultimately evolved from an Aristotelian principle: “*imperfectum appetit suam perfectionem*”. Other traces of this discourse are to be found in the work of Fétis (*affinites*, *tendances energiques*), Halm (drama of forces) or Kurth (interplay of potential and kinetic energies), and are clearly still present in recent work of Larson, Lerdahl, Krumhansl, and Thomson, amongst others. To this we should also add the names of Meyer, Narmour and Farbood, whose contributions are all framed in comparable terms. To be clear, the common thread that binds the parade of names is simply this: all acknowledged, at some level, a formal order that governs musical teleology. Notably, these ideas are predominantly framed in harmonic terms, with conspicuously less attention being afforded to matters of rhythm. Nonetheless, it should be quite clear that the fundamental principle, being that certain sonic configurations create a strong expectation that they are to be followed by other, particular sonic configurations, is perfectly congruent with the evidence produced by harmonic priming studies.

In seeking evidence of musical teleology outside of the Western canon, we have cast our net widely over sub-Saharan Africa, focusing first on Bantu, then Nguni, and then Xhosa music (see our Appendix C). Principally, we note the wide-spread derivation of tonal materials from alternating fundamentals of a hypothetical bow, these fundamentals most commonly spaced either a semitone apart (Zulu, Swazi), a tone apart (Xhosa), or possibly even a minor third. The sense of motion is likely driven, in part, by the combination of alternating fundamentals and descending melodies. “[i]t appears to be a rule of Xhosa music that melodies should begin on an off-beat; this coincides with the strong initial accent of the melodic pattern, and is always sung with loud emphasis.” (Hansen 1981:624–625) The ensuing rhythmic transitions from out-of-phase (off-beat, at the beginning of phrases) to in-phase (on-beat, at the end of phrases) appear to contribute significantly to a sense of goal-directed motion, and Rycroft notes with interest to what extent “... rhythmic concord coincides with the other characteristic 'punctuation signs' of final or semi-final tone and interval, and long note-duration, to achieve climax...” (Rycroft 1954:24) Regarding Xhosa music, “This is what Jones has described as the 'teleological trend' in African music: songs 'lean towards the end of the lines: it is at the end where they are likely to coincide with their time background'” (Jones 1959:49, cf. also 41, 84, 86, 128; Hansen 1981:624). We also take note of Jeff Pressing's exploration of isomorphisms between West African timelines and Western diatonicism, which have lead him to suggest that “... this commonality must tell us something underlying about perception and the mind.” (Pressing 1983:44) Initially, such studies were not particularly well received: “... recent scholarship, motivated by an empiricist illusion, has confined its field of enquiry to what may be termed the mechanical aspects of rhythmic organization.” (Agawu 1987:403) The reception has thawed somewhat since then: “Pressing's rules […] hold considerable promise for codifying certain aspects of African musical behavior and thereby promoting a cross-cultural understanding.” (Agawu 2006:29) Still, there remains strong opposition to the notion of quantitative analysis of African musics: “... the quantitative orientation that facilitates interdimensional – not intersemiotic – comparison is not characteristic of African musical discourse” (2006:17), and that its products, for all their purported objectivity, “... have so far not found any corroboration in indigenous African discourse” (2006:6). Regardless, we press on, arguing here not for a narrow search for the specific patterns discussed by Pressing, but rather for a more general awareness of the potential for common structural principles to be translated between spectral and temporal domains.

The evidence would appear to support, at the very least, the notion that musical expectation is a feature not only of Western music, but perhaps of some African music too. Moreover, it would seem that the interplay of super-particular ratios is equally evident, whether in apparently tonic-dominant alternation (ratio 3:2, or perhaps 4:3), alternation between fundamentals separated by a tone (ratio 9:8) or in the superposition of duple- and triple-time rhythmic divisions. We want to now suggest that such super-particular ratios are significant chiefly due to the phenomenon of entrainment.

3.7 Entrainment as Driver of Directed Motion

The term "entrainment" assumes more or less abstract meanings in various contexts, as is evidenced by its application to domains as diverse as hydrodynamics (Scase *et al.* 2006) and systems of cosmological and social order (Collier & Burch 1998). The specific meaning intended in this research is in line with that of Pikovsky (2003:11–14), being a specific characteristic of non-linear oscillators elsewhere termed mode-, phase-, or injection-locking.[54] Specifically, two oscillators having sufficiently close natural frequencies are found to settle on a common frequency of oscillation by virtue of their interaction.

Without necessarily undermining this more narrow reading, Collier & Burch posit "rhythmic entrainment", together with "symmetry breaking", as "... two processes [that] are responsible for much (if not all) of the complexity and organization in the Universe" (1998). Invoking Schrödinger's (1948) suggestion of "negative entropy", particularly as later formulated as the "Negentropy Principle of Information" (NPI) (Brillouin 1953), they argue that any formation of order inevitably dissipates exergy (energy available to do work). What is more, they distinguish between forced and spontaneous entrainment, emphasising the gross inefficiency of the former. In particular, "[f]orced entrainment always transfers preexisting information", while spontaneous entrainment is capable of creating new information. Abstract as their Shannon-esque approach might seem,[55] they nonetheless relate the same principles to various symmetries observed in the Solar System, and to energy levels in particular chemical compounds. "Spontaneous entrainment creates *new* symmetries via the dissipation of energy and/or information. Systems tend towards minimal energy and tend to organize themselves so as to minimize dissipation (and consequently loss of available energy within the system – self-organization tends to increase efficiency). This process increases higher level order, or symmetry, and is mutual among the parts of the system, with excess energy being dissipated externally, unlike many cases of forced resonance." Ultimately, though, they set their sights on human sensory perception, arguing provocatively that self-organisation (spontaneous

---

[54] Huygens described the phenomenon as "*une espece de sympathie*". Hoppensteadt & Izhikevich take pains to distinguish between frequency-locking, entrainment or phase-trapping, on the one hand, and phase-locking, on the other, employing the term "synchronization" only to describe simultaneous frequency- and phase-locking (1997:248–253). Other authors seem to be less concerned about such distinctions.

[55] Ongoing debate surrounds the coexistence of two distinct views of entropy, these being the classical thermodynamic definition (in terms of energy transfer per unit temperature) and the statistical mechanics definition (in terms of the number of potential microstates), the latter having found its way into information theory and beyond variously with or without the Boltzmann constant. The latter approach paves the way for such abstractions as are proposed above.

entrainment) might play an important role in forming gestalts (Collier & Burch 1998).

To clarify the link between the above account and our notion of directed motion, we return to Schrödinger (1948). What separates life from other forms of existence, he argues, is metabolism. And metabolism, he continues, is "... the marvellous faculty of a living organism, by which it delays the decay into thermodynamic equilibrium (death) … attracting, as it were, a stream of negative entropy upon itself, to compensate the entropy increase it produces by living and thus to maintain itself on a stationary and fairly low entropy level." Without delving into the subtleties of the argument, we can immediately see how Schrödinger distinguishes life by virtue of its active attempts to retard the inevitable progression toward increasing entropy, which it accomplishes by "... continually sucking orderliness from its environment." The most efficient means of accomplishing this, by Collier's & Burch's (1998) account, is by spontaneous entrainment, and is likely instrumental in much human sensory perception too. Thus entrainment would, inasmuch as it represents the perpetual re-creation of ever-higher levels of order, be an intimate expression of life-processes.

At a more concrete (and somewhat less philosophical) level, the mechanism of entrainment is arguably an expression of the "Principle of Stationary Action". A non-linear system driven by some external stimulus tends towards a state which results in minimal dissipation, thereby maximising the efficiency of energy transfer between itself and its stimulus (see, for example, (Susuki *et al.* 2007)). Could it be, then, that our musical expectations might be determined, to some extent, by a physiological imperative to maintain maximally efficient energy transfer as we listen to music? Might music's affective affordance not be determined, in part, by the psychic tension which might arise when such efficiency is denied?

We therefore speculate that, given a predisposition towards the extraction of order from our environment (to offset the entropy we naturally produce), and given that spontaneous entrainment constitutes a particularly efficient means of doing so, so too are our musical expectations profoundly influenced by the particular progression of patterns implied by such an account. Given any particular starting configuration of musical sounds, entrainment would suggest the next closest configuration, in the super-particular sense explored earlier, as the most desirable, since such a transition would constitute the most efficient option by virtue of minimising the dissipation of exergy.

In various cultural settings, this mechanism could play out in different ways, albeit through the relationships between pitch configurations, or the temporal patterning of events, or any hybrid of both. What remains common to all accounts is the particular harmonic ordering implied. However, this only speaks to our expectations. What imbues music with its particular emotive effect is the extent to which it validates such expectations at the highest level, whilst simultaneously challenging those same expectations at other levels. In effect, the musical man is being goaded into hunting for his food, and thereby to experience the thrill of that pursuit, rather than merely receiving rations. Thus, to whatever extent we are able to find evidence of directed motion in any particular music, we should also expect substantial attempts to confound those expectations.

3.8 Towards an Improved Distance Measure

Our earlier attempt to construct a distance measure between harmonic series was crude (see section 3.4), and though it appeared to confirm our intuitions, it proved to be analytically intractable. The link between entrainment and efficiency of energy supply might then be a more promising prospect.

As shown in (Susuki *et al.* 2007), maximally efficient energy transfer and minimal dissipation are markers of a non-linear system at equilibrium. Their analysis of a Van Der Pol oscillator, driven by a sinusoidal stimulus, places analytically derived amplitude- and phase-response curves alongside an analytical expression of supplied energy. The latter is derived by constructing an "energy balance relation" over the system, and separating the dissipation and driving terms by inspection. Comparison readily reveals that energy supply is most efficient within the range of entrainment.

Exploiting this principle in respect of arbitrary stimuli, though, requires that we assemble a bank of such oscillators, with each tuned to resonate maximally to a different range of frequencies. Such a configuration would closely resemble a conventional bank of overlapping filters, but would be distinguished by the addition of non-linear behaviours associated with non-linear oscillators, such as entrainment. In other words, given a bank of such non-linear oscillators which is already resonating to some particular musical stimulus, other stimuli could result in a more or less efficient energy transfer, depending on the proximity of the harmonic components in the incoming stimuli to those already resonating. A stimulus which presents a smoother mapping, we presume, will make for the most energy-efficient transition. Based on the results of our naïve simulation earlier, we would expect such an approach to further support the special status of superparticular ratios, but moreover to provide, by way of entrainment, a plausible "tolerance" mechanism (see section 3.5) whereby to abstract such ratios from the irrational relationships present between the components of real audio stimuli.

We therefore propose to monitor the progress of such an oscillator bank towards equilibrium, specifically in terms of energy supply efficiency, and thereby to characterise some musical continuations as being more efficient than others. From a conventional music-theoretic perspective, this is not a particularly surprising position to take, since this idea is already well embedded in the principles of parsimonious voice-leading. In that account, however, we only consider fundamentals, and are constrained by the lens of Western pitch classes. What we are proposing here is a mapping which speaks directly to the audio signal, without the intercession of Western music-theoretic constructs.

3.9 Concluding Remarks

After contextualising the contentious nature of our quest, this chapter has set out to develop our initial intuition about the nature of musical teleology into a general, cross-cultural construct. We propose that the super-particular ratios which permeate musical organisation as an ideal, even if not in practical implementation, point to the agency of entrainment as a key driver of musical expectation. Though entrainment has been considered with regards to ethnomusicological studies before (Clayton *et al.* 2005), it does not appear to have been invoked, as yet, to account for musical teleology in general. Interrogating this particular notion, then, would seem to be a key contribution to be made by this study.

Chapter 4

Operationalising the Construct

We will here develop the conceptual understanding formulated in the previous chapter in order to operationalise that construct. Specifically, we wish to exploit the phenomenon of entrainment in non-linear oscillators, and especially the implications thereof for the efficiency of energy transfer, as an indicator of musical teleology. We will briefly review two components which will contribute to this endeavour, being artificial neural networks (ANNs) and non-linear oscillators, before introducing their synthesis as “neural oscillators” in a complex-valued neural network (CVNN) with non-linear dynamics. We will then turn our attention to the question of “energy” in such a network.

4.1 Artificial Neural Networks

Indubitably, fascination with the mysteries of perception and cognition both fuelled and inspired the early development of Artificial Neural Networks (ANNs). McCulloch & Pitts' (1943) Threshold Logic Unit (TLU), featuring summed inputs to a simple binary threshold function, heralded two and a half decades of enthusiastic development and produced, as single examples, Hebbian Learning (Hebb 2002), Rosenblatt's (1962) Perceptron (featuring iterative weight adjustment with faster convergence than Hebb's), as well as Widrow & Hoff's (1960) Adeline (and later Madeline, featuring an important precursor to the backpropagation rule which would emerge in the 1980s). Some impetus was lost when Minsky & Papert (1988) pointed out some of the limitations of single-layer ANNs, including the inability to learn simple mappings such as the XOR function, this largely being due to the absence of sufficiently general training methods in multi-layer ANNs. Regardless, important developments emerged during the 1970s in the work of Kohonen, Anderson, Grossberg and Carpenter, particularly in the field of Self-Organising Maps (SOMs). A wider resurgence of interest followed in the 1980s, with the emergence of a robust method of feeding back errors at the output to earlier layers (the so-called backpropagation rule), and also due to advocacy by prominent physicist John Hopfield.[56]

In its modern, canonical form, an ANN consists of simple computational units which sum their weighted inputs, feeding the result to an activation function, which ultimately produces the output (cf. Figure 4.1).

---

[56] For a more complete, though brief exposition of the same, see (Fausett 1994:22–26)

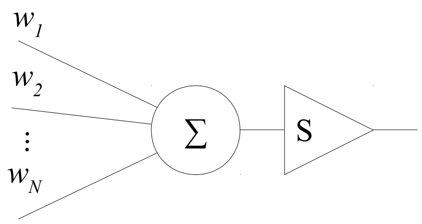


Figure 4.1: Schematic of an artificial neuron

This could be viewed as a two-step procedure:

$$a = \sum_i w_i x_i, \tag{4.1}$$

$$y = \frac{1}{1 + \mathrm{e}^{-a}} \tag{4.2}$$

where x_i are the various inputs, each weighted by a corresponding w_i, and the activation function S is the logistic sigmoid, which serves to asymptotically “squash” the output value y towards either 0 or 1.

Variations on this simple theme abound, but the core principles encapsulated in the above remain evident. An ANN, then, combines an arbitrary number of these computational units (or neurons), connected in any of a number of possible configurations, such that the outputs of certain neurons feed the inputs of others. Any output may simultaneously connect to inputs of multiple neurons (“fan out”), and a neuron's output may feed its own input (feedback). While in some cases it might be sufficient to employ a single, fully-connected layer of neurons (recurrent network), other applications are best served by multiple layers in a feed-forward configuration. In such a case, outputs of one layer feed the inputs of the next. Between the input and output layers, it might be necessary to introduce hidden layers.[57] Typically, the implementation of such an ANN is performed in terms of matrices of state variables and connection weights. In the case of feed-forward networks, a differentiable activation function allows one to perform backpropagation learning. Alternatively, methods such as genetic mutation and simulated annealing might be employed.

With regard to the kind of ANN described above, it is important to note that outputs are essentially binary. Notwithstanding the fact that common activation functions are asymptotic to their extrema, the outputs are nevertheless understood to represent discrete values, thereby representing the firing (or not) of a particular neuron. By incorporating bias and scaling parameters to the typical sigmoid function (Figure 4.2), the threshold can be centred arbitrarily and made as steep as might be required, becoming a discrete step function in the limit.

---

[57] Theoretically, no more than two hidden layers are ever required to solve a given problem. However, the desire to model physiological processes may motivate the inclusion of more layers.

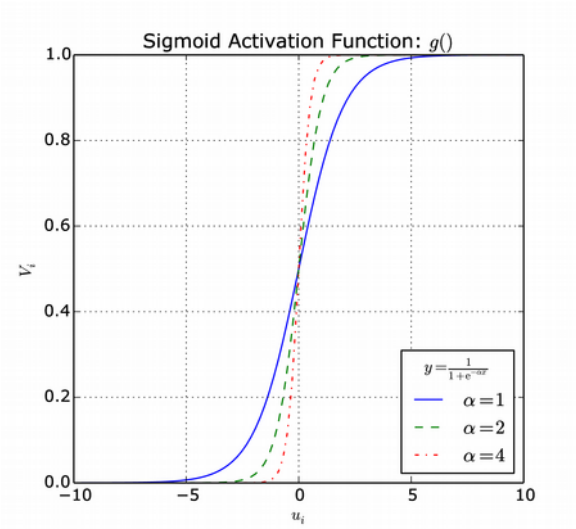


Figure 4.2: Sigmoid activation functions with three different values of α controlling steepness

$$y=\frac{1}{1+\mathrm{e}^{-\sigma(a-\theta)}} \tag{4.3}$$

However, smoothness is retained primarily to serve the differentiability requirements of backpropagation learning.

The binary nature of the activation function's output makes it difficult to conceive of such an output (from a single neuron) as oscillatory except, perhaps, in the sense of a periodic pulse train. Real, physiological neurons, however, do display a range of sub-threshold oscillatory behaviours, but these are lost in the simple model given above. Consider the simple unforced damped harmonic oscillator, with mass, friction and restoring force:

$$m\ddot{x}+\mu\dot{x}+kx=0. \tag{4.4}$$

This may be linearised by substituting $x_1=x$, $x_2=\dot{x}$, with the result that:

$$\begin{aligned}\dot{x}_1&=x_2\\ \dot{x}_2&=-\frac{\mu}{m}x_2-\frac{k}{m}x_1\end{aligned}. \tag{4.5}$$

Thus a network of two computational units, appropriately connected, and without activation threshold, would appear to be capable of typical oscillatory behaviour, given appropriate values for each of the parameters and initial conditions.[58] One might further notice that the connection from x_2 to x_1 is positive, whilst that from x_1 to x_2 is negative, and that x_2 furthermore receives negative feedback from its own output. It is precisely such pairing of excitatory and inhibitory neurons, as is supported by physiological evidence, which gives rise to

---

[58] In fact, Yang argues that at least three neurons are required (Yang 1995).

oscillation in ANNs, as will presently become evident when we review the Wilson-Cowan equations (see section 4.3).

4.2 Non-Linear Oscillation

The perception of music rests heavily on the presence of periodicity in the signal perceived, as has already been discussed earlier. Thus it is apt that oscillating ANNs should endear themselves to musical tasks. However, simple linear (or harmonic) oscillators have had little efficacy in elucidating some of the more enigmatic aspects of the musical experience. Very briefly, linear oscillators are subject to a restoring force which is directly (linearly) proportional to their state of displacement from rest, as may be expressed succinctly as $F=-kx$ (Hooke's Law), yielding the equation of motion $\ddot{x}+kx=0$, or more conveniently, $\ddot{x}+\omega^2 x=0$.[59] Solutions then take the form $x(t)=\alpha\cos(\omega t+\theta)$, where α and θ are determined by initial conditions. It is immediately clear that the radian frequency of oscillation is in this case $\sqrt{k}$, and that this oscillator is expected to resonate most strongly at only that one frequency. In the face of small perturbations, many real-world oscillatory systems do in fact behave adequately linearly, to all intents and purposes. However, at greater displacements, most natural systems deviate from linear predictions by evidencing more complex relationships between their displacement and their restoring forces.[60] In the spirit of a Taylor expansion of some such more complex relationship, let us consider the addition of just a single quadratic term into the relationship between the displacement and the restoring force, as in $F=-\omega^2 x+\alpha x^2$, which then gives the equation of motion as $\ddot{x}+(\omega^2+\alpha x)x=0$. Such systems do not yield solutions by conventional analytical methods, but a numerical simulation quickly shows what is happening.

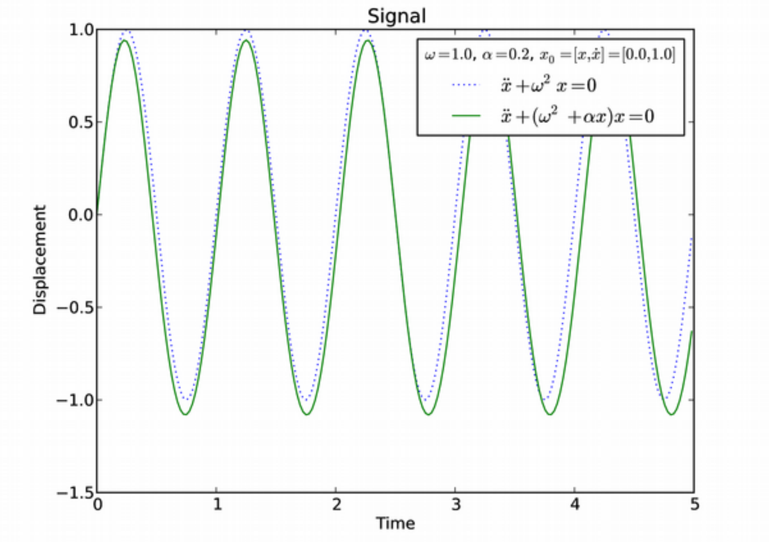


Figure 4.3: Linear vs. non-linear oscillation in the time domain

Apparently, these signals differ in frequency, but the non-linear signal also seems to be offset vertically, as if

---

[59] For the sake of both clarity and brevity, we have here dispensed with mass and friction.
[60] The humble pendulum is a common textbook example.

by bias. A view from the spectral domain sheds even more light.

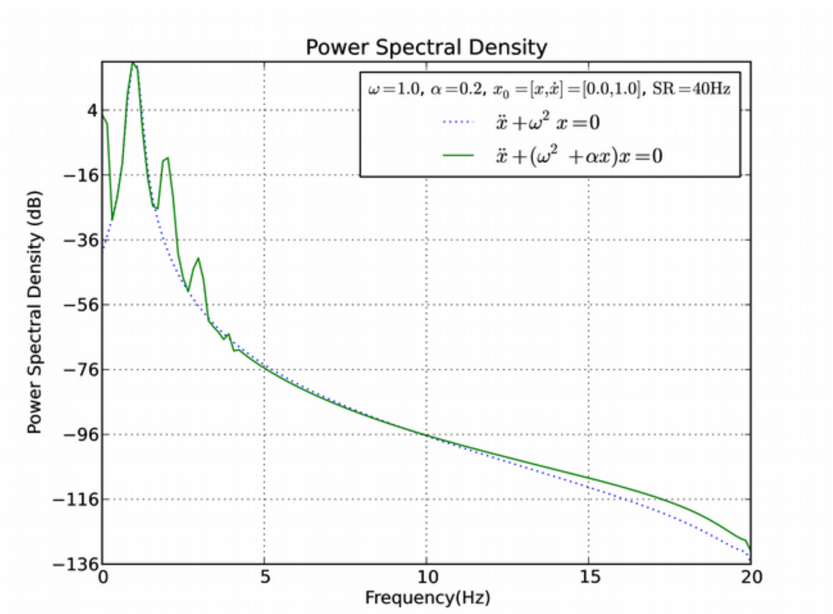


Figure 4.4: Linear vs. non-linear oscillation in the spectral domain

Here is becomes clear that, while the linear oscillator resonates at a single frequency, the non-linear oscillator resonates at multiple frequencies. Specifically, these are integer multiples of the apparent fundamental frequency (f_0, $2f_0$, $3f_0$ and $4f_0$ are all plainly visible). A phase plot similarly demonstrates the qualitative differences between the linear and non-linear cases.

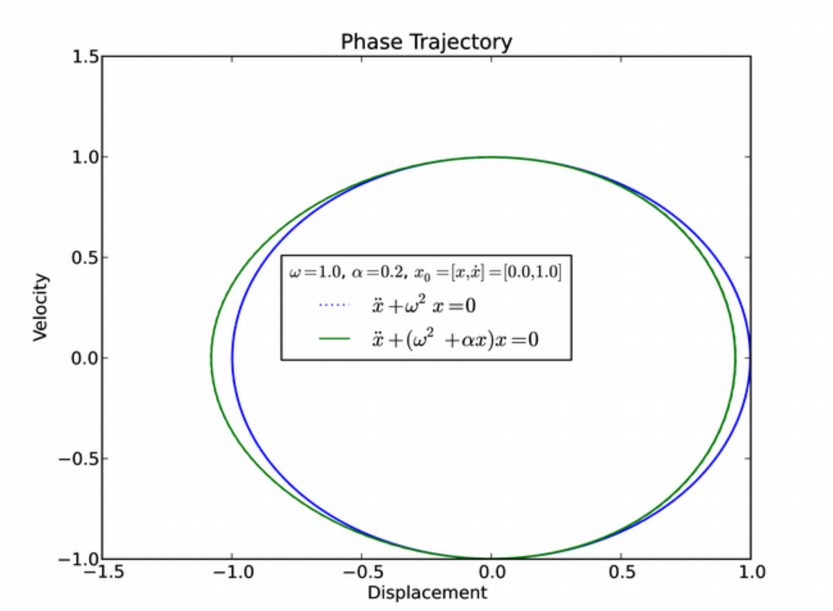


Figure 4.5: Linear vs. non-linear oscillation in phase space

As was already evident in the signal plot (Figure 4.6), we see that positive displacement is attenuated, while negative displacement is amplified. We can get a sense of why this is so by plotting the restoring force.

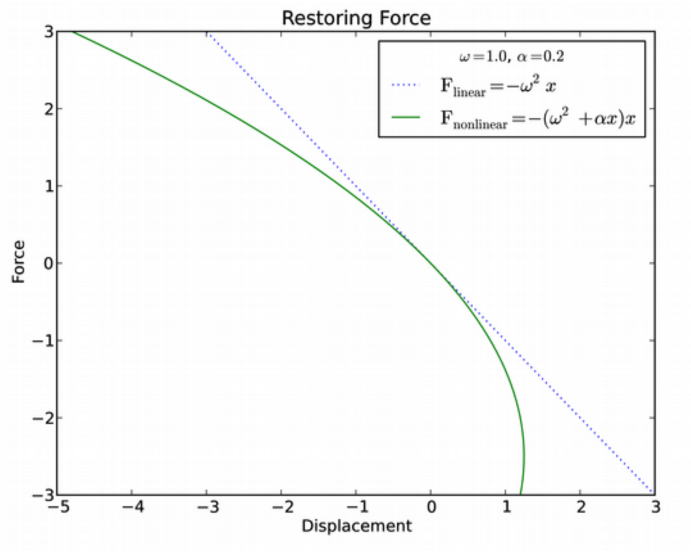


Figure 4.6: Restoring force in linear vs. non-linear oscillation

Here the simple Hooke's law relationship of the linear oscillator stands in contrast to the apparently parabolic relationship between force and displacement in the non-linear oscillator. In the latter case, greater positive force does not linearly increase the negative displacement, while greatest positive displacement occurs in response to a negative force of precisely -5/3, whereafter the displacement lessens. Note, though, that both behaviours are substantially similar near the origin, which is why linear methods are frequently found adequate.

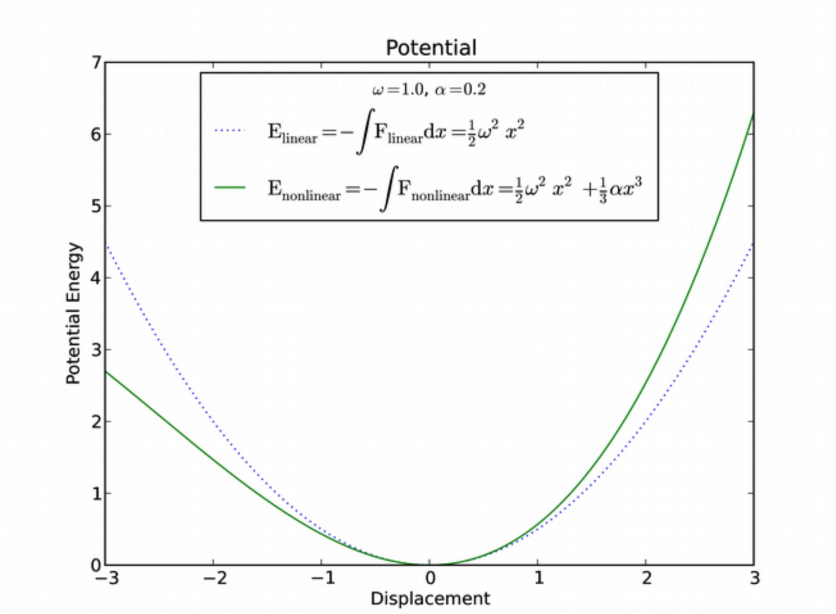


Figure 4.7: Potential function of linear vs non-linear oscillators

The above, necessarily brief, exposition of non-linear oscillation will now be supplemented by an eclectic sampling of other important characteristics which set such systems apart from their more widely understood linear counterparts.

The eigenfrequency of a linear oscillator is independent of its state, or of any driving amplitude. Not so for non-linear oscillators, which may be seen to have their resonances shift in response to a change in driving amplitude. This is at once of interest since it is known that musical pitch perception does, in fact, depend on the amplitude of the stimulus received (Rossing 1990). Furthermore, as has been demonstrated above, linear oscillators merely filter their stimuli, while non-linear oscillators effectively add to the signal. Given the persistent, though largely unsatisfactory attempts (to date) to relate pitch perception to some cognitive awareness of harmonic relations, whereby missing fundamentals are actively recomputed, it would seem far more parsimonious to see such reconstruction as being the inevitable, passive result of super-harmonic resonance in non-linear oscillators. Non-linear oscillators are furthermore capable of "entrainment", whereby they may phase-lock to a sufficiently close, yet quantitatively different stimulus frequency. This clearly speaks to the notion of categorical perception (or what Barlow (1987) has termed "tolerance"), though once again more parsimoniously, being by passive rather than active means. Again, this is not the case for linear oscillators.

The picture that should be emerging from all of the above is that a suitably crafted oscillating neural network, specifically one whose oscillation is suitably non-linear, should be capable of detecting harmonically related periodicities in complex signals. The same approach should scale equally well to both temporal and spectral domains, and should moreover be endowed with a suitable degree of tolerance (by virtue of their phase-locking behaviour) to be able to adapt to typical fluctuations in pitch or tempo. Over time, such a network (by virtue of super-harmonic resonance and Hebbian learning) should come to favour strong connections between harmonically related nodes. Within the Western context, this could proffer a alternative, plausible account of tonal and rhythmic hierarchy, as has been extensively argued by Large. This time, however, the account is rooted far more deeply in natural, rather than cognitive processes. In the context within which we are engaging this work, we see the potential of a slightly different reading of the same principles. In particular, we suspect that the role of phase-locking is even more substantial in the human musical experience than has hitherto been acknowledged, and perhaps, in the music of sub-Saharan Africa, lies at the core of the experience of directed motion.

4.3 Neural Oscillation

While previous models of neural activity (FitzHugh 1961; Hodgkin & Huxley 1952; McCulloch & Pitts 1943; Nagumo *et al.* 1962; Rosenblatt 1962) had concerned themselves with the activation of *individual* neurons, Wilson & Cowan's (1972) core innovation was to consider spiking *probability* in neural *populations*. Specifically, they concerned themselves with populations of excitatory/inhibitory pairs (1972:7, Eq. 3 & 4):

$$\begin{aligned} E(t+\tau) &= [1-\int_{t-r}^{t} E(t')dt']\cdot S_e\left\{\int_{-\infty}^{t} \alpha(t-t')[c_1E(t')-c_2I(t')+P(t')]dt'\right\} \\ I(t+\tau') &= [1-\int_{t-r'}^{t} I(t')dt']\cdot S_i\left\{\int_{-\infty}^{t} \alpha(t-t')[c_3E(t')-c_4I(t')+Q(t')]dt'\right\} \end{aligned}, \tag{4.6}$$

with

$$S(x)=\frac{1}{1+\mathrm{e}^{-a(x-\theta)}}-\frac{1}{1+\mathrm{e}^{a\theta}}, \tag{4.7}$$

where E and I represent the spiking probability in excitatory and inhibitory populations, respectively; P and Q represent input stimuli applied to the corresponding populations; c_1, c_2, c_3, c_4 represent connectivity weights; τ, τ', r, r' and $\alpha(x)$ together encapsulate the notions of membrane time constants and of synaptic refractory periods; and the sigmoid activation functions S_e and S_i may take different values of a and θ (Wilson & Cowan 1972:10). In their paper, Wilson & Cowan go on to introduce various assumptions and simplifications,[61] including time coarse graining[62] (thereby later eliminating the integrals) and dispensing with input to the inhibitory population (eliminating the Q term). Later authors have taken this even further, producing what is arguably the popular formulation of the Wilson-Cowan equations, namely (Hoppensteadt 1997:23–24,45,74):

$$\begin{aligned} \dot{x} &= -x+S(\rho_x+ax-by) \\ \dot{y} &= -y+S(\rho_y+cx-dy) \end{aligned} \tag{4.8}$$

Here, x and y have replaced E and I, with $\dot{x}$ and $\dot{y}$ being derivatives with respect to time, connectivity parameters expressed as a, b, c and d, and ρ_x and ρ_y are bifurcation parameters, selected so as to place the system in the neighbourhood of an Andronov-Hopf bifurcation.[63] The activation function S may now take various forms, though most commonly either of the binary or bipolar forms of the sigmoid function are employed, resulting in output ranges of [0,1] or [-1,1]. Specifically:

$$S(x)=\frac{1}{1+\mathrm{e}^{-\sigma x}}, \tag{4.9}$$

or

---

[61] Some of which have lately been described as “Draconian in the extreme” (Whittle 2010:135).

[62] By the “averaging method”, which reduces fluctuations at the small time scale to constants at the large time scale.

[63] A bifurcation point represents a critical parameter threshold separating behavioural regimes in a dynamic system. An Andronov-Hopf bifurcation specifically marks a transition to or from oscillatory behaviour (Hoppensteadt 1997:42,45–48).

$$S(x)=\frac{1-e^{-\sigma x}}{1+e^{-\sigma x}}, \tag{4.10}$$

respectively. The latter formulation of the Wilson-Cowan equations is intuitively more accessible with reference to Figure 4.8:

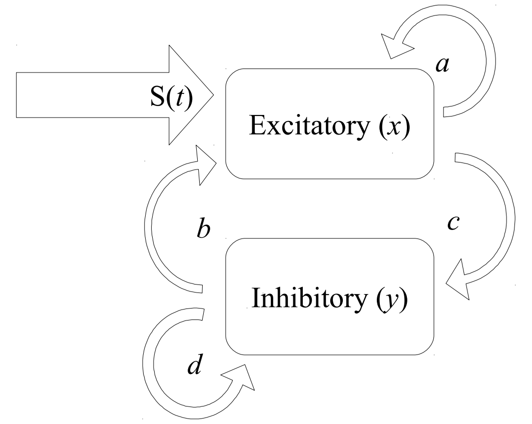


Figure 4.8: Schematic representation of a Wilson-Cowan oscillator

Here it is plain that the state of x at time t is a function of both its own positive internal feedback (weighted by a), as well as the inhibitory (negative) input received (weighted by b). Similarly so for the state of y. The selection of appropriate parameter values may bring about various oscillatory regimes (linear, critical and limit cycle). In the typical case, an additional time-varying stimulus $s(t)$ is introduced, representing external input from the environment.

4.4 Complex-Valued Neural Networks

Systems such as the above exhibit complex behaviours, such as stable and unstable equilibria, and are termed non-linear oscillators. Normal forms are derived from these in order to reduce analytical complexity by expressing such a system in simpler terms near its elementary solutions (Wiggins 2003:270). However, the application of normal form theory to the class of systems represented by the Wilson-Cowan equations[64] has typically disregarded higher-order terms (*h.o.t.*). In particular, given the following "general system of coupled neural oscillators":

$$\begin{aligned}\dot{u}_i&=f_i(u_i, v_i, \lambda)+\epsilon p_i(u_1, v_1, \ldots, u_n, v_n, \epsilon)\\ \dot{v}_i&=g_i(u_i, v_i, \lambda)+\epsilon q_i(u_1, v_1, \ldots, u_n, v_n, \epsilon)\end{aligned}, \tag{4.11}$$

with $\{u_i, v_i\}\in\mathbb{R}$ representing the state of the ith oscillator, λ as the set of parameters to functions f_i and

---

[64] As is demonstrated in (Wiggins 2003:278–283)

g_i, and $\epsilon > 0$ a connectivity parameter, normal form theory produces:

$$\dot{z}_i = z_i(a_i + b_i|z_i|^2) + x_i(t) + h.o.t., \qquad i \in \{1, \ldots, n \in \mathbb{Z}^+\} \tag{4.12}$$

where

$$x_i(t) = \sum_{j \neq i}^{n} c_{ij} z_j, \qquad \{a_i, b_i, c_{ij}, z_i\} \in \mathbb{C}. \tag{4.13}$$

The problem with this formulation, variously known as the Hopf normal form or Stuart-Landau equation, is not evident in applications which restrict themselves to networks having homogeneous (equal, or at least ϵ-close) eigenfrequencies. However, the desire to model banks of heterogeneous frequency oscillator networks, furthermore stimulated by complex acoustic stimuli (as is our own intention), introduces significant complexity into the standard procedure for obtaining normal forms. In particular, important interactions between the variously tuned oscillators are reflected in the very higher order terms ("*h.o.t.*") that are typically discarded. Large *et al.* (2010) thus propose a "fully expanded canonical model of a neural oscillator with an input":

$$\dot{z} = z\left(a + b|z|^2 + d\frac{\epsilon|z|^4}{1-\epsilon|z|^2}\right) + \frac{x}{1-\sqrt{\epsilon}x} \cdot \frac{1}{1-\sqrt{\epsilon}\bar{z}}, \qquad \{z, a, b, d\} \in \mathbb{C}, \quad |x| < \sqrt{1/\epsilon}, |z| < \sqrt{1/\epsilon}. \tag{4.14}$$

Most importantly, their model incorporates the resonant monomials which represent "harmonics, subharmonics, and higher order combinations of the input frequencies".[65] They conclude by comparing time frequency transformations of a simple stimulus, as produced by two gradient frequency neural networks (GFNNs), one based on Wilson-Cowan, the other on their expanded canonical model, and therein show high correlation between the results of the two methods. The true potential of the canonical model, however, is revealed in a more recent "pulse detection" task (Velasco & Large 2011), wherein rhythmic patterns were constructed so as to contain little or no physical energy at the humanly perceived pulse frequency. A two-layer GFNN was here shown to resonate strongly at that pulse frequency, despite there being no such frequency present in the stimulus, a result which would not have been obtained by traditional Fourier transform-based methods. This is offered as evidence that the musical experience of meter or pulse is largely due to the effects of non-linear resonance in the brain. Large has pursued a far broader program of research investigating the role of non-linear oscillation in musical perception generally, and has argued the merits of his approach in regards to matters both rhythmic and harmonic (e.g. Large 2010; Large & Palmer 2002; Large & Snyder 2009; Large & Tretakis 2005, etc.). Briefly citing a number of the auditory conundrums that have already been discussed in this work, all leading away from an understanding of music in terms of small

---

[65] The restrictions placed on $|x|$ and $|z|$ are necessitated by their derivation, which relies on the convergence of the three geometric series produced by the expansion of higher order terms.

integer ratios (SIR), he presents important evidence of non-linear response throughout the human auditory system, and reiterates the ubiquitous and persistent tendency towards SIR intervals in much of the world's music (Large 2011:117–118). Non-linear oscillators, he argues, effectively address much of this apparently enigmatic behaviour, being able to account for, amongst others, perceptual categorisation, tonal attraction and stability, and musical learning. To the latter point, Large also invokes the notion of Hebbian learning: “When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.” (Hebb 2002:62). An appropriate reformulation of Hebb's rule, in this case, is:

$$\dot{c}_{ij} = -\delta_{ij} c_{ij} + k_{ij} \frac{z_i}{1-\sqrt{\epsilon}\, z_i} \cdot \frac{\bar{z}_j}{1-\sqrt{\epsilon}\, \bar{z}_j}, \qquad \{c_{ij}, z_i, z_j\} \in \mathbb{C}, \ \{\delta_{ij}, k_{ij}\} \in \mathbb{R}, \tag{4.15}$$

where c_{ij} holds both phase and magnitude of the connection between any pair of non-linear oscillators, these in turn represented by the state variables z_i and z_j, while δ_{ij} and k_{ij} represent the learning rate in each case (cf. Hoppensteadt 1997:366; Large 2011:119–120). Thus, over time, “[neurons] that fire together wire together.” (Shatz 1992:21).

For purpose of analysis, a useful formulation is obtained by expressing the canonical model in polar form. By the application of Euler's formula, setting $x = F\,\mathrm{e}^{i\theta}$, $z = r\,\mathrm{e}^{i\varphi} \Rightarrow \bar{z} = r\,\mathrm{e}^{-i\varphi}$, $\dot{z} = \mathrm{e}^{i\varphi}(\dot{r} + ir\,\dot{\varphi})$, with $a = \alpha + i\,\omega$, $b = \beta_1 + i\,\delta_1$, $d = \beta_2 + i\,\delta_2$, and $\{\alpha, \omega, \beta_1, \beta_2, \delta_1, \delta_2, F, R, \theta, \varphi\} \in \mathbb{R}$, amplitude and phase response can be examined separately:

$$\dot{r} = r\left(\alpha + \beta_1 r^2 + \frac{\beta_2 \epsilon r^4}{1-\epsilon r^2}\right) + \frac{F(\epsilon F r + \cos(\varphi-\theta)) - \sqrt{\epsilon}(F\cos\varphi + r\cos\theta)}{(1+\epsilon F^2 - 2F\sqrt{\epsilon}\cos\theta)(1+\epsilon r^2 - 2r\sqrt{\epsilon}\cos\varphi)} \tag{4.16}$$

$$\dot{\varphi} = \omega + \delta_1 r^2 + \frac{\beta_2 \epsilon r^4}{1-\epsilon r^2} + \frac{F(\sin(\theta-\varphi) + \sqrt{\epsilon}(F\sin\varphi - r\sin\theta))}{(1+\epsilon F^2 - 2F\sqrt{\epsilon}\cos\theta)(1+\epsilon r^2 - 2r\sqrt{\epsilon}\cos\varphi)\,r} \tag{4.17}$$

From this vantage point it becomes clear how, for instance, instantaneous frequency no longer depends solely on the oscillator's natural frequency (ω), but on on its amplitude (r) as well, assuming non-zero coefficients in the implicated terms (Large 2008:202). Moreover, in the absence of any stimulus ($F=0$), the system is nonetheless able to oscillate spontaneously when afforded a negative damping parameter ($\alpha>0$), such oscillation stabilising at $r=\sqrt{\alpha/\beta}$ (Large 2008:203). Spontaneous oscillation has been posited as a plausible physiological mechanism whereby to account for memory, since oscillation continues after the removal of the stimulus. Taking ϵ to be small in the limit, one might get a sense of how the phase-locking phenomenon mentioned earlier comes into being:

$$\dot{\varphi} = \omega + \delta_1 r^2 + F\sin(\theta - \varphi) \tag{4.18}$$

Assume furthermore that the detuning parameter is also unused ($\delta_1 = 0$). It then becomes plain that, in the case of stimulus frequency which matches the natural frequency of the system ($\theta = \omega$), φ will assume the same frequency, the trigonometric term will become 0, and thus $\dot{\varphi} = \omega$, as is intuitively expected. On the other hand, given $\theta \neq \omega$, then the natural frequency will at each moment be modulated by the difference between the frequency of the driving stimulus and the current frequency of the system, ultimately driving φ progressively towards equilibrium, which would have been reached once $\ddot{\varphi} = 0$, in this case, $\varphi = \theta - \arcsin(-\omega / F)$. Of course, once all of the other parameters re-enter, the behaviour becomes somewhat more complex, though such complexity is constrained somewhat by the desire to keep ϵ small, as is typically desired. The periodic nature of the trigonometric term also implies limits to the effect.

4.5 Network Energy

The fundamental notion which we pursue here is that, given an arbitrary set of constraints, a network will converge towards a stable state of lowest "energy", arguably a reflection of the second law of thermodynamics. The update rule which governs the dynamics of a Hopfield network might be expressed as follows:

$$\tau \dot{v}_i = -v_i + \mathrm{S}(u), \tag{4.19}$$

$$u = b + \sum_j^n w_j v_j, \tag{4.20}$$

with S(•) typically sigmoid. Hopfield employed the term "energy" with reference to a circuit-based representation of these "neurons with graded response" (1984), itself a development of his earlier "two-state neurons" (1982).

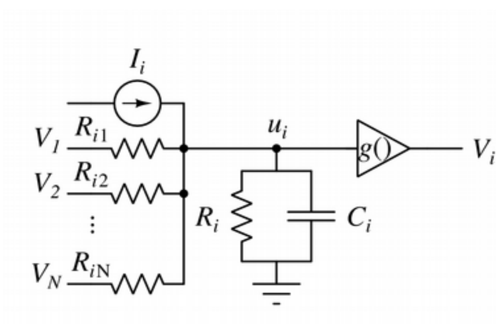


Figure 4.9: Hopfield neuron viewed as a circuit

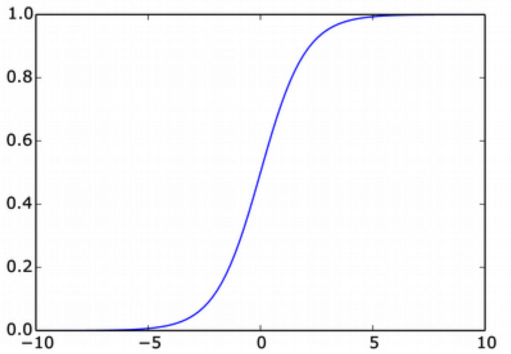


Figure 4.10: Sigmoid activation function g()

With the non-linear activation function $g(\bullet)$ represented by a high impedance operational amplifier (thus drawing effectively no current), analogies are then to be drawn between the weighted summing of inputs and the voltage adder on the left, with bias entering through I_i and the RC circuit in the centre represents the neural membrane's tendency to "leak". Ohm's law ($I = V / R$) gives one the individual currents arriving from the left, as well as that leaving through R_i, while the current though the capacitor may be readily obtained as $I = C\, du_i / dt$. Then, by Kirchoff's first law (KCL), these in- and out-flows sum to zero, and we so obtain:

$$C_i \frac{du_i}{dt} = \sum_j \frac{1}{R_{ij}} V_j - \frac{u_i}{R_i} + I_i \tag{4.21}$$

The resultant voltage then passes through the non-linear (typically sigmoid) activation function:

$$V_i = g(u_i) \tag{4.22}$$

Hopfield proposed, without explicit justification, the following "energy" function over the entire recurrent network:

$$E = -\frac{1}{2} \sum_i \sum_j T_{ij} V_i V_j + \sum_i \frac{1}{R_i} \int_0^{V_i} g_i^{-1}(V)\, dV + \sum_i I_i V_i, \tag{4.23}$$

where $T_{ij} = R_{ij}^{-1}$, which he then proved to be a Lyapunov function for that network: "...the time evolution of the system is a motion in state space that seeks out minima in E and comes to a stop at such points." (1984:3090) In fact, we might prefer to regard Hopfield's "energy" function as a "power" function (or energy rate function) instead, since it expresses the integral of flow (the various component currents) with respect to effort (the voltage at the output). To be precise, this quantity is properly termed the "co-content".[66] Nonetheless, Hopfield's E aptly demonstrates how the dynamics of an abstract neural network may feasibly be related to the physical constructs of power and energy. Hopfield's neurons, though, are single and real-valued, while ours consist of paired excitatory and inhibitory components, and so we require a somewhat different approach.

Given that the network update rule might be rearranged into the following general form:

$$\dot{x}_i = a_i(x_i) \left[b_i(x_i) - \sum_{j=1}^{n} w_{ij} d_j(x_j) \right] \quad 1 \le i \le n, \tag{4.24}$$

with W a symmetric matrix, Cohen and Grossberg (1983) proposed a Lyapunov function:

$$V(x) = -\sum_{i=1}^{n} \int^{x_i} b_i(\xi) d'_i(\xi)\, \mathrm{d}\xi + \frac{1}{2} \sum_{j,k=1}^{n} w_{jk} d_j(x_j) d_k(x_k). \tag{4.25}$$

This subsumes the more specific case above, and is extended by Mendes (1999) to provide a decomposition of such a system into "... one gradient and one Hamiltonian component[...], $\dot{x}_i = \dot{x}_i^{(G)} + \dot{x}_i^{(H)}$...", where:

---

[66] It would make for interesting further study to add the "content" and so obtain a complete expression of the "total power supplied to or extracted from the element" (Jeltsema & Scherpen 2009:31).

$$\dot{x}_i^{(G)} = -\frac{a_i(x_i)}{d'_i(x_i)} \frac{\partial V^{(S)}}{\partial x_i}, \text{ and} \tag{4.26}$$

$$\dot{x}_i^{(H)} = -\sum_j a_i(x_i) w_{ij}^{(A)}(x) a_j(x_j) \frac{\partial H}{\partial x_j}. \tag{4.27}$$

Here $V^{(S)}$ is essentially identical to Cohen and Grossberg's Lyapunov (Equation 4.25), and

$$H = \sum_{i=1}^{n} \int^{x_i} \frac{d_i(\xi)}{a_i(\xi)} \, \mathrm{d}\,\xi. \tag{4.28}$$

This is at once of interest since Mendes goes on to demonstrate the application of this decomposition to "... the Wilson-Cowan model of a neural oscillator without refractory periods, in the antisymmetric case considered by most authors" (1999:4):

$$\begin{aligned} \dot{x}_1 &= -x_1 + \mathrm{S}(\rho_1 + w_{11} x_1 + w_{12} x_2) \\ \dot{x}_2 &= -x_2 + \mathrm{S}(\rho_2 + w_{21} x_1 + w_{22} x_2) \end{aligned} \tag{4.29}$$

Mendes introduces the following change of variables:

$$z_i = \rho_i + \sum_{i=1}^{2} w_{ij} x_j, \tag{4.30}$$

thereby obtaining:

$$\begin{aligned} \dot{z}_1 &= -\frac{\partial V}{\partial z_1} + w_{12} \frac{\partial H}{\partial z_2} \\ \dot{z}_2 &= -\frac{\partial V}{\partial z_2} + w_{12} \frac{\partial H}{\partial z_1} \end{aligned} \tag{4.31}$$

where:

$$V = \frac{1}{2} \sum_i \{z_i^2 - \rho_i z_i + w_{ii} \log(1 - \mathrm{S}(z_i))\}, \text{ and} \tag{4.32}$$

$$H = \sum_i \log(1 - \mathrm{S}(z_i)). \tag{4.33}$$

Mendes' decomposition of an excitatory-inhibitory network into a dissipative gradient-*like* component and a conservative Hamiltonian-*like* component (Howse IV 1995) is suggestive of the kind of abstract energy landscape that we wish to exploit here, but the change of variables makes it less clear how one might interpret these components in terms of physical energy, *per se*. One view is that "... if each x_i denotes the average number of spikes per unit time, then $[z_i]$ could denote the average amount of neurotransmitter in the synaptic clefts at the dendrite endings. This is the same as the total excitation converging from the entire network and external receptors to the *i*th neuron." (Hoppensteadt 1997:22) In fact, the Wilson-Cowan equations have lately been characterised as "... the Euler-Lagrange solution of the minimization of an energy functional" (Bertalmío & Cowan 2009:72; cf. Bertalmío *et al.* 2007), though we have nowhere been able to discern that Lagrangian, explicitly stated, from the literature reviewed. Unfortunately, Mendes' sample application assumes infinitely fast inhibition by forgoing the refractory parameter which turns out to be essential to the task of constructing a graded-frequency neural network (GFNN). Moreover, our own, pedantic efforts to reproduce Mendes' example, as a prelude to our adaptation thereof, have not yielded the published result, undermining our ability to proceed with appropriate certainty.[67] This is a path of enquiry which might yet reward future work.

An even more general approach, though specifically seeking to construct a Lyapunov function for an excitatory-inhibitory network, is that of Seung *et al.* (1998). They too recognise "... what makes the Lyapunov function especially interesting, beyond the convergence results it yields: its role in a conceptual framework that relates excitatory-inhibitory networks to optimization theory and classical mechanics." (1998:330) They proceed from the following network definition:

$$\begin{aligned} \tau_x \dot{x} + x &= f(u + Ax - By) \\ \tau_y \dot{y} + y &= g(\upsilon + B^{\mathrm{T}} x - Cy) \end{aligned} \tag{4.34}$$

with $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$ as state vectors together representing any number of distinct excitatory and inhibitory neural populations, respectively. We can at once see the similarity to Equation 4.29, though this formulation speaks to more than just that more homogeneous case. Significantly, Seung *et al.*'s formulation includes a refractory parameter, τ. On the assumptions of smoothness, boundedness and invertibility in respect of f, they then propose function F as the antiderivative of f, and $\bar{F}$ as the Legendre transform $\bar{F}(x) = \max_p \{px - F(p)\}$, which results in $F'(x) = f(x)$ and $\bar{F}'(x) = f^{-1}(x)$. So too are the "conjugate convex pair G, $\bar{G}$" obtained. They then construct $\Phi : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$ and $\Gamma : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, being generalisations of the "standard" kinetic energies $1/2\ \tau_x \dot{x}^2$ and $1/2\ \tau_y \dot{y}^2$, in the following manner:

$$\begin{aligned} \Phi(p, x) &= \mathbf{1}^{\boldsymbol{T}} F(p) - x^T p + \mathbf{1}^{\boldsymbol{T}} \bar{F}(x) \\ \Gamma(q, y) &= \mathbf{1}^{\boldsymbol{T}} G(q) - y^T q + \mathbf{1}^{\boldsymbol{T}} \bar{G}(y) \end{aligned} \tag{4.35}$$

with the dimensionality of the **1** matrices to be trivially inferred from their contexts. They note that: "[s]etting $p = u + Ax - By$, we obtain the generalized kinetic energy $\tau_x^{-1} \Phi(u + Ax - By, x)$, which vanishes when $\dot{x} = 0$ and is positive otherwise. It reduces to $\tau_x \dot{x}^2/2$ in the special case where f is the identity function." (Seung *et al.* 1998:332) Their Lyapunov function is completed by adding "... a multiple of the saddle function":

---

[67] Mendes' response to our enquiry in this regard was that these decompositions are not unique. He did not venture to comment on the correctness, or otherwise, of the alternative solution that we proposed.

$$S=-u^T x-\frac{1}{2}x^T A\,x+\upsilon^T y-\frac{1}{2}y^T C\,y+\mathbf{1}^T\bar{F}(x)+y^T B^T x-\mathbf{1}^T\bar{G}(y) \tag{4.36}$$

The complete, composite Lyapunov function is then:

$$L=\tau_x{}^{-1}\Phi(u+Ax-By,\,x)+\tau_y{}^{-1}\Gamma(\upsilon+B^T x-Cy,\,y)+rS \tag{4.37}$$

Now, though the validation which is offered in respect of the "kinetic energy" component (setting $f(x)=x$ to produce $\tau_x\dot{x}^2/2$) checks, the rationale which informs Equation 4.35 is not at all clear. We would expect that, given the claimed generality of the formulation, we should obtain reasonable results in respect of simple classical mechanical systems too. Consider $p=m\dot{x}$ then:

$$x+\dot{x}=x+\frac{p}{m} \tag{4.38}$$

$$F(p)=\int x+\frac{p}{m}dp=xp+\frac{p^2}{2m} \tag{4.39}$$

$$\bar{F}(p)=xp-F(p)=\frac{1}{2}m\dot{x}^2 \tag{4.40}$$

$$F(p)-xp+\bar{F}(p)=\frac{p^2}{2m}+\frac{1}{2}m\dot{x}^2 \tag{4.41}$$

Seung *et al.*'s antiderivatives and Legendre transforms invite us to interpret the activation function as defining the constitutive relationship between generalised flow ($\dot{x}$) and generalised momentum (p). This, in turn, suggests that their formulation is not the generalised kinetic energy, as stated, but rather the sum of generalised kinetic energy and generalised kinetic co-energy (Jeltsema & Scherpen 2009:31). In the case of a simple, linear relationship between flow and momentum, kinetic energy and kinetic co-energy are equal, and so we simply obtain double the quantity we'd expect. A non-linear relationship, though, would result in different quantities in each case, with perhaps no relationship evident between the quantities so produced.

As regards the "saddle function", Seung *et al.* give no justification, though it appears that this has been constructed to satisfy:

$$\begin{array}{ccccccc} -\dfrac{\partial S}{\partial x} & = & u+Ax-By-f^{-1}(x) & = & p-f^{-1}(x) & = & \dot{p} \\ \dfrac{\partial S}{\partial y} & = & \upsilon+Bx-Cy-g^{-1}(x) & = & p-g^{-1}(x) & = & \dot{q} \end{array} \tag{4.42}$$

If we are correct in this interpretation, then this "saddle function" incorporates both integrals, with respect to generalised displacement, over the effort components of each of the excitatory and inhibitory neural populations, i.e. potential energy. It is then not clear why a multiplier is required.

Despite our concerns about the details of their implementation, Seung *et al.*'s approach would seem to provide a reasonable roadmap, illustrating as it does the summation of kinetic and potential energy components in the fashion of a Hamiltonian. They purport to take their inspiration from a classical mechanical framework, and ultimately prove convergence of their Lyapunov function over an abstract network. Unfortunately, though a refractory parameter is now included, it seems that the intention is only to cater to homogeneous, single-frequency neural networks. All external input, it would seem, is to be directed via the u and υ parameters, and recurrence managed by way of matrices A, B, and C. Here too, future work might well be rewarded.

Both Mendes and Seung *et al*. approached excitatory-inhibitory networks as pairs of real components. Hoppensteadt & Izhikevich (1996) employ a complex-valued, canonical model instead:

$$z_i{}'=b_i z_i+d_i z_i|z_i|^2+\sum_{j=1}^{n} c_{ij} z_j, \qquad i=1,\ldots,n \tag{4.43}$$

"... where $'=d/d\tau$ and $\tau=\epsilon t$ is 'slow' time; $z_i \in \mathbb{C}$ describes activity of the *i*th neural oscillator; $c_{ij} \in \mathbb{C}$ describes the synaptic connection from the *j*th on the *i*th neural oscillator and $b_i, d_i, \in \mathbb{C}$ are parameters." (1996:129) Adopting complex values seems prudent given that we are interested in oscillatory behaviours. However, the derivation of a normal form, passing as it does through a coordinate transform, followed by a Taylor expansion up to some order, seeks to preserve only the essential characteristic behaviour of a system. We would hope to find the notion of network energy well preserved, regardless. Hoppensteadt & Izhikevich rescale their canonical model as $z_i \rightarrow z_i/\sqrt{|d_i|}$, $c_{ij} \rightarrow c_{ij}\sqrt{|d_j|/|d_i|}$ to produce:

$$z_i{}'=(\rho_i+\mathrm{i}\,\omega_i)z_i-z_i|z_i|^2+\sum_{j=1}^{n} c_{ij} z_j \tag{4.44}$$

where $b_i=\rho_i+\mathrm{i}\,\omega_i$ (1996:129). By introducing a "rotating coordinate system" $\mathrm{e}^{\mathrm{i}\,\omega\tau} z_i(\tau)$, they obtain:

$$z_i{}'=\rho_i z_i-z_i|z_i|^2+\sum_{j=1}^{n} c_{ij} z_j, \qquad i=1,\ldots,n \tag{4.45}$$

Summing integrals taken over all of $z_i{}'$ with respect to its complex conjugate[68] yields the mapping $U:\mathbb{C}^{2n}\rightarrow\mathbb{R}$:

$$U(z_1,\ldots,n_n,\bar{z}_1,\ldots,\bar{z}_n)=-\sum_{i=1}^{n}\left(\rho_i|z_i|^2-\frac{1}{2}|z_i|^4+\sum_{j=1}^{n} c_{ij}\bar{z}_i z_j\right) \tag{4.46}$$

which is shown to be a global Lyapunov function for the system. The proof hereof assumes a homogeneous network with all oscillators having equal frequencies. This, together with a Hermitian connection matrix, allows one to satisfy:

---

[68] These are plainly inverse Wirtinger derivatives, though such is not explicitly mentioned by the authors.

$$\begin{aligned} z_i' &= -\frac{\partial U}{\partial \bar{z}_i}, \\ \bar{z}_i' &= -\frac{\partial U}{\partial z_i} \end{aligned} \tag{4.47}$$

and therefore:

$$\frac{dU}{d\tau}=\sum_{i=1}^{n}\left(\frac{\partial U}{\partial z_i}z_i'+\frac{\partial U}{\partial \bar{z}_i}\bar{z}_i'\right)=-2\sum_{i=1}^{n}|z_i'|^2\leq 0 \tag{4.48}$$

Our attempts to generalise this specific method to the gradient-frequency case have so far been unsuccessful, but we note here, as in most of the methods reviewed thus far, an important feature: all, at some point, integrate with respect to a conjugate quantity. Hopfield's E integrates current with respect to voltage (by analogy, in any event), Seung *et al.* introduces an antiderivative over the activation function, while Hoppensteadt & Izhikevich integrate their complex state value with respect to its complex conjugate.

Hopfield's original energy function for a real-valued, recurrent neural network has more lately been extended to complex-valued networks (Kuroe *et al.* 2002), though with specific constraints. The very endeavour is inherently problematic since: "[c]omplex cost functions are of no interest, because in the field of complex numbers no ordering … is defined and thus minimization or maximization makes no sense." (Fischer 2005:407) Consider their complex reformulation of a Hopfield network in vector notation:

$$\begin{aligned} \mathbf{T}\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} &= -\mathbf{u}+\mathbf{W}\mathbf{x}+\boldsymbol{\theta} \\ \mathbf{x} &= \mathbf{f}(\mathbf{u}) \end{aligned} \tag{4.49}$$

Where $\mathbf{u},\mathbf{x},\boldsymbol{\theta}\in\mathbb{C}^N$, $\mathbf{T}\in\mathbb{R}^{N\times N}$, and $\mathbf{W}\in\mathbb{C}^{N\times N}$.

We therefore require a mapping $\mathbb{C}^n\rightarrow\mathbb{R}$, which is realised by requiring a Hermitian connection matrix $\mathbf{W}^*=\mathbf{W}$. Further, the activation function $\mathbf{f}(\mathbf{u}):\mathbb{C}\rightarrow\mathbb{C}$ must be bounded, but not analytic,[69] and must furthermore be separable into continuously differentiable real and imaginary parts $f^R(u^R,u^I)$ and $f^I(u^R,u^I)$, respectively. Now, if $\mathbf{f}$ is one-to-one and satisfies:

$$\begin{aligned} &\frac{\partial f^R}{\partial u^R}\neq 0 \\ &\frac{\partial f^R}{\partial u^I}=\frac{\partial f^I}{\partial u^R} \\ &\frac{\partial f^R}{\partial u^R}\frac{\partial f^I}{\partial u^I}-\frac{\partial f^R}{\partial u^I}\frac{\partial f^I}{\partial u^R}>0 \end{aligned} \tag{4.50}$$

---

[69] As explained in Kuroe *et al.*, Liouville's theorem implies that a bounded, complex analytic function would be constant, and thus inherently unsuitable as an activation function (2002:1080).

for all $\mathbf{u} \in \mathbb{C}$, then a function $G(x^R, x^I)$: $\mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ exists:

$$G(x^R, x^I) := \int_0^{x^R} g^R(x, 0)\, dx + \int_0^{x^I} g^I(x^R, y)\, dy \tag{4.51}$$

where $\mathbf{g}(\mathbf{x}) = g^R(x^R, x^I) + \mathrm{i}\, g^I(x^R, x^I)$ and $\mathbf{g} = \mathbf{f}^{-1}$. And finally:

$$E(\mathbf{x}) := -\frac{1}{2}\left(\mathbf{x}^* \mathbf{W}\mathbf{x} + \boldsymbol{\theta}^* \mathbf{x} + \mathbf{x}^* \boldsymbol{\theta}\right) + \sum_{j=1}^{n} G(x_j^R, x_j^I) \tag{4.52}$$

is an energy function for the network (Kuroe *et al.* 2002:1080–1).

A considered comparison of Equations 4.52 and 4.23 will readily reveal their commonalities. Summation has partly been replaced by matrix multiplication, but one should still be able to discern the earlier terms $1/2 \sum_i \sum_j T_{ij} V_i V_j$ and $\sum_i I_i V_i$ as $1/2(\mathbf{x}^* \mathbf{W}\mathbf{x})$ and $1/2(\boldsymbol{\theta}^* \mathbf{x} + \mathbf{x}^* \boldsymbol{\theta})$, respectively. It would then appear that Kuroe *et al.* have not extended Hopfield's circuit metaphor to incorporate complex currents, voltages and impedances, but have simply replaced the real values of the original energy formulation with complex values, noting the mathematically-motivated constraints above. Once again, this suggests an interesting avenue for further research.

A key challenge to our adoption of Kuroe *et al.*'s energy function lies in trying to rearrange the canonical model (more particularly, as extended by Large *et al.*, Equation 4.14) to conform to their prototype (Equation 4.49). Our substitutions:

$$\begin{aligned} \mathbf{W} &= \mathbf{c} \\ \mathbf{x} &= \mathbf{f}(\mathbf{u}) = \frac{\mathbf{z}}{\mathbf{1} - \sqrt{\epsilon}\,\mathbf{z}} \\ \boldsymbol{\Theta} &= \mathbf{z}\left(\mathbf{1} + \mathbf{a} + \mathbf{b}|\mathbf{z}|^2 + \frac{\mathbf{d}\,\epsilon|\mathbf{z}|^4}{\mathbf{1} - \epsilon|\mathbf{z}|^2}\right) + \frac{\mathbf{1}}{\mathbf{1} - \sqrt{\epsilon}\,\bar{\mathbf{z}}} \circ \left(\frac{\mathbf{s}}{\mathbf{1} - \sqrt{\epsilon}\,\mathbf{s}} + \mathbf{W}\mathbf{x}\right) - \mathbf{W}\mathbf{x} \end{aligned} \tag{4.53}$$

yield:

$$G(x^R, x^I) = \frac{1}{\sqrt{\epsilon}} x^R - \frac{1}{\epsilon} \ln\left(1 + \sqrt{\epsilon}\, x^R\right) + \frac{1}{2\epsilon} \ln\left(1 + \left(\frac{\sqrt{\epsilon}\, x^I}{1 + \sqrt{\epsilon}\, x^R}\right)^2\right) \tag{4.54}$$

which, together with said substitutions (Equation 4.53), might be used to construct the energy function (Equation 4.52). Given the contortions required of our $\boldsymbol{\Theta}$ term, however, and especially its dependence on $\mathbf{z}$ and $\mathbf{W}\mathbf{x}$, we should not not be entirely surprised if the results obtained were less than satisfactory. Ultimately though, it is the failure of our activation function $\mathbf{f}(\mathbf{u})$ to satisfy conditions 4.50 which sabotages our progression along this path.

The general strategy of integrating by a conjugate quantity bears closer scrutiny. At first, we consider an integral over only the intrinsic terms of the expanded canonical model:

$$\begin{aligned} U := & \quad -\int z_i \left(a + b\,|z_i|^2 + \frac{d\,\epsilon |z_i|^4}{1 - \epsilon |z| + i^2} \right) d\,\bar{z}_i \\ = & \quad \frac{1}{\epsilon^2} d \ln(1 - \epsilon |z_i|^2) - (a + \frac{d}{\epsilon}) |z_i|^2 - \frac{b-d}{2} |z_i|^4 \end{aligned} \tag{4.55}$$

where we adopt the typical parameters as a constraint, namely that $a = 0 + \mathrm{i}\,2\pi$, $b, d \in \mathbb{R}$. By design, this satisfies Equation 4.47, and may therefore trivially be proven to be a Lyapunov function (by Equation 4.48). The moment we incorporate recurrent terms and the refractory parameter τ, though, hermiticity is lost, and the proof breaks down.

Though no extant method could be found whereby to satisfactorily operationalise our notion of energy in the network, we have encountered indications that a first principles approach might bear fruit. We now dispense with the goal of formulating a Lyapunov function *per se*, and simply follow the general pattern established in the approaches reviewed above in pursuit of a plausible formulation of network energy.

In classical mechanics, kinetic energy is the integral of the flow, expressed in terms of the momentum, with respect to that momentum (Jeltsema & Scherpen 2009:30). In the conventional linear case, this yields the familiar quadratic term[70]:

$$\begin{aligned} T(p) & = \int_0^p \hat{f}(\rho)\, d\rho \\ & = \int_0^p \frac{\rho}{m} d\rho \\ & = \frac{p^2}{2m} \end{aligned} \tag{4.56}$$

Integrating our state variable, a complex phasor, by its conjugate obtains:

$$\int z\, d\bar{z} = |z|^2 \tag{4.57}$$

This is a reasonable formulation for the energy of our phasor given that, in conventional cases, energy is the square of amplitude. In our case, we presume that contributions from both real and imaginary components account for the absence of the ½ multiplier.[71] Also, in the absence of a coefficient of proportionality such as mass, a factor of one is assumed instead. Now suppose:

---

[70] The more well-known ½ mv^2 is the kinetic co-energy, equal to the kinetic energy in the linear case only (no relativistic effects).

[71] We have noted, in the course of our own experimentation, that a real signal passed through a Hilbert transform (and thus made analytic) has twice as much energy as the real signal alone. Thus our conclusion above.

$$E = |z|^2 = r^2 = x^2 + y^2. \tag{4.58}$$

If this represents the total energy of the phasor at time t, and noting that r, x and y are functions of t, then taking the time derivative yields an expression for the instantaneous power transfer at time t:

$$\frac{d}{dt}|z|^2 = 2r\dot{r} = 2x\dot{x} + 2y\dot{y} \tag{4.59}$$

We deduce, in general, that each infinitesimal power increment amounts to the sum of products of state variables and their respective rates of change, and so we set out to apply this principle to our network, following somewhat in the footsteps of Susuki *et al.* (2007):

Starting with Large's fully expanded canonical model with an input (Lerud *et al.* 2014):

$$\tau_i \dot{z}_i \;=\; z_i\left(\mathrm{i}\,\omega + \beta_1|z_i|^2 + \frac{\beta_2\,\epsilon|z_i|^4}{1-\epsilon|z_i|^2}\right) + \frac{1}{1-\sqrt{\epsilon}\,\bar{z}_i}\left(\frac{s}{1-\sqrt{\epsilon}\,s}\cdot\frac{1}{1-\sqrt{\epsilon}\,\bar{s}}\right) \tag{4.60}$$

we split into real and imaginary components:

$$\begin{aligned}
\tau_i \dot{x}_i &= \left(\beta_1|z_i|^2 + \frac{\beta_2\,\epsilon|z_i|^4}{1-\epsilon|z_i|^2}\right)x - \omega y + \frac{(1-\sqrt{\epsilon}\,x_i)\,\Re(s) + \sqrt{\epsilon}\,y_i\,\Im(s)}{(1-\sqrt{\epsilon}\,z_i)(1-\sqrt{\epsilon}\,\bar{z}_i)(1-\sqrt{\epsilon}\,s)(1-\sqrt{\epsilon}\,\bar{s})}\\
\tau_i \dot{y}_i &= \omega y + \left(\beta_1|z_i|^2 + \frac{\beta_2\,\epsilon|z_i|^4}{1-\epsilon|z_i|^2}\right)y + \frac{(1-\sqrt{\epsilon}\,x_i)\,\Im(s) - \sqrt{\epsilon}\,y_i\,\Re(s)}{(1-\sqrt{\epsilon}\,z_i)(1-\sqrt{\epsilon}\,\bar{z}_i)(1-\sqrt{\epsilon}\,s)(1-\sqrt{\epsilon}\,\bar{s})}
\end{aligned}. \tag{4.61}$$

We now apply Equation 4.59 and simplify[72] to obtain the following expression for the instantaneous power transfer:

$$P = \frac{1}{\tau_i}\left\{\underbrace{\beta_1|z_i|^4 + \frac{\beta_2\,\epsilon|z_i|^6}{1-\epsilon|z_i|^2}}_{\text{intrinsic}} + \underbrace{\frac{(\Re(z_i) - \sqrt{\epsilon}|z_i|^2)\,\Re(s) + \Im(z_i)\,\Im(s)}{(1-\sqrt{\epsilon}\,z_i)(1-\sqrt{\epsilon}\,\bar{z}_i)(1-\sqrt{\epsilon}\,s)(1-\sqrt{\epsilon}\,\bar{s})}}_{\text{stimulatory}}\right\}. \tag{4.62}$$

We have separately indicated the contributions made by intrinsic dynamics of the system from that made by the external stimulus. Setting the non-linearity parameter $\epsilon = 0$ produces:

$$P = \frac{1}{\tau_i}\left\{\beta_1|z_i|^4 + \Re(z_i)\,\Re(s) + \Im(z_i)\,\Im(s)\right\}, \tag{4.63}$$

which reaffirms the principle that power transfer resides in the product of real components, summed with the product of imaginary components. For real stimulus, the final term is zero. The intrinsic term(s) are typically characterised by $\beta_1, \beta_2 < 0$, and so result in dissipation (given $0 \le \epsilon \ll 1$). Equilibrium is reached when these two contributions balance. Integrating Equation 4.62 over a time interval, say t_0 to t_1, will then

---

[72] We dispense with the factor of 2, since our abstract formulation does not relate to physical units.

give us the system's change in energy over that period. At equilibrium, on average, this will be zero. This relation is termed the “energy balance relation” for the system (Jordan & Smith 2007:125–129; Susuki *et al.* 2007), here framed in terms of a complex state variable and complex stimulus, with separate integrals indicating the energy dissipated by internal dynamics and energy supplied by the external stimulus:

$$\begin{aligned} \Delta E &= S(z_{t_1}) - S(z_{t_0}) \\ &= \frac{1}{\tau_i}\int_{t_0}^{t_1} \beta_1 |z_i|^4 + \frac{\beta_2 \epsilon |z_i|^6}{1-\epsilon |z_i|^2}\, dt \\ &+ \frac{1}{\tau_i}\int_{t_0}^{t_1} \frac{\left(\Re(z_i) - \sqrt{\epsilon}|z_i|^2\right)\Re(s) + \Im(z_i)\Im(s)}{(1-\sqrt{\epsilon}z_i)(1-\sqrt{\epsilon}\bar{z}_i)(1-\sqrt{\epsilon}s)(1-\sqrt{\epsilon}\bar{s})}\, dt \end{aligned} \qquad (4.64)$$

Equation 4.64, by our observations, behaves as expected. As the system approaches equilibrium, dissipation tends to a minimum, and thus the system requires less stimulation in order to maintain a given energy level. In other words, energy transfer efficiency attains its maximum at equilibrium. The amplitude and phase response curves of each neuron may be analytically determined with respect to a simple sinusoidal stimulus as follows (Jordan & Smith 2007:29). From Equation 4.59 we trivially derive a general expression for the amplitude response of a phasor:

$$\dot{r} = \frac{x\dot{x} + y\dot{y}}{r} \qquad (4.65)$$

Similarly, given $\tan\theta = y/x$, we may derive an expression for the phase response:

$$\begin{aligned} \frac{d}{dt}\tan\theta &= \frac{d}{dt}\left(\frac{y}{x}\right) \\ \dot{\theta} sec^2\theta &= \frac{x\dot{y} - \dot{x}y}{x^2} \\ \dot{\theta} &= \frac{x\dot{y} - \dot{x}y}{r^2} \end{aligned} \qquad (4.66)$$

Consider a real stimulus $\alpha\cos(\omega t + \varphi)$, where both α and φ change slowly enough so that they might each depend on t over the long term, but may be treated as constants in the short term.[73] Since our stimulus is periodic, we wish to even out short-term fluctuations by averaging over one complete cycle of the stimulus. Our stimulus term, taken from Equation 4.63 (only the real part) and averaged over one cycle, yields the following:

$$\frac{\omega}{2\pi}\int_0^{\frac{2\pi}{\omega}} \cos(\omega t)\cos(\omega t + \varphi)\, dt = \frac{1}{2}\cos\varphi \qquad (4.67)$$

Therefore, we can substitute Equations 4.64 and 4.67 into 4.65, set $\dot{r} = 0$, and so obtain:

---

[73] This is the essence of the so-called “two-timing” method (Strogatz 1994:218–223).

$$\frac{\alpha}{2r}\cos\varphi = -\beta_1 r^2 - \frac{\beta_2 \epsilon r^4}{1-\epsilon r^2} \tag{4.68}$$

Similarly, for the phase response Equation 4.66, we set the response $\dot{\varphi} = \tau\nu$, so to obtain:

$$\frac{\alpha}{2r}\sin\varphi = \omega - \tau\nu \tag{4.69}$$

With two equations in two unknowns, we can now alternately eliminate φ and r in turn to obtain implicit response curves for oscillators at equilibrium:

$$\left(\frac{\alpha}{2r}\right)^2 = \left(\beta_1 r^2 + \frac{\beta_2 \epsilon r^4}{1-\epsilon r^2}\right)^2 + (\omega - \tau\nu)^2 \tag{4.70}$$

$$\frac{4(\omega-\tau\nu)^2}{\alpha^2 \sin^2\varphi} + \frac{\beta_1 \tan\varphi}{(\omega-\tau\nu)} = \epsilon\left(1 + \frac{\beta_1 - \beta_2}{4(\omega-\tau\nu)^3}\,\alpha^2 \sin^2\varphi \tan\varphi\right) \tag{4.71}$$

We plot these for arbitrary, typical parameter values ($\beta_1 = -1$, $\beta_2 = -1$, $\epsilon = 0.1$), and for different stimulus amplitudes ($\alpha = 0.1$, $\alpha = 1.0$) to obtain the following:

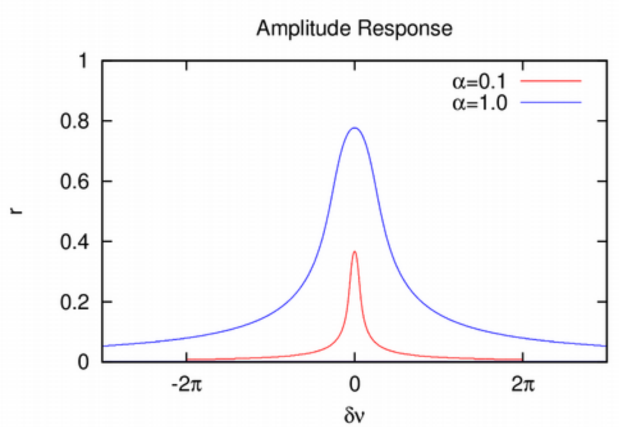


Figure 4.11: Amplitude response

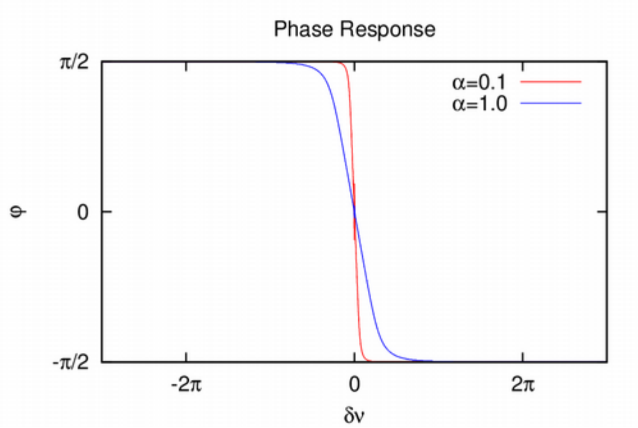


Figure 4.12: Phase response

Varying the parameters, we find that the marked resemblance to a probability density function (PDF) and cumulative distribution function (CDF) persists. Most notably, the response to low-amplitude stimuli is quite sharply tuned, while the response to a higher-amplitude stimulus is comparatively broader. Our Python numerical simulations reveal more of the dynamic nature of the neural oscillator. Under the same parameters, we see the build-up to equilibrium, and the sawtooth-like waveform which results there, followed by exponential decay.

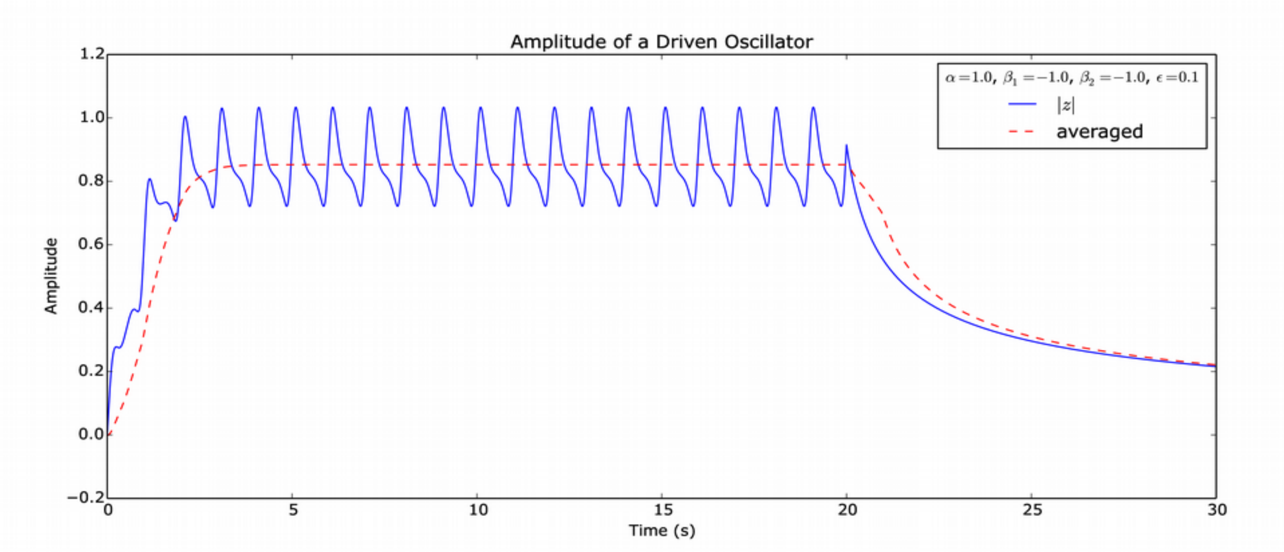


Figure 4.13: Numerical simulation amplitude plot

We separately plot the dissipated (intrinsic) and driven (stimulatory) components of Equation 4.62:

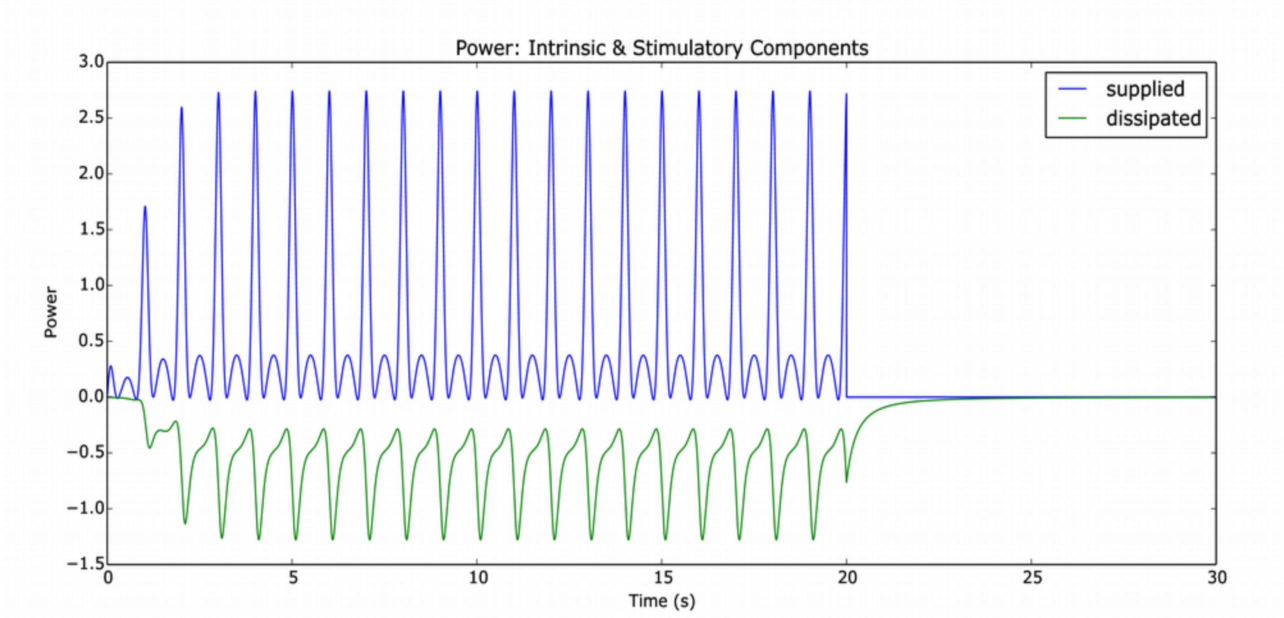


Figure 4.14: Intrinsic (dissipatory) & stimulatory (driven) components

At equilibrium, these two components, on average (over a complete stimulus cycle), balance each other out, implying that power supplied by the stimulus matches what is being lost to dissipation.

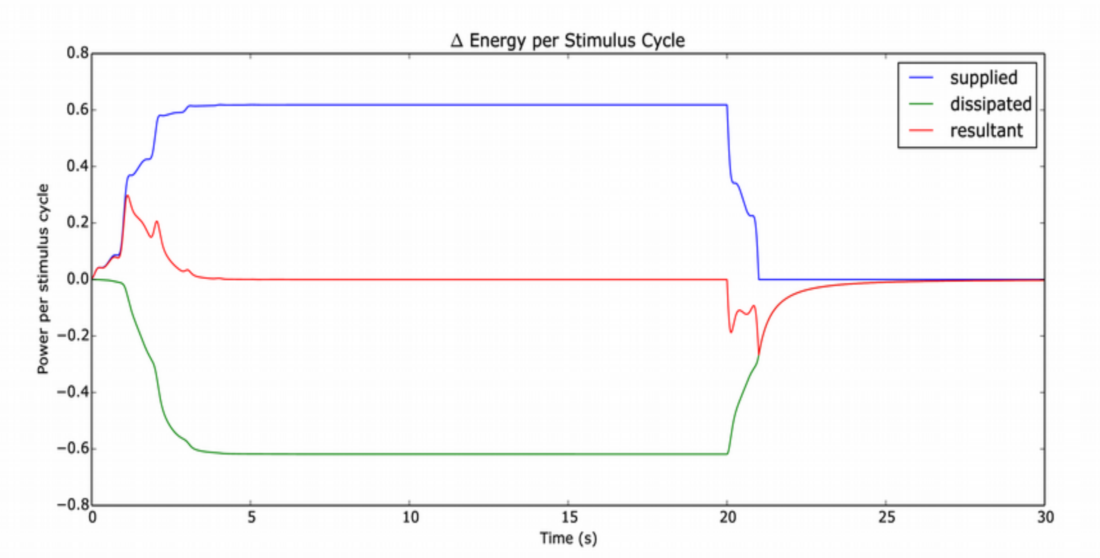


Figure 4.15: Energy flow into and out of a single oscillator

Obtaining analytic expressions which describe the neural oscillator's response to more complex real-world stimuli is beyond our reach, but we take the above as reassurance that the equation performs as expected. Equation 4.64 then serves to operationalise our notion of energy flow into and out of the network.

4.6 Concluding Remarks

We have discussed the various qualities which endear a complex-valued, non-linear, artificial neural network to our purpose. Moreover, we have surveyed various approaches which seek to account for energy flows in comparable networks. In the absence of any suitable extant approach, we have extracted general principles and formulated our own. Analysis of simple cases, together with numerical simulation, support our formulation as appropriate to the goal. We trust that our implementation will further sustain this impression.

Chapter 5

Designing & Implementing Experiments

In the preceding chapters we have framed musical teleology in terms of an abstract "energy landscape" and proposed a mechanism (a complex-valued, non-linear, artificial neural network) and a measure (our energy formulation) whereby to explore that landscape. We hypothesise that this measure will show a correlation between musical expectation and efficiency of energy transfer. In this chapter we will describe the development of appropriate apparatus and experimental procedures for the purpose of testing that hypothesis.

5.1 Implementation of an Experimental Tool

We constructed a software tool with which to perform exploratory experiments in the problem space. Specifically, we implemented a modular processing "pipeline" in C++, essentially an expanded derivative of the "Producer-Consumer" pattern.[74] This pipeline allows us to arbitrarily chain various functional modules in whatever order we wish, each of these modules performing its particular processing function on a designated buffer which is external to the module itself. Identically-sized buffers are provided to each module, and all buffers reside in a rotating double-ended queue. Thus the queue of buffers may be viewed as a data conveyor-belt of sorts, with processing modules operating somewhat independently of each other, and ultimately enabling me to take advantage of parallelism in modern processor architectures. Input options include basic synthetic tones generated on demand, pre-recorded audio files, and real-time input from installed audio devices. Output may be streamed to the outputs of audio devices, and/or written to a file on disk in standard formats such as WAV. The most important intermediate modules, and the *raison d'être* of this tooling, are those which pass the incoming data through a "gradient-frequency neural network", or GFNN (Large *et al.* 2010).

Addressing a preliminary issue: one might expect that such a framework as this has already been implemented numerous times, and rightly ask why it be done yet again?[75] Certainly, the MARSYAS tool-kit (Tzanetakis n.d.) featured prominently in our early explorations, and it was our intention to contribute the necessary *MarSystem(s)* to that project in order to leverage the advantages of proven stability and cross-

---

[74] We have yet to properly study and distinguish between Mattson's (and others') "pipeline" pattern, and the producer/consumer pattern posited by, for example, Doug Lea or Mark Grand. What we have implemented is according to our own naïve design, rather than a strict implementation of either of these.

[75] The implementation by Yuan (Yuan n.d.) warrants further study, and (time permitting) a comparative evaluation, though this came to our attention only after our current implementation had been completed.

platform convenience already present therein. However, once we looked closely at the details of the internal "implicit" patching of MARSYAS[76], we found that substantial changes would need to be made to the current architecture in order to accommodate recurrent networks.[77] We came to similar conclusions in our cursory reviews of various alternatives, and thus resolved to implement a leaner solution to address our own requirements more directly. We hope, thereby, to be assured of greater insight into the mechanisms investigated by denying ourselves, to a greater extent than would have been so otherwise, some of the typical "black-box" luxuries of extant frameworks. We also believe this approach to be necessary in order to fruitfully pursue any possibility of real-time system operation, a highly desirable feature in such an exploratory study, and one which had remained elusive in our prototyping thus far. Nonetheless, our own design is clearly influenced by what we have learned from our time spent with MARSYAS, in particular. It remains to be rigorously investigated, but we do believe that this is a unique contribution: that this framework eclectically combines and addresses specific requirements of both an audio processing pipeline and those of a recurrent neural network implementation in a way that is currently not well serviced by any single framework. However, whilst endeavouring to keep the system architecturally quite open and extensible, we have given little attention to the development of modules required by any exploration of issues beyond the focus of this research: being the behaviour of a GFNN consisting of non-linear oscillators of a sort recently proposed by Edward Large, and particularly in the ability of such to account for important musical phenomena in a novel, culture-neutral way.

Our "pipeline" then encapsulates a dynamic,[78] ordered list of "processing modules", together with an accompanying double-ended queue of data buffers. When the pipeline receives a "tick" signal, it checks that each buffer is valid: meaning that it exists, and has not been exhausted by the previous tick. Each buffer has an attendant counter which monitors the point in that buffer to which processing had last proceeded, and from which it should now continue. When not restricted by external factors, modules are constructed to consume entire buffers at a single tick, but some modules employ the buffers in smaller increments by virtue of their dependence on factors such as hardware design. Such is the case, for instance, with the real-time audio input and output modules, with RtAudio (Scavone n.d.) (upon which these are based) specifying its own buffers (and their sizes) which need to be filled by the callbacks provided in our modules. Nonetheless, for each non-exhausted buffer, a tick to the network issues a tick to the attendant processing module, which results in a further section (or perhaps all) of the buffer being processed accordingly. Once all buffers have been exhausted, the next tick rotates the queue and resets the counters of each buffer before continuing as above. Two special cases merit noting: the incremental initial allocation of buffers; and their incremental

---

[76] This, as opposed to the more ubiquitous "explicit" patching approach (Bray & Tzanetakis 2005).

[77] Feed-forward networks, on the other hand, are trivially implemented in MARSYAS.

[78] We use the term "dynamic" only in reference to our perspective as programmer. At compile-time, we are able to construct processing pipelines according to arbitrary requirements. To the user of such a program, the pipeline is static, of course, and no provision is made for dynamically altering its structure at run-time, such functionality falling wholly outside of the purpose for which this tool is designed and implemented.

deallocation. Initially, the queue has length equal to the number of processing modules in the pipeline, but contains empty place-holders at all positions. The first tick, detecting an unallocated buffer in the last position along the queue, allocates a new buffer in that position, before rotating the queue and proceeding to process the first module. Since all other buffers are unallocated at that point, their ticks are not processed. Subsequent ticks will process only the first *n* modules, allocating new buffers as required in precisely the same way until all buffers are allocated. As a simple optimisation when rotating, buffers are not re-allocated or re-initialised, but are simply re-used with counters reset to initial positions. A single flag is raised when an end-of-file condition is reached, whether by encountering the actual end of an input file, or having produced a synthetic tone of the desired length, or by way of a signal from the user. Once this flag is raised, buffers are deallocated in much the same fashion as they have been allocated above.

When it became clear that real-time computation of the evolving artificial neural network's (ANN) state would easily stretch our computing resources beyond their means, we resolved that the solution would require a multi-threaded approach. Initially, this was implemented by means of the thread class in C++11, along with attendant mutexes, locks and condition variables, but we subsequently replaced that implementation by exploiting OpenMP instead (OpenMP ARB n.d.), which we believe to be a far more elegant solution within the context of this research. As it now stands, each network tick initiates an OpenMP task in respect of each processing module, which is assigned to an available thread in OpenMP's thread pool. Threads may execute in any order, exploiting whatever parallelism is available on the system, but the pipeline's tick blocks until all threads are complete, thence proceeding to the next tick. This approach favours the use of small buffers and reasonably long processing chains, resulting in lower latency as measured from the input to the output of the pipeline as a whole (as compared to a strictly serial, single-threaded approach, that is). However, the performance gains turned out to be wholly inadequate to address the demands introduced by the computation of network state in a GFNN of any complexity. Thus we resolved that parallelism should also be introduced, as far as possible, into the computation of that state. Parallelisation of the primary loop which iterates over each of the oscillators in the GFNN, as measured on an 8-core CPU, saw computation time falling to approximately 25% of what had been measured for the same code running on one processor only. By all accounts, this seems to be the upper end of what one reasonably expects, considering the additional overhead incurred by thread management. There remained some smaller loops which were tempting to parallelise even further, but to date such attempts have only resulted in degraded performance, likely due to additional thread management overhead not being offset by the relatively smaller gains in such tight loops. In attempting to further reduce the time taken to do that computation, we then allowed the network to take as its buffer size the size of an entire input file, or of an entire generated signal of a specified length (at a given sampling rate). This, of course, effectively reduced the potentially parallel computation of individual processing modules in the pipeline (as implemented above) to a serial one again, and thereby eliminated the possibility of real-time input, in exchange for the maximum available computing capability afforded to the state computation. This approach yielded the most promising results, and set the standard against which the later introduction of an Eigen-based implementation was measured (see below).

The specifics of the GFNN now merit attention. Our earlier review of the ongoing program of research by Large and associates, focusing more or less on the explanatory and predictive capabilities represented by their expanded canonical model (as applied to the domain of music), convinced us of the value of pursuing this approach. We set about attempting to duplicate a number of their published results, discovering in due course that the computational requirements were substantial, even by current standards. Our first partially successful attempt to reproduce the graph in Large, Almonte & Velasco (2010),[79] for instance, required slightly less than a week of run-time to compute (and that only after many failed attempts). Edward Large kindly pointed out that we should consider abandoning our reliance on a fixed-step, fourth order Runge-Kutta solver in favour of an adaptive method, after which we adopted the Dormand-Prince (dopri5) method, which appeared to most closely match the adaptive Runge-Kutta (RK45) method which Large *et al* had employed in their own studies, using MatLab (personal communication). This reduced our own computation time to minutes, but had implications for the architecture we had adopted. Though our initial fixed-step RK4 approach was our own trivial implementation, we thereafter elected to exploit the ODEINT library (Ahnert & Mulansky n.d.), a recently introduced component of the popular BOOST libraries, for its DOPRI5 implementation, not least of all because it showed itself to be faster than any implementation we were able to build, and moreover came highly recommended. Nonetheless, the learning curve was steep.

The Large, Almonte & Velasco (2010) study employed a GFNN consisting of 361 oscillators, these tuned to span a range from 3 octaves below 1 Hz, to 3 octaves above. Thus each was spaced regularly (logarithmically) at a 60th of an octave from the next, giving exactly 5 oscillators covering the range between any two adjacent notes of the modern piano keyboard. In other words, each oscillator was spaced 20 cents from the next. Significantly, human pitch perception, under ideal conditions, is quite capable of detecting differences on the order of 10 cents, and even less. Moreover, the range of hearing is typically cited as 20 Hz – 20 kHz[80], which spans 10 octaves. Thus, though the centering of their GFNN on 1 Hz was plainly a mathematical convenience, the span and spacing of their oscillators was quite conservative. Moreover, as we discovered, raising the centre frequency of both the network and stimulus by two orders of magnitude (to 100 Hz) required shortening the length of the stimulus by the same (to 0.5 sec) in order to maintain comparable computation time (approximately 2.5 sec, still a far cry from real-time).[81] Since the GFNN is

---

[79] The stimulus in that study is a 50 sec sine tone at 1 Hz.

[80] ... though 16 Hz – 18 kHz is a far more realistic specification, and that only for a brief portion of an average human's lifespan, starting in puberty.

[81] This statement must further be qualified by stating that the particular set of parameters employed were somewhat arbitrary at this point, and with minimal neural interconnectivity, the intention merely being to illustrate a key difference between the time scales employed in the studies under review, as opposed to the application of those methods to real audio stimuli.

recurrent, connections within any layer are $O(n^2)$, and that in addition to the $O(n^2)$ afferent and efferent connections between layers. The computation of these matrices, then, potentially presents a significant bottleneck. Considering that, in one of our early implementations, such a matrix had dimensions of 361 by 361, consisting of complex doubles (each occupying 16 bytes), the total ask might be 3 × 2,085,136 bytes, or just shy of 6 MB of stack allocation per layer. Of course, heap allocation then presents as an option, though only by accepting other trade-offs (e.g. paging, fragmentation, etc.) associated with that strategy.

Having explored such performance issues from various perspectives, we have lately come across the Eigen library (Guennebaud *et al.* 2010), wherein we believe we have found something approaching an optimal solution. Eigen provides a suite of tools for linear algebra. Most importantly, it prioritises speedy computation, which it achieves primarily by exploiting three distinct mechanisms. Firstly, it is a header-only, template-based library, implying that only the necessary components are ever compiled and linked into the user's executable, rather than carrying the bloat of a pre-compiled library loaded with unnecessary functionality. Secondly, it is built on the technique of "expression-templates", meaning that the compiler is coerced (by lazy evaluation) into compiling maximally efficient code from the potentially verbose representations employed by a human programmer. Thirdly, Eigen aggressively exploits the availability of vectorisation in particular computing architectures, and otherwise falls back to sane defaults. Eigen is also reportedly aware of multiple processors, thought the details hereof remain sketchy. What is quite clear is that implementations of our use case based on Eigen, when compared to earlier implementations which simply looped over the individual state computations, albeit supported by parallelisation via OpenMP, fared at least comparably well if not better. These earlier implementations, however, were incapable of scaling to encompass the full range of inter-neural connectivity that was required (as discussed above). Eigen, on the other hand, presented the option of employing sparse matrices, an eminently sensible technique given the specifics of the problem space. To clarify, we expect that Hebbian learning, combined with the harmonic resonance which is characteristic of non-linear oscillators, will favour the development of connections between oscillators which are tuned to integer multiples of each other, and thus the expectation of sparse connection matrices. Of course, it lies amongst the aims of this research to validate this assumption, rather than to impose it *a priori* on the problem space, and so we have approached this matter with some circumspection.

Exploiting Eigen allowed us to treat the entire GFNN more or less as a single entity, delegating determination of the optimal iteration strategy to the compiler. This stands in contrast to the canonical approach of solving individual neural states within an explicitly supplied iterative loop, and moreover makes for a somewhat cleaner representation of the system at hand. In particular, the expansion of our earlier experiments, having single-layer ensembles of non-linear oscillators, to the multi-layer case would have become wholly intractable without the availability of something akin to sparse matrices. It should be noted, though, that there is little reliance here on the most idiosyncratic linear algebra operations, such as matrix multiplication, for instance. Instead, most operations are simply element-wise applications of their common forms, as seen in the extensive use of Hadamard products (i.e. $A \circ B$ rather than AB). On the other hand, outer products are exploited in the computation of the various connection matrices. If one understands the Hebbian learning rule, $\dot{c} = -\delta c + \kappa z \bar{z}$, as the composition of a "forgetting process" with a rate controlled by δ and a "learning process" with rate controlled by κ, then $z \bar{z}$, considered in linear algebra terms, becomes the product of a vector-encapsulated layer of states and its conjugate transpose.

The other significant bottleneck pertains directly to the adoption of an adaptive solver. As mentioned above, earlier experiments with fixed-step ODE solvers yielded disappointing results which would have rendered much of this study computationally intractable. Adaptive methods fared much better, but presented two specific challenges: they require an essentially continuous stimulus input; and they favour the largest buffer sizes possible. Continuous input is required because the solver, in response to the error computed at any particular step of its iteration, is free to reject that computation, roll back, and take a smaller step in the hope of attaining acceptable error the next time around. In the alternate, if error is acceptably low, the step size is increased, resulting in faster completion. Thus it is quite conceivable that the solver might ask to take a step that is in no way related to the sampling rate of the digital stimulus provided. Two solutions are apparent: either the solver must merely be presented with the closest sample to that requested, or an appropriately interpolated value must be computed, given the samples available. Both methods have been explored, and all indications are that, while the latter method is computationally more intensive, the former results in greater error, in turn resulting in the solver requiring a vastly greater number of steps to reach a solution. Nonetheless, we have taken care to implement a sinc interpolation,[82] optimised as best as we are able, and have made it possible to trivially switch between these two strategies in a bid to explore this issue further. To the other challenge, it is clear that the DOPRI5 solver is most efficient when it has unconstrained control of its own step size. The partitioning of input into buffers, whilst reducing latency and enhancing the potential for effective parallelisation, invariably counteracts many of the very same gains achieved in this way, since step size is reset at each tick, and can in any event not overstep the boundaries of the buffer. Thus, while we have retained this mode of operation as an option for further investigation, we have resolved that our experiments will be conducted by way of pre-recorded stimuli, rather than live real-time input (as was initially hoped).

Future work will revisit some of the avenues that have been considered in trying to circumvent the issues outlined above. NVIDIA's Compute Unified Device Architecture (CUDA), for instance, was considered as an alternative to OpenMP, which would have required us to obtain appropriate hardware. Our cursory

---

[82] Upon considering the available alternatives, we did not think it appropriate to employ either spline or linear interpolation under these specific circumstances. Sinc interpolation is a reasonable choice given its direct relation to and derivation from the Fourier Transform. See also the compelling case made in Yaroslavsky (2007).

reading, however, indicates that CUDA really only yields the advertised benefits once 1,000,000 or more calculations are performed in parallel (Demidov *et al.* 2012). For that matter, Field Programmable Gate Arrays (FPGA) warrant further investigation too. However, we do not believe that this investigation has reached that point yet. The biggest advances remain in carefully considered parallelisation of the code, and particularly of those computations involving the connection matrices. Furthermore, the issue of "false sharing" requires further investigation. This commonly occurs when a cache line is shared amongst several processors, as easily happens when multiple threads are simultaneously working on the potentially adjacent elements of an array (as might occur in the ODE computations or connection matrix updates). A write access to one element of that cache line "dirties" the line, signalling that all other processors must first retrieve a fresh copy before doing their own writes. This slows progress substantially, and is typically avoided by either artificially spacing array elements far enough apart to avoid their coexistence on one cache line (resulting in tremendously inefficient memory use), or by assigning threads to blocks of iterations (as we have done – see above). Ideally, this issue, too, will be revisited.

After all of the aforementioned, it might now be pointed out that the GFNN modules are not here being employed as pipeline processing modules *per se*. Rather, they are merely hooking into the "data conveyor belt" described earlier in order to have access to basic file-reading / writing / auditioning and signal generation capabilities.[83] For the moment, at least, the GFNN simply reads the data in the buffer as an external stimulus, fed into the network along with the various weighted inputs arriving from the outputs of all *other* oscillators in the *same* network layer, as well as the weighted inputs arriving from the outputs of *all* oscillators in the *precedent* and *antecedent* layers. The output of the final layer is three-fold. Firstly, the states of the individual oscillators may be viewed as the outputs of a filter bank, with appropriate graphing of the magnitude and phase of each indicating the degree of resonance experienced by that filter in response to the stimulus, considering its time evolution. In effect (and in the case of a single-layer architecture), this is substantially similar to a spectrogram of sorts. Secondly, the implementation of an appropriate learning rule introduces dynamic changes to the network which, it is hypothesised, ultimately reinforce optimal response to audio stimuli that are routinely recognised as "musical" (Large 2011:121–123). Plotting the connection matrix after it has been allowed to evolve in this way is likely to lend credence to a notion that certain musical relations are innately privileged, given what we currently think we know about human auditory perception beyond the auditory nerve. The third aspect of the GFNN's output, which is likely unique to this study, is the set of observations relating to the flow of energy into and out of the network. In all of these cases, a suitable graphing solution is required. Having explored various options[84], we have ultimately settled

---

[83] This is not to say that the output of the GFNN might not be useful to pass on for further processing, just that it is not required for the moment.

[84] Two other options have been investigated beyond a cursory review: libpng (Schalnat et al. n.d.); and QCustomPlot (Emanuel Eichhammer n.d.). The former serves as a rapid prototyping tool, giving a reasonably efficient way to generate raster plots of the various outputs discussed above. The latter has been studied and extended in order to provide a drop-in Qt widget that would enhance a future packaging of the pipeline framework as a GUI application.

on gnuplot-cpp, a C++ adaptation of Nicolas Devillard's pipe-based C interface to gnuplot (Conlin & Guha 2009).[85]

5.2 Validation of the Tool

The software tool was validated by presenting various ideal stimuli and comparing the output so obtained to our theoretical expectations. To begin with, an earlier numerical simulation (in Python, presented in Figures 4.13, 4.14, and 4.15) was repeated with our tool, again producing results congruent with our discussion in Section 4.5. Beyond the earlier single-neuron results, though, we can now observe the behaviour of an ensemble of neural oscillators, each resonating to a different frequency, and all potentially interconnected in arbitrarily complex configurations. We build, for instance, a single-layer network consisting of non-linear oscillators of the expanded canonical sort presented earlier, here tuned across a range of 6 octaves, centred at 1Hz, and spaced at intervals of 20 cents (60 logarithmically-spaced bands per octave). As before, we adopt arbitrary, though somewhat typical parameter values ($\beta_1 = -1$, $\beta_2 = -1$, $\epsilon = 0.1$), and drive the network with a 20 second sinusoidal stimulus having frequency equal to the centre frequency of the network (octave offset 0), and unit amplitude ($\alpha = 1.0$).

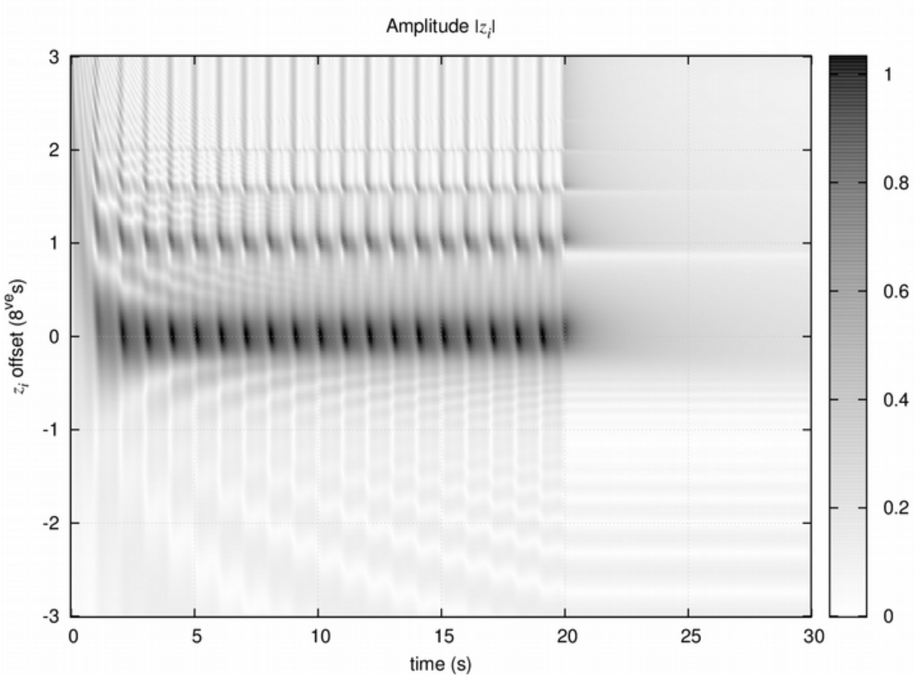


Figure 5.1: Network amplitude response to simple, sinusoid stimulus

It is at once apparent (see Figure 5.1) that oscillators tuned to frequencies other than that of the stimulus

---

[85] gnuplot-cpp was initially written by Rajarshi Guha, and later passed on to Jeremy Conlin, the current maintainer. See http://gnuplot.10905.n7.nabble.com/Announcing-C-interface-to-Gnuplot-td12436.html [Accessed October 30, 2015].

are being activated. A broad band of oscillators around the centre frequency are all seen to resonate, not at their own frequency, but at the frequency of the stimulus, albeit with progressively larger phase-shift further from that centre frequency. This is agrees with our earlier analytic formulations of amplitude and phase response (see Figure 4.11 and Figure 4.12). Also, resonant bands appear at integer multiples of the resonant frequency, again with phase shift as above. Given this particular parameter regime, all oscillator amplitudes decay exponentially to zero at the cessation of the stimulus. We re-run the same simulation over a somewhat longer period, this time summing the stimulatory and dissipatory components of each neuron in the ensemble to show that network equilibrium is indicated in the same way as for individual neurons (see Figure 5.2).

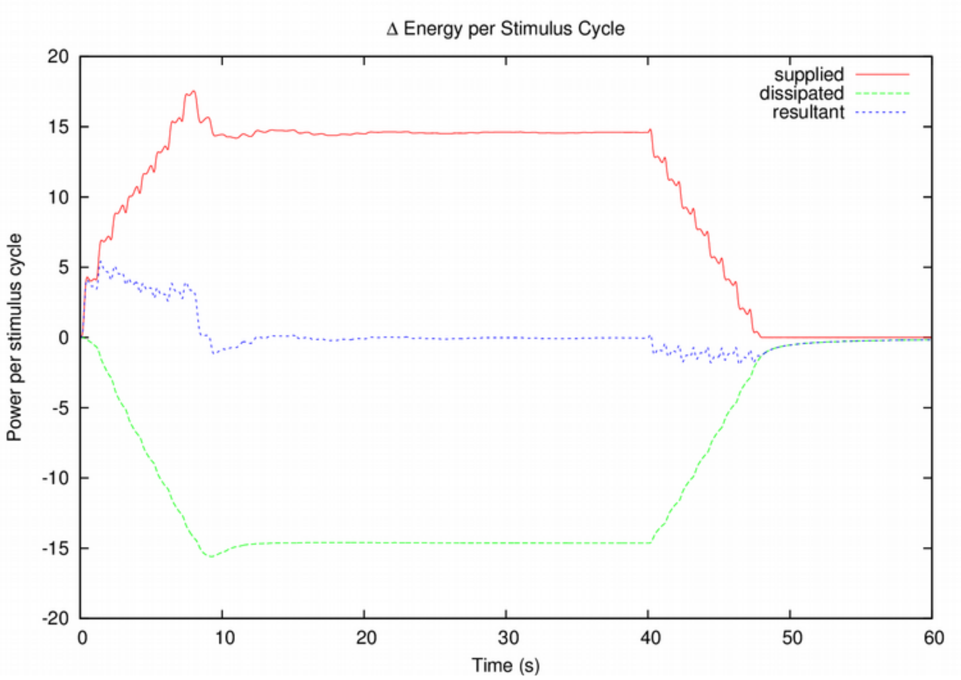


Figure 5.2: Network energy flows under simple sinusoidal stimulus

It bears noting that, in order to obtain a relatively flat resultant at equilibrium, the averaging period in this case had to be extended to 8 times that of the stimulus: an indication of low-amplitude sub-harmonics added by non-linear interactions. Still, caveats notwithstanding, our energy balance relation appears to perform as expected under theoretically ideal conditions. In particular, we are able to thereby discern the net inflow or outflow of "energy" as the network tends towards equilibrium upon a change of stimulus, and further to discern that equilibrium by the "energy" balance attained.

5.3 Experimental Objectives

Our aim is now to discover whether the transition from dominant to tonic harmonic function is most efficient, in the physical sense described earlier, amongst all possible harmonic continuations. Given the "energy" measure developed herein, can we show a correlation between efficiency and expectancy? In order to answer this question, we intend to “prime” the network with a dominant quartad and then to present, in turn, a range of possible harmonic continuations whilst measuring the “energy” required to bring the network to equilibrium in each case.

5.4 Experimental Design

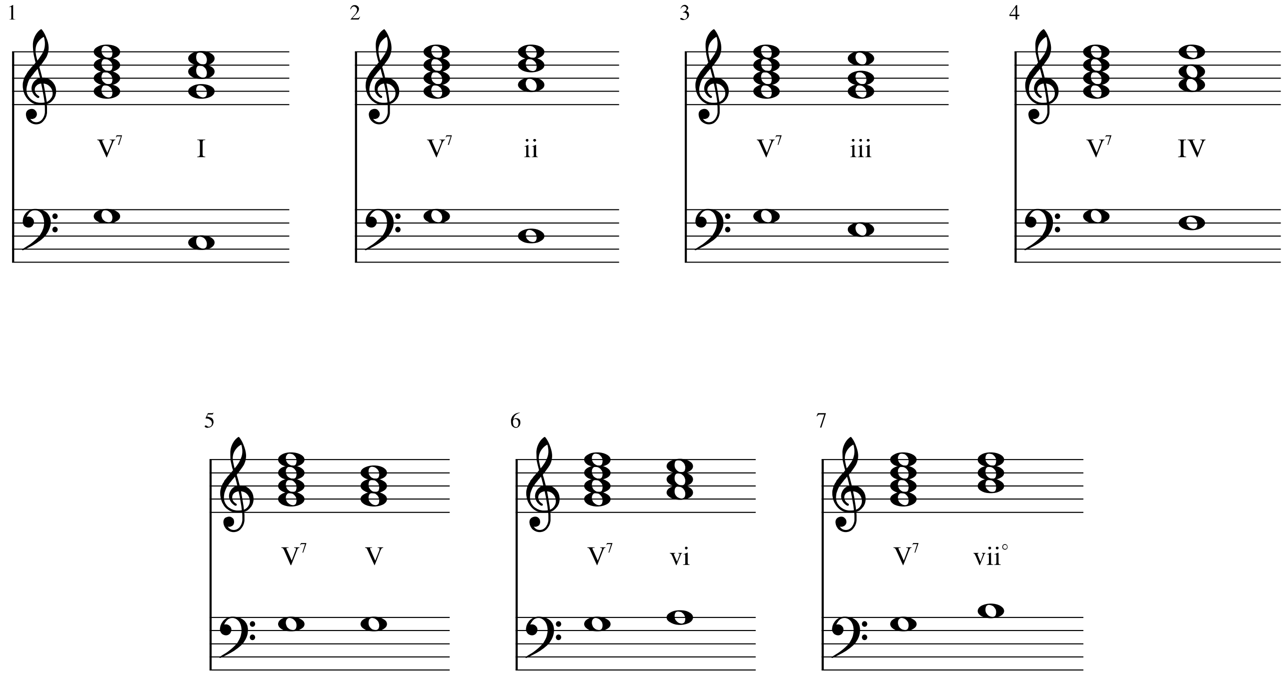


Figure 5.3: Test Progressions

From a canonical music-theoretic perspective, the top stave of each of the progressions in Figure 5.3 is a maximally smooth progression from the antecedent dominant quartad to a consequent triad. The bottom stave is included only to indicate the implied harmonic root movement, and does not form part of the experimental stimuli Once the visual symmetries of the above have been appreciated, the “smoothest” progression might appear to be V^7 – vi (numbered 6 in Figure 5.3), though this progression is, in fact, the musical archetype of “surprise”, forming the duly named “interrupted” cadence. Rather, the most expected progression amongst those presented, canonically speaking, is V^7 – I (numbered 1 in Figure 5.3), termed the “perfect” cadence. Without recourse to Western music theory, there seems to be no reason why this should be so.

5.5 Experimental Method

Our experimental tool is a custom test-bench implemented in C++ and running on an Ubuntu 14.04 LTS Linux PC (Intel® Core™ i7-4770 CPU @ 3.40 GHz, 16 GB RAM). For all of the following experiments, the system is configured as a single-layer GFNN spanning 3 octaves above and below 441 Hz (4 cents sharp of A4), and each octave is further subdivided into 60 geometrically equal frequencies, each to which an oscillator is tuned: thus a total of 361 uniquely tuned resonators. Besides the unique resonant frequency of each oscillator, all other parameters were set identically to each other, specifically to operate within a Hopf regime ($\alpha=0$, $\beta_1=\beta_2=-1$, $\epsilon=0.1$). Stimuli were initially synthesised as pure sinusoids at each component frequency, tuned to 12-tone equal-temperament with A4 = 440 Hz, with real-world audio introduced in the final experiment. In each case, the experiment proceeds in two phases. First the network is primed by the introduction of a priming stimulus, specifically a dominant quartad in closed position on G4, lasting 1 second (see Figure 5.3). The network state, specifically the complex state values of each oscillator, as well as all complex-valued connectivity between and within layers, is then saved. In the second phase, a single chord from a set of possible diatonic continuations, again lasting 1 second, is presented to the primed network. The response is monitored and recorded at each step of the numerical solver in terms of an energy balance relation developed earlier. At the end of each test case, the network state is restored to what it was directly after priming. This is repeated for each of the seven target stimuli.

Succeeding experiments varied the network configuration in order to glean some understanding of the mechanism underlying the results obtained. Experiment 1 employed no recurrent connectivity and no learning. Experiment 2 introduced a non-linear learning rule for the duration of the prime, once again operating in a Hopf regime ($\lambda=0$, $\mu_1=\mu_2=-100$, $\epsilon=0.1$, $\kappa=0.1$), but then disabled the learning rule and held the connectivity configuration static for the duration of the target. Experiment 3 allowed the learning rule to continue to modify connectivity during the stimulus period, resetting connectivity at the end of each target stimulus to the freshly primed state. Experiment 4 repeated the protocol with PCM recordings of a piano performing precisely the same notes.

5.6 Data Analysis

For each of the stimuli, appropriate averaging periods were determined partly by inspection, partly by autocorrelation, but ultimately by selecting the fundamental implied by the components of each stimulus. Generalised power metrics obtained at each step of the numerical solver were then averaged, interpolated and plotted in order to explore the flow of energy into and out of the network. In particular, the measure representing the supply of energy by the stimulus to the network was fitted to a regression model in order to identify its likely equilibrium. The model-fitting parameters obtained in this way are to be presented as primary evidence in this study.

Chapter 6

Results, Analysis & Discussion

The preceding chapter presented the experimental methods that were employed in collecting the data which form the subject of the present chapter. Here we report selected results and subject the same to analytic scrutiny. In each experiment, our goal was to establish a correlation between greatest musical expectancy for a tonic triad following a dominant quartad, as opposed to any other diatonic continuation, and our measure of energy supply by the stimulus to the network.

6.1 Experiment 1: No Neural Plasticity

This experiment, featuring a network without any neural interconnection and without the capability to develop such by dynamic learning processes (neural plasticity), sought to establish whether the effect might reside purely within the intrinsic dynamics of the non-linear oscillators. The priming phase is illustrated in the spectrogram below (Figure 6.1).

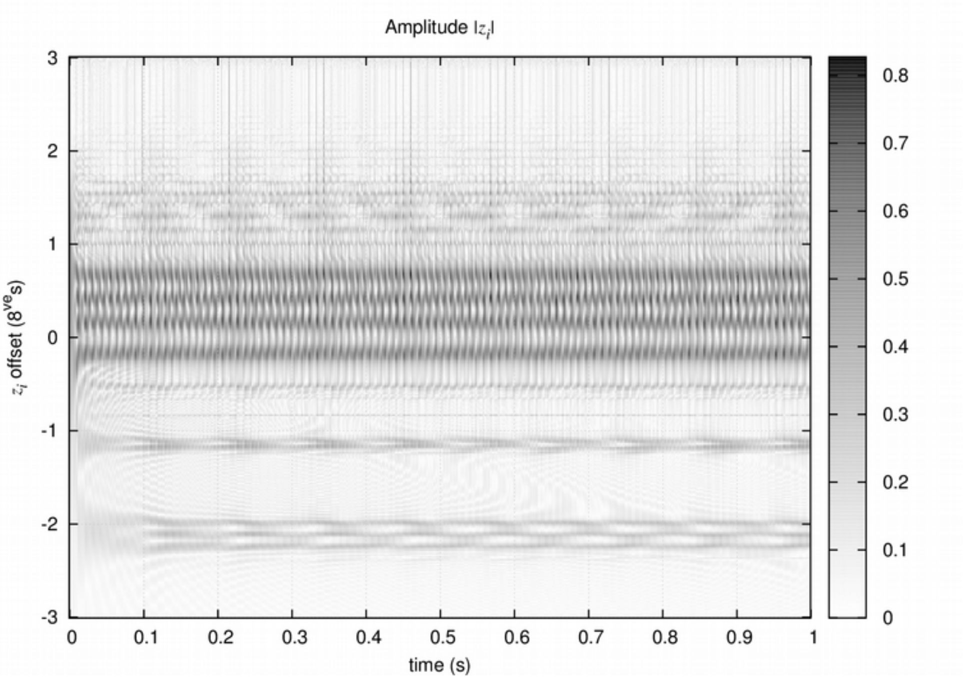


Figure 6.1: Spectrogram of priming by dominant quartad

We note that relative stability appears to have been reached in under 200 ms, and super- and sub-harmonic components are plainly evident. Presentation of a tonic chord to the primed network results in some oscillators' amplitudes decaying, some stirring up, whilst others continue in much the same fashion as before (see Figure 6.2). The principle which holds our interest here is how the cumulative waxing and waning of oscillators at such transitions (from prime to target), viewed in terms of our energy balance relation with respect to the entire neural ensemble, might serve up evidence of the tonic continuation being "most efficient" in some sense.

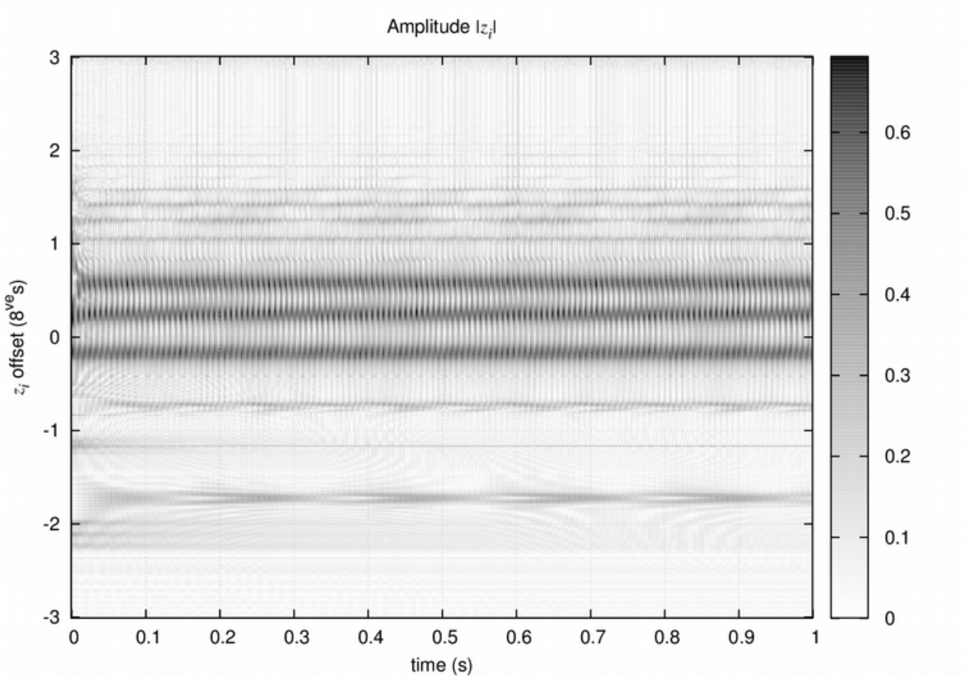


Figure 6.2: Spectrogram of tonic resolution without neural plasticity

In Figure 6.3 we see that all supplied energy comes from the stimulus, since the absence of interconnections eliminates the possibility of any contribution due to feedback. Stimulatory and dissipative components quickly settle into relatively slower oscillatory balance, whose resultant averages to zero over that slower period. However, we found that this is essentially the same for every other of the seven targets, save for variation in the period of oscillation. These lower oscillations are sub-harmonics of the fundamentals being employed, artefacts of our non-linear processing. We ignore these as we set about fitting a straight-line model,

$$y = mx + c \tag{6.1}$$

to the "externally supplied energy" data.

Stimulus	***m***	***c***	***r*** ^{2}
I	-0.035	2.017	0.007
ii	0.108	1.977	0.035
iii	0.051	1.969	0.009
IV	-0.003	2.061	0.000
V	0.016	2.090	0.001
vi	0.022	2.093	0.001
vii°	0.005	2.141	0.000

Table 6.1: Model fit parameters for externally supplied energy in experiment 1

It is evident from the above that, though the model fit is not particularly close (having ignored the periodicities still evident in the data), there is also little reason to believe that any of these targets would settle to levels far different from 2. Slope values close to 0 would seem to further corroborate that view. Thus we find nothing here to sustain the belief that any one target enjoys any efficiency advantage over any other in this network configuration.

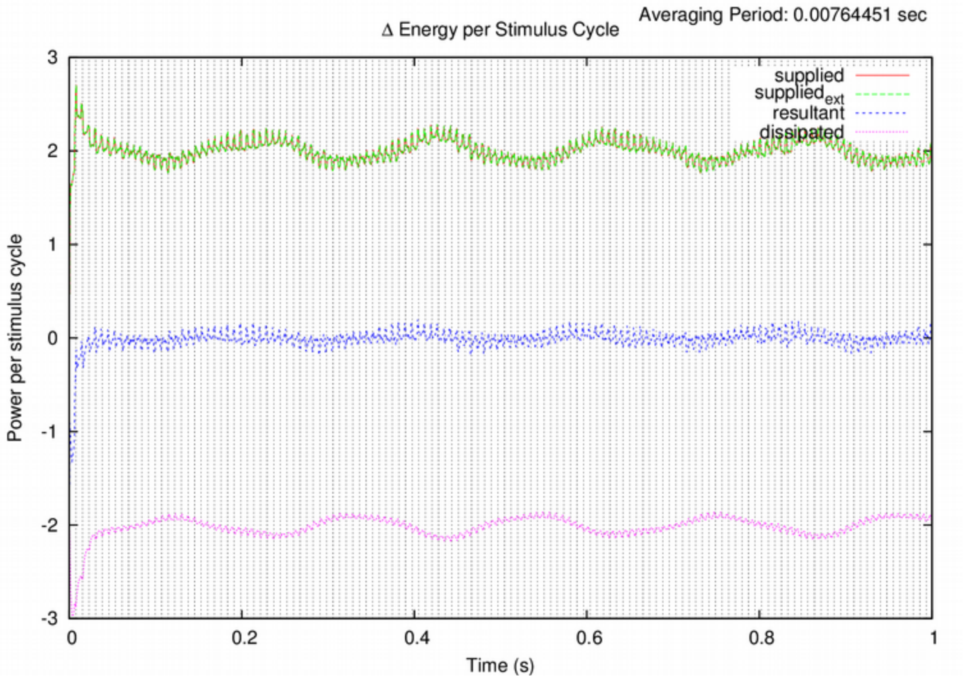


Figure 6.3: Energy balance relation for tonic target without neural plasticity

6.2 Experiment 2: Long-Term Neural Plasticity

We next considered the impact of connections forged between neurons by Hebbian learning. Connections were allowed to form during the priming phase, with that connectivity configuration being held static during the target phase. This closely resembled the typical train-and-predict cycle, and sought to test the notion of connectivity as a passive component, essentially static as regards immediate musical perception, but nonetheless a factor over the longer term by virtue of slow learning processes. The connectivity which emerged evidently favours harmonic relatedness, with a particularly strong association between frequency components of the prime and those lying an octave above each of these (see Figure 6.4). Closer inspection reveals emerging connections at intervals of a twelfth too (third harmonic) and possible higher. The white diagonal illustrates that self-connection was not permitted.

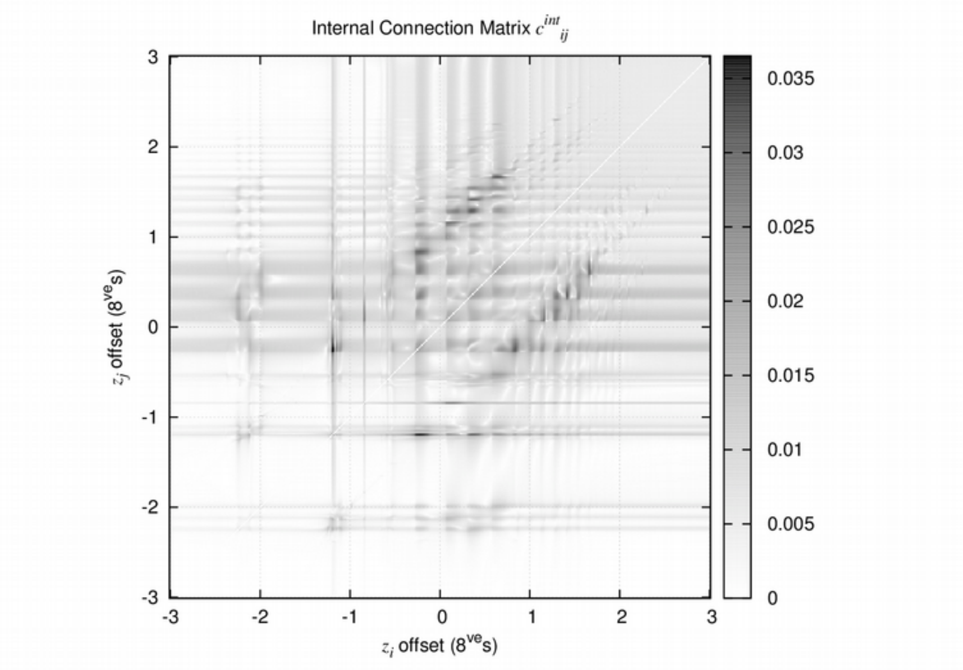


Figure 6.4: Internal connectivity arising due to priming by a dominant quartad

In Figure 6.5 we plot a shorter time-frame at the onset of the target stimulus, in which we now see the lesser contribution being made to the energy balance relation by the external stimulus. A relatively constant contribution, accordingly, is being made by feedback due to the internal connectivity developed between neurons during priming.

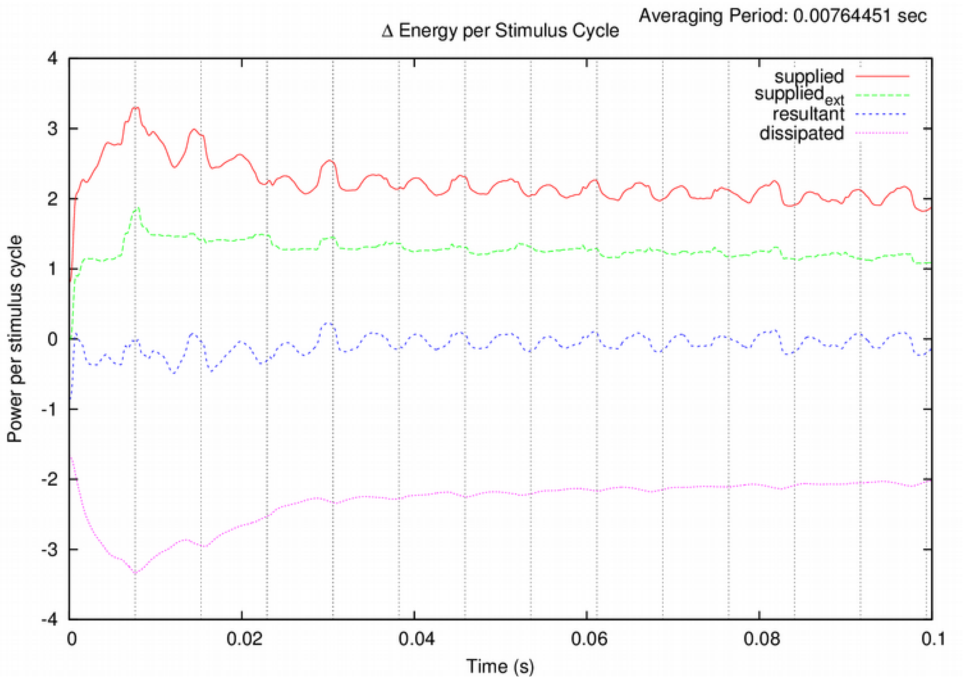


Figure 6.5: Energy balance relation for tonic target with priming plasticity

Over the full term of the stimulus, every one of the test cases quickly settles to a different equilibrium, which we yet again model with a straight line (Equation 6.1).

Stimulus	m	c	r^2
I	-0.027	1.269	0.006
ii	0.040	1.230	0.015
iii	0.001	0.874	0.000
IV	0.012	1.624	0.001
V	-0.049	0.636	0.040
vi	0.042	1.786	0.005
vii°	-0.063	0.896	0.027

Table 6.2: Model fit parameters for externally supplied energy in experiment 2

As before, the model fit is not particularly impressive given that we have ignored slower periodicities which are still plainly evident in the data, but the slope values suggest that these intercepts are, in fact, close to whatever equilibrium would ultimately be reached. Most importantly, these equilibria are now different in respect of the various targets. Specifically, we find that the continuation of the priming dominant quartad by a dominant triad requires the least energy supply, whereas subdominant (IV) and submediant (vi) continuations require considerably more. This is sensible given that we are holding static those connections forged by the dominant quartad, thereby introducing bias against any stimulus which does not exploit that connectivity to increase its own efficiency. The dominant triad (V) does this best, since all three of its notes have already been imprinted upon the connectivity configuration by the prime. The leading tone triad (vii°), though also sharing all of its notes with the dominant quartad, is slightly less efficient, being overtaken by the mediant (iii), which share two of its three notes with the dominant quartad. Between these highest and lowest values lie the tonic (I) and supertonic (ii) triads. These results do not support our hypothesis, but do allow us to put aside these more simplistic accounts for the moment. Barring the eschewing of our selected network architecture for another, we either require appropriate connectivity to arise from some other, longer term process than a simple 1 second prime, or learning must be allowed to continue during the target phase. We next elected to explore the latter option.

6.3 Experiment 3: Short-Term Neural Plasticity

Whereas the energy balance relation contributions due to internal network connectivity remained fairly constant during the target phase of the previous experiment (due to connectivity being held fixed), ongoing learning during the target phase of this experiment caused the relative contributions of internal feedback and external stimuli to shift over the course of the target phase as neural connectivity continued to evolve. Once again we present the response to the tonic triad as an example, though the general characteristics are found to be common to each of the target cases. See Figure 6.6.

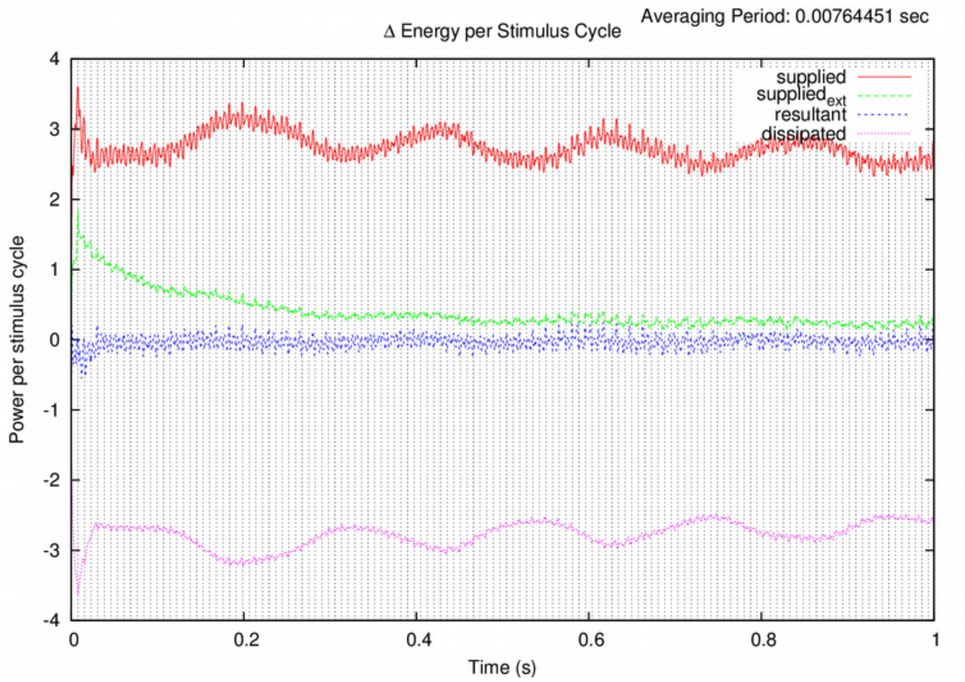


Figure 6.6: Energy balance relation for tonic target with sustained plasticity

The result is a marked exponential decay as the externally supplied energy decreases in response to the increasing contribution made by newly formed neural connections. Accordingly, we adopt the following model with a view to discovering the asymptotic lower limit of the decay:

$$y = N_0 e^{-\lambda t} + \mu. \tag{6.2}$$

Once again, we choose to ignore slower periodicities which are still evident in the data. It is therefore quite surprising to find relatively good r^2. Moreover, our parameter of interest (μ) turns up precisely the result we have been searching for:

Stimulus	N_0	λ	μ	r^2
I	1.124	7.100	0.243	0.941
ii	0.876	5.328	0.369	0.799
iii	0.601	5.867	0.295	0.857
IV	1.371	5.945	0.325	0.918
V	0.448	4.469	0.343	0.819
vi	1.441	6.724	0.327	0.868
vii°	0.734	6.810	0.417	0.781

Table 6.3: Model fit parameters for externally supplied energy in experiment 3

By this analysis, a neural network of the sort employed here, primed by a dominant quartad, does indeed favour the progression to its tonic triad above all other diatonic triads by virtue of the greater efficiency enjoyed by the tonic triad in supplying energy to the network. This is indicated by the lowest μ-value being associated with the tonic stimulus (I). Still, the case presented is highly idealised, and with no reasonable account to be given of the other values obtained for μ, we would do well to proceed cautiously.

6.4 Experiment 4: Real-World Stimuli

We take a leap forward now to consider the response to real-world stimuli. For the prime and each target, dry, monophonic PCM audio recordings were made of a piano (44.1 kHz, 16 bit). The protocol of experiment 3 was repeated, this time substituting the audio recordings for the sinusoidal stimulus employed previously.

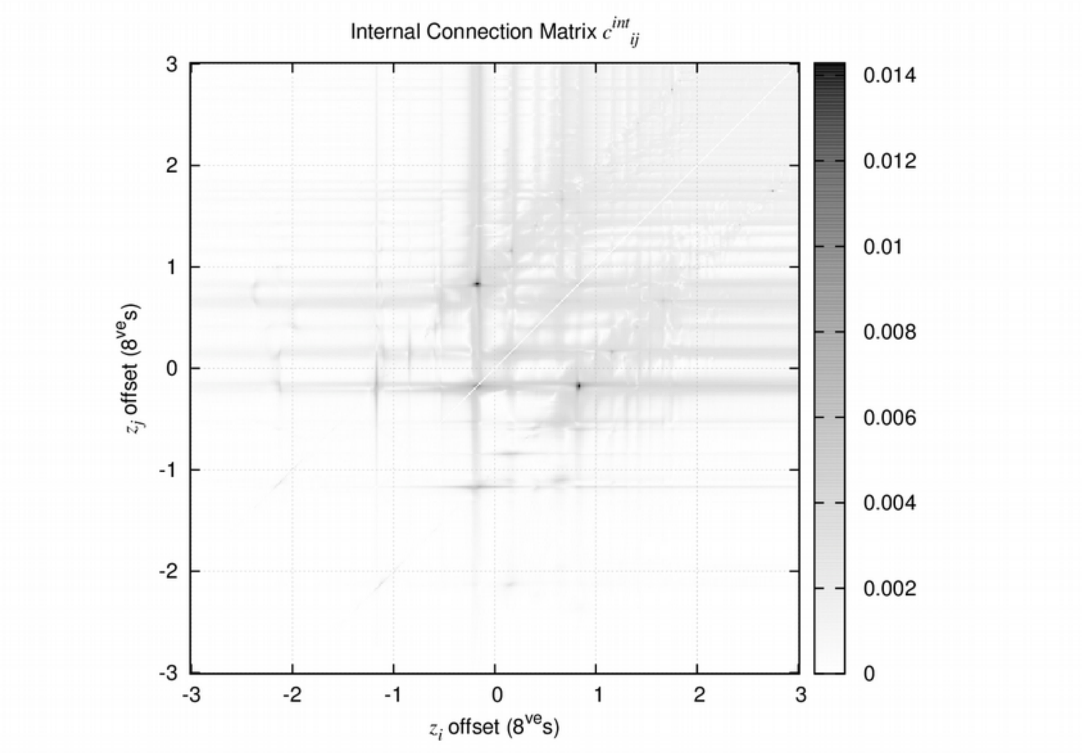


Figure 6.7: Internal connectivity arising due to priming by a dominant quartad played on a piano

Once more, learning favoured harmonic relations, and particularly octaves, though the details were less pronounced than in the earlier, more idealised cases. As before, each of 7 diatonic triads were presented as possible continuations, and the same energy metrics recorded in each case. The energy balance relation produced in this case, though, differs in notable respects from those produced in the preceding experiments (see Figure 6.8). Firstly, the stimulus employed here decays, as opposed to the constant amplitude of the sinusoids employed thus far, and this is evident from the graph. Secondly, we know that the piano tone is spectrally more complex than the sinusoidal stimuli, though we do not yet know precisely how this impacts on the response. We do notice that the externally supplied energy quickly falls to zero, as it does for every one of the other targets in this experiment. This undermines rationale of our experiment, since it implies that all target stimuli are maximally, and thus equally favoured by this particular arrangement. Prolonging training, too, might have produced results more congruent with our previous experiment.

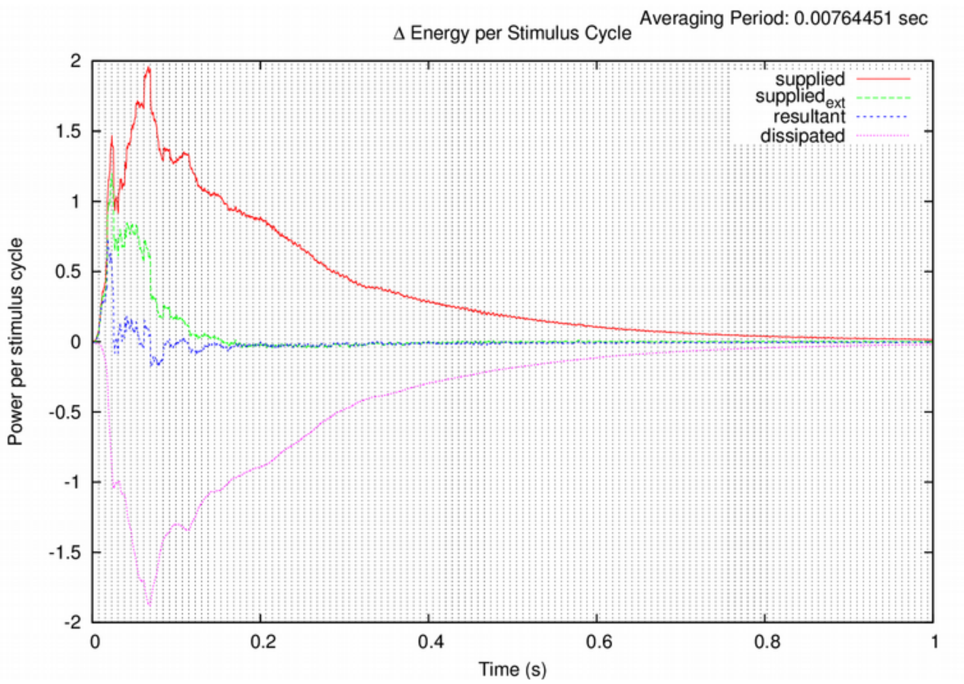


Figure 6.8: Energy balance relation for tonic target employing piano timbre

We confirm this result by the fitting the same model as before (Equation 6.2).

Stimulus	N_0	λ	μ	r^2
I	0.740	11.449	-0.016	0.610
ii	0.462	9.699	-0.005	0.602
iii	0.883	11.146	-0.028	0.517
IV	0.804	9.539	-0.009	0.610
V	0.658	11.232	-0.018	0.579
vi	0.604	8.974	-0.007	0.655
vii°	0.867	13.846	-0.011	0.504

Table 6.4: Model fit parameters for externally supplied energy in experiment 4

Again, though the fit is far from ideal, the most important observation concerns the values obtained for μ. Specifically, while these values bring contradictory evidence to that observed in the previous experiment, all further suggest a strong tendency to zero. This, of course, is the result of employing decaying stimuli. We will need to carefully reconsider our methods in order to extend our tentative findings to more robust contexts such as these.

6.5 Discussion

The results presented in this chapter fall far short of providing a universal account of the mechanisms supporting teleology in tonal music. However, they do make a strong case to sustain further inquiry along these lines. Experiments 1 and 2 established, within the limits of our design and selected parameters, that no target enjoyed any efficiency benefit purely as a result of the intrinsic dynamics of the oscillators. The introduction of full neural plasticity in experiment 3, on the other hand, served up a marked differentiation between different targets, and moreover singled out the most expected target, canonically speaking, as the most efficient continuation, given the particular prime employed.

Our finding is that it now appears quite feasible that musical expectation be viewed, to a far more significant extent than ever before, as a low-level, data-driven process, as against the prevailing account of a top-down, cognitive process. Considering that this result was obtained for a single-layer network with quite arbitrarily selected parameters, it would be reasonable to consider whether closer attention to these details might yield more robust results with more general application. Velasco & Large (2011), for instance, found enhanced ability to follow syncopated rhythms in a double-layer (as opposed to single-layer) network. The greater computation time required precludes us from following that line of enquiry in this study, but we will doubtless return to this in later work. Besides the Hopf parameter regime employed throughout our experiments, simple linear, limit cycle, or other novel regimes may yet be found to play a role in such a multi-layer, ostensibly more complete model.

Does this account, tentative as it is, seek to subsume the totality of musical teleology? It would be ludicrous to dismiss the role of the socio-cultural scaffolding which exists around every single musical endeavour. Expectations arise in numerous ways, but our research has sought to reaffirm schematic, rather than veridical expectations in tonal music. Most importantly, though, a positive result in this research would lay the foundation for a new, cross-cultural approach to musical teleology beyond the confines of common music notation. That positive result is not entirely forthcoming as yet, but it does appear to be close at hand.

6.6 Concluding Remarks

We have presented our experimental results and considered their implications. Though the response to our research question is essentially inconclusive from a scientific realist's point-of-view, we have encountered evidence which demands further investigation. In the final chapter, we will regroup and consider the way forward.

Chapter 7

Conclusions, Recommendations & Future Work

This research set out by considering the vital role of musical teleology in enabling affective response within the context of Western tonal music, but noted that our understanding thereof was bound to a culture-specific theoretical framework, and thus not generaliseable. We therefore proceeded to construct an alternative which would reveal that teleology at the level of the audio signal, forgoing the need to transcribe the signal into an intermediate, higher-level representation and thus avoiding the cultural bias inherent therein. After introducing the problem area more generally in Chapter 1, we motivated our enquiry by further elaborating on the themes of cultural lensing and musical teleology in Chapter 2. In Chapter 3 we conducted a naïve exposition of the intuition which leads us to believe that our quest has any merit whatsoever. We also report, in the same chapter, on two extensive literature studies (included in full as Appendix B and Appendix C), both conducted within the scope of this project, in which we reviewed evidence for the existence of the phenomena under scrutiny beyond being a mere theoretical convenience. Specifically, harmonic priming indicates a cognitive processing advantage conferred on particular musical progressions, while our review of African music theory suggests abstract structural parallels, though manifesting differently, between musical schemata in Western and African music. Our experimental construct emerged from these investigations as a maximally efficient mapping of one musical sonority to another. In Chapter 4 we set about operationalising that construct by leveraging the unique properties of a recurrent, complex-valued, gradient-frequency artificial neural network, furthermore electing to employ an abstract energy balance relation to account for the relative efficiency of one musical progression relative to another. The resulting *quasi*-Lyapunov has never been proposed before in respect of a complex-valued neural network, nor for any other network that we are aware of. In Chapter 5 we described our custom implementation of an experimental tool, a digital signal processing pipeline incorporating options supporting our enquiry, followed by a description of experiments to test our hypothesis. Chapter 6 presented results deemed pertinent to the analysis and discussion, with a more complete set of raw results relegated to Appendix D. The present chapter considers the implications of all the foregoing.

7.1 Conclusions of this Research

Under our selected network architecture and one very specific set of network parameters, and by employing sustained, purely sinusoidal stimuli, we were able to distinguish the most commonly used target chord (a tonic triad in second inversion) from all other diatonic completions that might follow on from a dominant quartad in root position. Given that this was done without reference to the actual notes employed, but rather in terms of overall network energy, we suspect that our method might reflect a fundamental truth about the nature of musical perception. In particular, we suggest that musical expectation is shaped by network dynamics, and especially by a preference for a most efficient continuation. This study fell short of interrogating this notion rigorously and generally, serving up only limited support for the idea, but the scene is certainly set for a more extensive enquiry. On the back of the results obtained thus far, we now recognise the central role of non-linear dynamics and neural plasticity in such a mechanism. We also recognise that we have explored only the smallest part of the parameter space, one particular network architecture, and only a very particular kind of musical expectation. Still, complex-valued neural networks such as ours have elsewhere shown themselves to be equally responsive to matters of rhythm as to matters of harmony, and so we believe that our method will be equally adept in dealing with music whose teleological implications might be embedded in the rhythmic domain as in the spectral domain.

The most profound implication of this research, tenuous as it is, is that tonality might not be an arbitrary cultural construct after all. Perhaps tonal music is the way it is precisely because of how we respond to it physiologically. If our argument can be sustained by way of more rigorous experimentation, we might yet catch a glimpse of precisely how such music comes to evoke expectations prior to the onset of cultural conventions, and thus to better understand its affective agency. This might even lead to a radical reassessment of precisely what is meant by the terms “tonal music”.

7.2 Recommendations & Future Work

First, we recommend careful comparison of our results against those obtained by means of Large *et al.*'s GrFNN Toolkit for MATLAB. Since this toolkit was released late in our own work, and since we did not have the use of MATLAB, we elected to defer such comparison to later work. Further study should then be invested into extending and generalising the results obtained thus far. Various other conventions in Western tonal music may be understood as teleological, in the sense that they engender strong expectations, both harmonic and rhythmic, and our model should be reviewed to account as broadly as possible for these. We must also explore the implications of alternative harmonic voicings, and particularly the effects of timbre. Most importantly, we must incorporate a mechanism to deal with decaying sounds. This would likely be an additional “memory” layer, added to the network and operating in a limit cycle parameter regime. In examining the time-progression of learning in the network, it became evident that the details of connection phase response are significant, and so require more careful analytic study. It then seems plausible that we might find a way to relate connectivity configurations to the Krumhansl-Kessler tone profiles, given that we properly understand that phase response. We are also not entirely happy with our system's handling of transitions between stimuli, and lack of windowing. Since our method currently relies on averaging by stimulus period, and since periodicity detection is non-trivial given real-world stimuli, we expect that some effort will required in developing a more robust solution. Ideally, we would like to see real-time operation, and a number of optimisation opportunities do remain that might advance that goal. Alongside these developments, we must adapt the harmonic priming paradigm to the rhythmic domain, so that any signs of teleology in non-Western music, albeit harmonic or rhythmic, can be verified according to that method.

References

Agawu, V. K. (1986) 'Gi Dunu,' 'Nyekpadudo,' and the study of West African Rhythm. *Ethnomusicology*. 30(1), pp. 64–83.

Agawu, V. K. (1987) The rhythmic structure of West African music. *The Journal of Musicology*. 5(3), pp. 400–418.

Agawu, V. K. (1992) Representing African music. *Critical Inquiry*. 18(2), pp. 245–266.

Agawu, V. K. (1995a) The invention of 'African rhythm'. *Journal of the American Musicological Society*. 48(3), pp. 380–395.

Agawu, V. K. (1995b) *African rhythm: a Northern Ewe perspective*. Cambridge, UK: Cambridge University Press.

Agawu, V. K. (2001) An African understanding of African music. *Research in African Literatures*. 32(2), pp. 187–194.

Agawu, V. K. (2003) *Representing African music: postcolonial notes, queries, positions*. London: Routledge.

Agawu, V. K. (2006) Structural analysis or cultural analysis? Competing perspectives on the 'standard pattern' of West African rhythm. *Journal of the American Musicological Society*. 59(1), pp. 1–46.

Ahnert, K. & Mulansky, M. (n.d.) *ODEINT* [online]. Available from: http://www.odeint.com (Accessed 6 November 2014).

Anku, W. (1997) Principles of rhythm integration in African drumming. *Black Music Research Journal*. 17(2), pp. 211–238.

Anku, W. (2000) Circles and time: a theory of structural organization of rhythm in African music. *Music Theory Online*. 6(1), . [online]. Available from: http://www.mtosmt.org/issues/mto.00.6.1/mto.00.6.1.anku_essay.html (Accessed 5 May 2013).

Anku, W. (2007) Inside a master drummer's mind: a quantitative theory of structures in African music. *TRANS - Transcultural Music Review*. 11(5), . [online]. Available from: http://www.sibetrans.com/trans/article/122 (Accessed 5 May 2013).

Antonopoulos, I. *et al.* (2007) Music retrieval by rhythmic similarity applied on Greek and African traditional music. In *Proceedings of the 8th International Conference on Music Information Retrieval*. pp. 297–300.

Arom, S. *et al.* (2004) *African polyphony and polyrhythm: musical structure and methodology*. Cambridge, UK: Cambridge University Press.

Avorgbedor, D. (1987) The construction and manipulation of temporal structures in 'Yeʋe' cult music: a multi-dimensional approach. *African Music*. 6(4), pp. 4–18.

Baily, J. (1996) Commentary. *Psychology of Music*. 24(2), pp. 114–117.

Barbour, J. M. (1972) *Tuning and temperament: a historical survey*. New York: Da Capo Press.

Bargh, J. A. & Chartrand, T. L. (1999) The unbearable automaticity of being. *American Psychologist*. 54(7), pp. 462–479.

Barlow, C. & Lohner, H. (1987) Two essays on theory. *Computer Music Journal*. 11(1), pp. 44–60.

Belet, B. & Tenney, J. (1987) An Interview with James Tenney. *Perspectives of New Music*. 25(1), pp. 459–466.

Benson, D. J. (2006) *Music: a mathematical offering*. Cambridge, UK: Cambridge University Press.

Bent, I. D. & Pople, A. (2001) 'Analysis'. Stanley Sadie & John Tyrrell (eds.). The New Grove Dictionary of Music and Musicians

Berger, J. & Gang, D. (1997) A neural network model of metric perception and cognition in the audition of functional tonal music. In *Proceedings of the International Computer Music Association, Thessaloniki, Greece*.

Berger, J. & Gang, D. (1998) A computational model of meter cognition during the audition of functional tonal music: modeling a-priori bias in meter cognition. In *Proceedings of the International Computer Music Association, Michigan*.

Bertalmío, M. *et al.* (2007) Perceptual color correction through variational techniques. *IEEE Transactions on Image Processing*. 16(4), pp. 1058–1072.

Bertalmío, M. & Cowan, J. D. (2009) Implementing the Retinex algorithm with Wilson–Cowan equations. *Journal of Physiology-Paris*. 103(1), pp. 69–72.

Bharucha, J. J. (1987) Music cognition and perceptual facilitation: a connectionist framework. *Music Perception*. 5(1), pp. 1–30.

Bharucha, J. J. & Stoeckig, K. (1986) Reaction time and musical expectancy: priming of chords. *Journal of Experimental Psychology: Human Perception and Performance*. 12(4), pp. 403–410.

Bigand, E. *et al.* (1999) Effect of global structure and temporal organization on chord processing. *Journal of Experimental Psychology: Human Perception and Performance*. 25(1), pp. 184–197.

Bigand, E. *et al.* (2001) The effect of harmonic context on phoneme monitoring in vocal music. *Cognition*. 81(1), pp. B11–B20.

Bigand, E. *et al.* (2003) Sensory versus cognitive components in harmonic priming. *Journal of Experimental Psychology: Human Perception and Performance*. 29(1), pp. 159–171.

Bigand, E., Tillmann, B., *et al.* (2005) Repetition priming: is music special? *The Quarterly Journal of Experimental Psychology Section A*. 58(8), pp. 1347–1375.

Bigand, E., Vieillard, S., *et al.* (2005) Multidimensional scaling of emotional responses to music: the effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion*. 19(8), pp. 1113–1139.

Bigand, E. & Pineau, M. (1997) Global context effects on musical expectancy. *Perception & Psychophysics*. 59(7), pp. 1098–1107.

Bigand, E. & Poulin-Charronnat, B. (2006) Are we 'experienced listeners'? A review of the musical capacities that do not depend on formal musical training. *Cognition*. 100(1), pp. 100–130.

Blacking, J. (1955) Some notes on a theory of African rhythm advanced by Erich von Hornbostel. *African Music*. 1(2), pp. 12–20.

Blacking, J. (1959) Problems of pitch, pattern and harmony in the ocarina music of the Venda. *African Music*. 2(2), pp. 15–23.

Blacking, J. (1965) Music in Uganda. *African Music*. 3(4), pp. 14–17.

Blacking, J. (1967) *Venda children's songs: a study in ethnomusicological analysis*. Johannesburg: Witwatersrand University Press.

Blacking, J. (1969) *Process and product in human society: inaugural lecture delivered 13 September, 1967*. Johannesburg: Witwatersrand University Press.

Blacking, J. (1995) *Music, culture, and experience: selected papers of John Blacking*. University of Chicago Press.

Blackman, D. (1996) Commentary. *Psychology of Music*. 24(2), p. 117.

Borwein, P. *et al.* (2006) *The Riemann hypothesis*. Berlin: Springer Verlag.

Brandel, R. (1959) The African hemiola style. *Ethnomusicology*. 3(3), pp. 106–117.

Brannon, E. M. *et al.* (2008) Electrophysiological measures of time processing in infant and adult brains: Weber's law holds. *Journal of Cognitive Neuroscience*. 20(2), pp. 193–203.

Bray, S. & Tzanetakis, G. (2005) Implicit patching for dataflow-based audio analysis and synthesis. In *Proceedings of the International Computer Music Conference (ICMC)*.

Bregman, A. S. (1994) *Auditory scene analysis: the perceptual organization of sound*. Cambridge, MA: MIT Press.

Bregman, A. S. & Campbell, J. (1971) Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*. 89(2), pp. 244–249.

Bregman, A. S. & Doehring, P. (1984) Fusion of simultaneous tonal glides: the role of parallelness and simple frequency relations. *Perception & Psychophysics*. 36(3), pp. 251–256.

Brillouin, L. (1953) The negentropy principle of information. *Journal of Applied Physics*. 24(9), pp. 1152–1163.

Browne, R. (1981) Tonal implications of the diatonic set. *In Theory Only: Journal of the Michigan Music Theory Society*. 5(6-7), pp. 3–21.

Brown, H. (1988) The interplay of set content and temporal context in a functional theory of tonality perception. *Music Perception*. 5(3), pp. 219–249.

Burford, M. (2006) Hanslick's idealist materialism. *19th-Century Music*. 30(2), pp. 166–181.

Burns, E. M. & Ward, W. D. (1978) Categorical perception–phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. *The Journal of the Acoustical Society of America*. 63(2), pp. 456–468.

Chernoff, J. M. (1979) *African rhythm and African sensibility: aesthetics and social action in African musical idioms*. Chicago, IL: University of Chicago Press.

Cherryholmes, C. H. (1992) Notes on pragmatism and scientific realism. *Educational Researcher*. 21(6), pp. 13–17.

Cheveigné, A. de & Kawahara, H. (2002) Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*. 111(4), pp. 1917–1930.

Chew, E. (2000) *Towards a mathematical model of tonality*. DPhil thesis. Massachusetts Institute of Technology.

Chew, G. (1983) The spice of music: towards a theory of the leading note. *Music Analysis*. 2(1), pp. 35–53.

Chew, G. (1991) Ernst Kurth, music as psychic motion and'Tristan und Isolde': towards a model for analysing musical instability. *Music Analysis*. 10(1), pp. 171–193.

Chowning, J. M. (1980) Computer synthesis of the singing voice. In Johan Sundberg (ed.) *Sound generation in winds, strings, computers*. Stockholm: Royal Swedish Academy of Music. pp. 4–13.

Church, R. M. & Deluty, M. Z. (1977) Bisection of temporal intervals. *Journal of Experimental Psychology: Animal Behavior Processes*. 3(3), pp. 216–228.

Clarkson, I. V. L. (1997) *Approximation of linear forms by lattice points with application to signal processing*. PhD thesis. The Australian National University.

Clayton, M. *et al.* (2005) In time with the music: the concept of entrainment and its significance for ethnomusicology. *European Meetings in Ethnomusicology*. 11, pp. 3–142.

Cohen, D. (2001) 'The imperfect seeks its perfection': harmonic progression, directed motion, and Aristotelian physics. *Music Theory Spectrum*. 23(2), pp. 139–169.

Cohen, M. A. & Grossberg, S. (1983) Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Transactions on Systems, Man and Cybernetics*. SMC-13(5), pp. 815–826.

Collier, J. & Burch, M. (1998) Order from rhythmic entrainment and the origin of levels through dissipation. In *Symmetry: Culture and Science Order/Disorder, Proceedings of the Haifa Congress*.

Collins, J. (1992) *West African Pop roots*. Philadelphia, PA: Temple University Press.

Conlin, J. & Guha, R. (2009) *gnuplot-cpp* [online]. Available from: https://code.google.com/p/gnuplot-cpp/ (Accessed 4 October 2015).

Cook, N. (2001) Explaining harmony. *Annals of the New York Academy of Sciences*. 2(1), pp. 382–385.

Cornelis, O. *et al.* (2010) Access to ethnic music: advances and perspectives in content-based music information retrieval. *Signal Processing*. 90(4), pp. 1008–1031.

Cross, I. (2001) Music, cognition, culture, and evolution. *Annals of the New York Academy of Sciences*. 930, pp. 28–42.

Cross, I. (2003) Music as a biocultural phenomenon. *Annals of the New York Academy of Sciences*. 999(1), pp. 106–111.

Curtis, N. (1920) *Songs and tales from the dark continent: the authoritative 1920 classic, recorded from the singing and the sayings of C. Kamba Simango, Ndau tribe, Portuguese East Africa, and Madikane Čele, Zulu tribe, Natal, Zululand, South Africa*. Mineola, NY: Courier Dover Publications.

Cuthbert, M. S. (2006) Generalized set analysis of sub-Saharan African rhythm? Evaluating and expanding the theories of Willie Anku. *Journal of New Music Research*. 35(3), pp. 211–219.

Dargie, D. (1991) Umngqokolo: Xhosa overtone singing and the song Nondel'ekhaya. *African Music*. 7(1), pp. 33–47.

DEKKMMA (2003) *DEKKMMA: Digitalisatie van het Etnomusicologisch Klankarchief van het Koninklijk Museum voor Midden-Afrika* [online]. Available from: http://music.africamuseum.be/ (Accessed 2 February 2014).

Delaere, M. & Daly, P. H. (1990) Mutations in systems in the natural sciences and music in the first half of the twentieth century. *International Review of the Aesthetics and Sociology of Music*. 21(1), pp. 3–28.

Demidov, D. *et al.* (2012) Programming CUDA and OpenCL: a case study using modern C++ libraries. *arXiv:1212.6326 [physics]*. [online]. Available from: http://arxiv.org/abs/1212.6326 (Accessed 18 October 2014).

Descoeudres, A. (1921) *Le développement de l'enfant de deux à sept ans: recherches de psychologie expérimentale*. Delachaux & Niestlé.

Deutsch, D. (ed.) (1982) *The psychology of music*. New York: Academic Press.

Dewey, J. (1938) *Logic, the theory of inquiry*. H. Holt and Company.

DeWitt, L. A. & Crowder, R. G. (1987) Tonal fusion of consonant musical intervals: the oomph in Stumpf. *Perception & Psychophysics*. 41(1), pp. 73–84.

Drake, C. & Bertrand, D. (2001) The quest for universals in temporal processing in music. *Annals of the New York Academy of Sciences*. 930, pp. 17–27.

Eerola, T. *et al.* (2009) Expectancy in Sami Yoiks revisited: the role of data-driven and schema-driven knowledge in the formation of melodic expectations. *Musicae Scientiae*. 13(2), pp. 231–272.

Ekwueme, L. E. N. (1974) Concepts of African musical theory. *Journal of Black Studies*. 5(1), pp. 35–64.

Ekwueme, L. E. N. (1975a) Structural levels of rhythm and form in African music: with particular reference to the West Coast. *African Music*. 5(4), pp. 27–35.

Ekwueme, L. E. N. (1975b) Guest editorial: African musicological investigation in a culture conscious era. *African Music*. 5(4), pp. 4–5.

Ekwueme, L. E. N. (1980) Analysis and analytic techniques in African music: a theory of melodic scales. *African Music*. 6(1), pp. 89–106.

Eliade, M. (1959) *The sacred and the profane: the nature of religion*. Vol. 77. Boston, MA: Houghton Mifflin Harcourt.

England, N. M. (1967) Bushman counterpoint. *Journal of the International Folk Music Council*. 19(1967), pp. 58–66.

Farbood, M. M. (2006) *A quantitative, parametric model of musical tension*. PhD thesis. Massachusetts Institute of Technology.

Fausett, L. V. (1994) *Fundamentals of neural networks: architectures, algorithms, and applications*. Englewood Cliffs, NJ: Prentice-Hall.

Ferris, J. (1959) The evolution of Rameau's 'harmonic theories'. *Journal of Music Theory*. 3(2), pp. 231–256.

Ferris, T. (1978) Voyager's music. In *Murmurs of Earth: The Voyager Interstellar Record*. New York, NY: Random House.

Firmino, É. A. *et al.* (2009) Travelling through pitch space speeds up musical time. *Music Perception*. 26(3), pp. 205–210.

Fischer, R. F. H. (2005) *Precoding and signal shaping for digital transmission*. New York, NY: John Wiley & Sons.

Fitch, W. T. (2006) The biology and evolution of music: a comparative perspective. *Cognition*. 100(1), pp. 173–215.

Fitts, P. M. (1954) The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*. 47(6), pp. 381–391.

FitzHugh, R. (1961) Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*. 1(6), pp. 445–466.

Gedik, A. C. & Bozkurt, B. (2010) Pitch-frequency histogram-based music information retrieval for Turkish music. *Signal Processing*. 90(4), pp. 1049–1063.

Gherman, S. (2008) Harmony and its functionality: a gloss on the substantial similarity test in music copyrights. *Fordham Intellectual Property, Media and Entertainment Law Journal*. 19(2), pp. 482–517.

Gibbon, J. (1981) On the form and location of the psychometric bisection function for time. *Journal of Mathematical Psychology*. 24, pp. 58–87.

Gibbon, J. & Church, R. M. (1981) Time left: linear versus logarithmic subjective time. *Journal of Experimental Psychology: Animal Behavior Processes*. 7(2), pp. 87–108.

Gilmore, B. (1995) Changing the Metaphor: Ratio Models of Musical Pitch in the Work of Harry Partch, Ben Johnston, and James Tenney. *Perspectives of New Music*. 33(1), pp. 458–503.

Gjerdingen, R. O. (1994) Apparent motion in music? *Music Perception*. 11(4), pp. 335–370.

Glasberg, B. R. & Moore, B. C. (1990) Derivation of auditory filter shapes from notched-noise data. *Hearing Research*. 47(1-2), pp. 103–138.

Goldman, R. F. (1968) *Harmony in Western music*. London: Barrie & Rockliff.

Gómez, E. & Herrera, P. (2008) *Comparative analysis of music recordings from Western and non-Western traditions by automatic tonal feature extraction*. 3(3), pp. 140–156.

Gómez, F. *et al.* (2007) Mathematical models for binarization and ternarization of musical rhythms. In *BRIDGES: Mathematical Connections in Art, Music and Science*. pp. 99–108.

Green, D. M. & Swets, J. A. (1966) *Signal detection theory and psychophysics*. New York, NY: John Wiley and Sons.

Greenwood, D. D. (1961) Critical bandwidth and the frequency coordinates of the basilar membrane. *The Journal of the Acoustical Society of America*. 33(10), pp. 1344–1356.

Griffiths, T. D. (2001) The neural processing of complex sounds. *Annals of the New York Academy of Sciences*. 930(191), pp. 133–142.

Guennebaud, G. *et al.* (2010) *Eigen v3* [online]. Available from: http://eigen.tuxfamily.org.

Gurney, E. (1880) *The power of sound*. Smith, Elder.

Hansen, D. (1981) *The music of the Xhosa-speaking people*. PhD thesis. Johannesburg: University of the Witwatersrand.

Hanslick, E. (1974) *The beautiful in music: a contribution to the revisal of musical aesthetics*. New York: Da Capo Press.

Harley, R. (1991) The role of critical listening in evaluating audio equipment quality. In *Audio Engineering Society Convention 91*. Audio Engineering Society.

Harley, R. (2004) *The complete guide to high-end audio*. 3rd ed. Tijeras, NM: Acapella Publishing.

Hauser, M. D. & McDermott, J. (2003) The evolution of the music faculty: a comparative perspective. *Nature Neuroscience*. 6(7), pp. 663–668.

Hebb, D. O. (2002) *The organization of behavior: a neuropsychological theory*. Taylor & Francis.

Helmholtz, H. von (1954) *On the sensations of tone as a physiological basis for the theory of music*. Mineola, NY: Courier Dover Publications.

Herrera, P. *et al.* (2009) The discipline formerly known as MIR. In *International Society for Music Information Retrieval (ISMIR) Conference, special session on The Future of MIR (fMIR)*. Kobe, Japan: . pp. 2–4.

Herskovits, M. J. (1972) Cultural relativism and cultural values. In Frances Herskovits (ed.) *Cultural Relativism: Perspectives in Cultural Pluralism*. New York: Random House.

Hindemith, P. (1961) *A composer's world: horizons and limitations*. Garden City: Doubleday.

Hodgkin, A. L. & Huxley, A. F. (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*. 117(4), p. 500.

Hood, M. (1971) *The ethnomusicologist, Volume 27*. New York, NY: McGraw-Hill Book Co.

Hopfield, J. J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*. 79(8), pp. 2554–2558.

Hopfield, J. J. (1984) Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the national academy of sciences*. 81(10), pp. 3088–3092.

Hoppensteadt, F. C. & Izhikevich, E. M. (1996) Synaptic organizations and dynamical properties of weakly connected neural oscillators: II. learning phase information. *Biological Cybernetics*. 75(2), pp. 129–135.

Hoppensteadt, F. C. & Izhikevich, E. M. (1997) *Weakly connected neural networks*. New York, NY: Springer-Verlag.

Hornbostel, E. M. von (1928) African Negro music. *Africa: Journal of the International African Institute*. 1(1), pp. 30–62.

Howse IV, J. W. (1995) *Gradient and Hamiltonian dynamics: some applications to neural network analysis and system identification*. PhD thesis. Albuquerque, NM: The University of New Mexico.

Hsu, D. (1966) Ernst Kurth and his concept of music as motion. *Journal of Music Theory*. 10(1), pp. 2–17.

Huron, D. (1996) Commentary. *Psychology of Music*. 24(2), p. 120.

Huron, D. (2001) Tone and voice: a derivation of the rules of voice-leading from perceptual principles. *Music Perception*. 19(1), pp. 1–64.

Hurwitz, R. I. (1970) *An investigation into the perception of root in harmonic intervals*. PhD thesis. Indiana University.

Hutchins, S. & Palmer, C. (2008) Repetition priming in music. *Journal of Experimental Psychology: Human Perception and Performance*. 34(3), pp. 693–707.

ISMIR (n.d.) *The International Society for Music Information Retrieval: conferences, publications and related activities* [online]. Available from: http://www.ismir.net/ (Accessed 2 February 2014).

Iyer, V. S. (1998) *Microstructures of feel, macrostructures of sound: embodied cognition in West African and African-American musics*. PhD thesis. University of California, Berkeley.

Jazayeri, M. & Shadlen, M. N. (2010) Temporal context calibrates interval timing. *Nature Neuroscience*. 13(8), pp. 1020–1026.

Jeltsema, D. & Scherpen, J. M. A. (2009) Multidomain modeling of nonlinear networks and systems: energy- and power-based perspectives. *IEEE Control Systems Magazine*. pp. 28–59.

Johnson, R. B. & Onwuegbuzie, A. J. (2004) Mixed methods research: a research paradigm whose time has come. *Educational Researcher*. 33(7), pp. 14–26.

Jonaitis, E. M. & Saffran, J. R. (2009) Learning harmony: the role of serial statistics. *Cognitive Science*. 33(5), pp. 951–968.

Jones, A. M. (1934) African drumming: a study in the combination of rhythms in African music. *Bantu Studies*. 8(1), pp. 1–16.

Jones, A. M. (1937) The study of African musical rhythm. *Bantu Studies*. 11(1), pp. 295–319.

Jones, A. M. (1949) *African music in Northern Rhodesia and some other places*. Livingstone Northern Rhodesia: Rhodes-Livingstone Museum.

Jones, A. M. (1954) African rhythm. *Africa: Journal of the International African Institute*. 24(1), pp. 26–47.

Jones, A. M. (1959a) *Studies in African music, Volume I*. London: Oxford University Press.

Jones, A. M. (1959b) *Studies in African music, Volume II*. London: Oxford University Press.

Jones, A. M. & Kombe, L. (1952) *The Icila dance, old style: a study in African music and dance of the Lala tribe of Northern Rhodesia*. Longmans, Green for African Music Society.

Jones, M. R. (1981) Music as a stimulus for psychological motion: Part I. Some determinants of expectancies. *Psychomusicology: A Journal of Research in Music Cognition*. 1(2), pp. 34–51.

Jones, M. R. (1982) Music as a stimulus for psychological motion: Part II. An expectancy model. *Psychomusicology: A Journal of Research in Music Cognition*. 2(1), pp. 1–13.

Jones, M. R. (1987) Dynamic pattern structure in music: recent theory and research. *Perception & Psychophysics*. 41(6), pp. 621–634.

Jones, M. R. & Boltz, M. (1989) Dynamic attending and responses to time. *Psychological Review*. 96(3), pp. 459–491.

Jordan, D. & Smith, P. (2007) *Nonlinear ordinary differential equations:an introduction for scientists and engineers: an introduction for scientists and engineers*. Oxford, UK: Oxford University Press.

Joseph, R. (1983) Zulu women's music. *African Music*. 6(3), pp. 53–89.

Jung, C. G. *et al.* (1964) *Man and his symbols*. Carl Gustav Jung & M. -L. von Franz (eds.). New York: Anchor Press/Doubleday.

Justus, T. C. & Bharucha, J. J. (2001) Modularity in musical processing: the automaticity of harmonic priming. *Journal of Experimental Psychology: Human Perception and Performance*. 27(4), pp. 1000–1011.

Kameoka, A. & Kuriyagawa, M. (1969) Consonance theory Part II: consonance of complex tones and its calculation method. *The Journal of the Acoustical Society of America*. 45(6), pp. 1460–1469.

Kauffman, R. (1980) African rhythm: a reassessment. *Ethnomusicology*. 24(3), pp. 393–415.

Kehinde, O. (2011) Dissonance and chromatic inflections in traditional African music. *British Journal of Arts and Social Sciences*. 3(1), pp. 98–109.

King, A. (1960) Employments of the 'standard pattern' in Yoruba music. *African Music*. 2(3), pp. 51–54.

Kirby, P. R. (1926) Some problems of primitive harmony and polyphony, with special reference to Bantu practice. *South African Journal of Science*. 23, pp. 951–970.

Kirby, P. R. (1932) Musical origins in the light of the musical practices of Bushman, Hottentot and Bantu. In *Proceedings of the Musical Association, 59th Session*. pp. 23–33.

Kirby, P. R. (1933) The reed-flute ensembles of South Africa: a study in South African native music. *The Journal of the Royal Anthropological Institute of Great Britain and Ireland*. 63, pp. 313–388.

Kirby, P. R. (1965) *The musical instruments of the native races of South Africa*. 2nd edition. Johannesburg: Witwatersrand University Press.

Klapuri, A. & Davy, M. (2007) *Signal processing methods for music transcription*. New York, NY: Springer.

Kline, M. (1964) *Mathematics in Western culture*. New York: Oxford University Press.

Koetting, J. (1970) Analysis and notation of West African drum ensemble music. *Selected Reports in Ethnomusicology*. 1(3), pp. 115–146.

Koetting, J. & Knight, R. (1986) What do we know about African rhythm? *Ethnomusicology*. 30(1), pp. 58–63.

Kolinski, M. (1973) A cross-cultural approach to metro-rhythmic patterns. *Ethnomusicology*. 17(3), pp. 494–506.

Konečni, V. J. (2008) Does music induce emotion? A theoretical and methodological analysis. *Psychology of Aesthetics, Creativity, and the Arts*. 2(2), pp. 115–129.

Körte, A. (1915) Kinematoskopische untersuchungen. *Zeitschrift für Psychologie*. 72, pp. 193–296.

Krumhansl, C. L. *et al.* (2000) Cross-cultural music cognition: cognitive methodology applied to North Sami Yoiks. *Cognition*. 76(1), pp. 13–58.

Krumhansl, C. L. & Keil, F. C. (1982) Acquisition of the hierarchy of tonal functions in music. *Memory & Cognition*. 10(3), pp. 243–251.

Krumhansl, C. L. & Kessler, E. J. (1982) Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological review*. 89(4), pp. 334–368.

Krumhansl, C. L. & Shepard, R. N. (1979) Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance*. 5(4), pp. 579–594.

Kubik, G. (1965) Transcription of Mangwilo xylophone music from film strips. *African Music*. 3(4), pp. 35–51.

Kubik, G. (1972) Transcription of African music from silent film: theory and methods. *African Music*. 5(2), pp. 28–39.

Kubik, G. (1985) African tone-systems: a reassessment. *Yearbook for Traditional Music*. 17(1985), pp. 31–63.

Kubik, G. (1999) *Africa and the blues*. Jackson, MS: University Press of Mississippi.

Kubik, G. (2010a) *Theory of African music, Volume I*. Chicago, IL: University of Chicago Press.

Kubik, G. (2010b) *Theory of African music, Volume II*. Chicago, IL: University of Chicago Press.

Kunej, D. & Turk, I. (2001) New perspectives on the beginnings of music: archeological and musicological analysis of a Middle Paleolithic bone ‘flute’. In Nils Lennart Wallin *et al.* (eds.) *The Origins of Music*. A Bradford Book. Cambridge, MA: MIT Press. pp. 235–268.

Kuroe, Y. *et al.* (2002) On energy function for complex-valued neural networks and its applications. In *Proceedings of the 9th International Conference on Neural Information Processing*. pp. 1079–1083.

Kurth, E. (1922) *Grundlagen des linearen Kontrapunkts: Bachs melodische Polyphonie*. 3. Aufl. Berlin: Hesse.

Langner, G. (1997) Temporal processing of pitch in the auditory system. *Journal of New Music Research*. 26, pp. 116–132.

Langner, G. *et al.* (1997) Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography. *Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology*. 181(6), pp. 665–676.

Langner, G. (2005) Neuronal mechanisms underlying the perception of pitch and harmony. *Annals of the New York Academy of Sciences*. 1060, pp. 50–52.

Large, E. W. (2006) A generic nonlinear model for auditory perception Alfred L. Nuttall *et al.* (eds.). *Auditory Mechanisms: Processes and Models*. pp. 516–517.

Large, E. W. (2008) Resonating to music rhythm: theory and experiment. In Simon Grondin (ed.) *The Psychology of Time*. Emerald Group Publishing. pp. 189–231.

Large, E. W. (2010a) A dynamical systems approach to musical tonality. In R Huys & V. K. Jirsa (eds.) *Nonlinear Dynamics in Human Behavior*. Berlin: Springer-Verlag. pp. 193–211.

Large, E. W. (2010b) Neurodynamics of music. In Mari Riess Jones *et al.* (eds.) *Music Perception*. Springer Handbook of Auditory Research. Springer New York. pp. 201–231.

Large, E. W. *et al.* (2010) A canonical model for gradient frequency neural networks. *Physica D: Nonlinear Phenomena*. 239(12), pp. 905–911.

Large, E. W. (2011) Musical tonality, neural resonance and Hebbian learning. In C Agon *et al.* (eds.) *Mathematics and Computation in Music*. Berlin: Springer-Verlag. pp. 115–125.

Large, E. W. *et al.* (2014) *GrFNN toolbox GitHub repository* [online]. Available from: https://github.com/MusicDynamicsLab/GrFNNToolbox (Accessed 20 August 2015).

Large, E. W. & Kolen, J. F. (1994) Resonance and the perception of musical meter. *Connection Science*. 6(2&3), pp. 177–208.

Large, E. W. & Palmer, C. (2002) Perceiving temporal regularity in music. *Cognitive Science*. 26(1), pp. 1–37.

Large, E. W. & Snyder, J. S. (2009) Pulse and meter as neural resonance. *Annals of the New York Academy of Sciences*. 1169, pp. 46–57.

Large, E. W. & Tretakis, A. E. (2005) Tonality and nonlinear resonance. *Annals of the New York Academy of Sciences*. 1060, pp. 53–56.

Lartillot, O. *et al.* (2008) Commentary on 'Comparative analysis of music recordings from Western and non-Western traditions by automatic tonal feature extraction' by Emilia Gómez, and Perfecto Herrera. *Empirical Musicology Review*. 3(3), pp. 157–160.

Leman, M. (2000) An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception*. 17(4), pp. 481–509.

Leman, M. (2008) *Embodied music: cognition and mediation technology*. MIT Press.

Lerdahl, F. (2001) *Tonal pitch space*. New York: Oxford University Press.

Lerdahl, F. A. & Jackendoff, R. S. (1983) *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Lerdahl, F. & Krumhansl, C. L. (2007) Modeling Tonal Tension. *Music Perception*. 24(4), pp. 329–366.

Lerud, K. D. *et al.* (2014) Mode-locking neurodynamics predict human auditory brainstem responses to musical intervals. *Hearing Research*. 308, pp. 41–49.

Lewis, C. I. (1970) Logical positivism and pragmatism. In *Collected Papers*. Stanford University Press. pp. 92–112.

Lidy, T. *et al.* (2010) On the suitability of state-of-the-art music information retrieval methods for analyzing, categorizing and accessing non-Western and ethnic music collections. *Signal Processing*. 90(4), pp. 1032–1048.

Locke, D. (1982) Principles of offbeat timing and cross-rhythm in Southern Eve̱ dance drumming. *Ethnomusicology*. 26(2), pp. 217–246.

Locke, D. (2011) The metric matrix: simultaneous multidimensionality in African music. *Analytical Approaches to World Music*. 1(1), pp. 48–72.

London, J. (2002) Some non-isomorphisms between pitch and time. *Journal of Music Theory*. 46(1), pp. 127–151.

Lorca, F. G. (1998) Theory and play of the Duende. In Christopher Maurer (ed.) *In Search of Duende*. English an New York, NY: New Directions Publishing. pp. 48–62.

Lucia, C. (ed.) (1986) *Proceedings of the first national music educators' conference*. Durban: University of Natal.

Lyon, R. F. *et al.* (2010) History and future of auditory filter models. In *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*. pp. 3809–3812.

Malmberg, C. (1918) The perception of consonance and dissonance. *Psychological Monographs*. pp. 93–133.

Margulis, E. H. (2005) A model of melodic expectation. *Music Perception*. 22(4), pp. 663–714.

Marmel, F. *et al.* (2010) Priming in melody perception: tracking down the strength of cognitive expectations. *Journal of Experimental Psychology: Human Perception and Performance*. 36(4), pp. 1016–1028.

Marmel, F. *et al.* (2011) Harmonic relationships influence auditory brainstem encoding of chords. *Neuroreport*. 22(10), pp. 504–508.

Mazzola, G. (2002) *The topos of music: geometric logic of concepts, theory, and performance*. Basel: Birkhauser Verlag.

McAdams, S. (1982) Spectral fusion and the creation of auditory images. In Manfred Clynes (ed.) *Music, Mind, and Brain: The Neuropsychology of Music*. New York: Plenum Press. pp. 279–298.

McAdams, S. (1989) Segregation of concurrent sounds. I: effects of frequency modulation coherence. *The Journal of the Acoustical Society of America*. 86(6), pp. 2148–2159.

McAdams, S. *et al.* (1995) Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychological Research*. 58(3), pp. 177–192.

McAdams, S. & Bregman, A. (1979) Hearing musical streams. *Computer Music Journal*. 3(4), pp. 26–43.

McCulloch, W. S. & Pitts, W. H. (1943) A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*. 5(4), pp. 115–133.

McLean, B. (1981) Symbolic extension and its corruption of music. *Perspectives of New Music*. 20(1), pp. 331–356.

Mendes, R. V. (1999) Dynamics of networks and applications. In John W. Clark *et al.* (eds.) *Scientific Applications of Neural Nets*. Lecture Notes in Physics. Springer Berlin Heidelberg. pp. 257–288.

Merriam, A. P. (1959) Characteristics of African music. *Journal of the International Folk Music Council*. 11, pp. 13–19.

Merriam, A. P. (1962) The African idiom in music. *The Journal of American Folklore*. 75(296), pp. 120–130.

Merriam, A. P. (1964) *The anthropology of music*. Evanston, IL: Northwestern University Press.

Meyer, L. B. (1956) *Emotion and meaning in music*. Chicago, IL: University of Chicago Press.

Miller, G. A. & Heise, G. A. (1950) The trill threshold. *The Journal of the Acoustical Society of America*. 22(5), pp. 637–638.

Minsky, M. L. (1988) Expanded ed. *Perceptrons: an introduction to computational geometry*. Cambridge, MA: MIT Press.

MIREX (2011) *2010:MIREX2010 results - MIREX wiki* [online]. Available from: http://www.music-ir.org/mirex/wiki/2010:MIREX2010_Results (Accessed 2 February 2014).

Moelants, D. *et al.* (2007) The problems and opportunities of content-based analysis and description of ethnic music. *International Journal of Intangible Heritage*. 2, pp. 57–67.

Moore, B. C. & Glasberg, B. R. (1983) Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *The Journal of the Acoustical Society of America*. 74(3), pp. 750–753.

Mudimbe, V. Y. (1988) *The invention of Africa: gnosis, philosophy, and the order of knowledge*. Bloomington, IN: Indiana University Press.

Nagumo, J. *et al.* (1962) An active pulse transmission line simulating nerve axon. *Proceedings of the IRE*. 50(10), pp. 2061–2070.

Narmour, E. (1990) *The analysis and cognition of basic melodic structures: the implication-realization model*. Chicago, IL: University of Chicago Press.

Neisser, U. (1967) *Cognitive psychology*. New York, NY: Appleton-Century-Crofts.

Nettl, B. (1964) *Theory and method in ethnomusicology*. New York, NY: Free Press.

Nettl, B. (2001) An Ethnomusicologist Contemplates Universals in Musical Sound and Musical Culture. In Nils Lennart Wallin *et al.* (eds.) *The Origins of Music*. A Bradford Book. Cambridge, MA: MIT Press. pp. 463–472.

Newman, E. (1919) *A musical motley*. London, UK: John Lane.

Nketia, J. H. K. (1962) The hocket-technique in African music. *Journal of the International Folk Music Council*. 14(1962), pp. 44–52.

Nketia, J. H. K. (1963) *African music in Ghana*. Evanston, IL: Northwestern University Press.

Nketia, J. H. K. (1974) *The music of Africa*. New York, NY: W. W. Norton.

van Noorden, L. P. A. S. (1975) *Temporal coherence in the perception of tone sequences*. PhD thesis. Eindhoven, Netherlands: Technische Hogeschool Eindhoven.

Nzewi, M. (1997) *African music: theoretical content and creative continuum: the culture-exponent's definitions*. Oldershausen: Institut für Didaktik populärer Musik.

Oliver, P. (1970) *Savannah syncopators: African retentions in the blues*. Studio Vista.

OpenMP ARB (n.d.) *OpenMP.org* [online]. Available from: http://openmp.org (Accessed 6 November 2014).

Ören, T. & Yilmaz, L. (2013) Philosophical aspects of modeling and simulation. In Andreas Tolk (ed.) *Ontology, Epistemology, and Teleology for Modeling and Simulation*. Intelligent Systems Reference Library. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 157–172.

Pantaleoni, H. (1972) Three principles of timing in anlo dance drumming. *African Music*. 5(2), pp. 50–63.

Parncutt, R. (1989) *Harmony: a psychoacoustical approach*. Berlin: Springer-Verlag.

Parncutt, R. (1994) A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*. 11(4), pp. 409–464.

Parncutt, R. & Strasburger, H. (1994) Applying psychoacoustics in composition: 'harmonic' progressions of 'nonharmonic' sonorities. *Perspectives of New Music*. 32(2), pp. 88–129.

Patel, A. D. *et al.* (1998) Processing syntactic relations in language and music: an event-related potential study. *Journal of Cognitive Neuroscience*. 10(6), pp. 717–733.

Patel, A. D. & Daniele, J. R. (2003) An empirical comparison of rhythm in language and music. *Cognition*. 87, pp. B35–B45.

Peretz, I. & Kolinsky, R. (1993) Boundaries of separability between melody and rhythm in music discrimination: a neuropsychological perspective. *The Quarterly Journal of Experimental Psychology Section A*. 46(2), pp. 301–325.

Pike, K. L. (1967) Emic and etic standpoints for the description of behaviour. In *Language in Relation to a Unified Theory of the Structure of Human Behavior*. Mouton.

Pikovsky, A. (2003) *Synchronization: a universal concept in nonlinear sciences*. Cambridge, UK: Cambridge University Press.

Pikrakis, A. *et al.* (2004) Music meter and tempo tracking from raw polyphonic audio. In *Proceedings of the International Conference on Music Information Retrieval*. pp. 192–197.

Pinker, S. (1999) *How the mind works*. 1st ed. London: Penguin.

Plomp, R. (1965) Detectability threshold for combination tones. *The Journal of the Acoustical Society of America*. 37, pp. 1110–1123.

Plomp, R. (1976) *Aspects of tone sensation: a psychophysical study*. Waltham, MA: Academic Press.

Plomp, R. & Levelt, W. J. (1965) Tonal consonance and critical bandwidth. *The Journal of the Acoustical Society of America*. 38(4), pp. 548–560.

Popper, K. R. (2002) *The logic of scientific discovery*. London, UK: Routledge.

Poulin-Charronnat, B. *et al.* (2005) Musical structure modulates semantic priming in vocal music. *Cognition*. 94(3), pp. B67–B78.

Pressing, J. (1983) Cognitive isomorphisms between pitch and rhythm in world musics: West Africa, the Balkans and Western tonality. *Studies in Music*. pp. 38–61.

Pressing, J. *et al.* (1996) Cognitive multiplicity in polyrhythmic pattern performance. *Journal of Experimental Psychology: Human Perception and Performance*. 22(5), pp. 1127–1148.

Provasi, J. *et al.* (2011) Temporal bisection in 4-month-old infants. *Journal of Experimental Psychology: Animal Behavior Processes*. 37(1), pp. 108–113.

Quine, W. V. (1951) Main trends in recent philosophy: two dogmas of empiricism. *The Philosophical Review*. 60(1), pp. 20–43.

Rameau, J. P. (1971) *Treatise on harmony*. Mineola, NY: Dover.

Riemann, H. (1896) *Harmony simplified: or, the theory of the tonal functions of chords*. London: Augener.

Roeder, J. & Schoenberg, A. (1989) Harmonic implications of Schoenberg's 'Observations of Atonal Voice Leading'. *Journal of Music Theory*. 33(1), pp. 27–62.

Rosenblatt, F. (1962) *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. Wachington, DC: Spartan Books.

Rossing, T. D. (1990) *The science of sound*. 2nd ed. Reading, MA: Addison-Wesley.

Rouget, G. (1961) Un chromatisme africain. *L'Homme*. 1(3), pp. 32–46.

Rougier, N. P. & Fix, J. (2012) DANA: distributed numerical and adaptive modelling framework. *Network: Computation in Neural Systems*. 23(4), pp. 237–253.

Rycroft, D. K. (1954) Tribal style and free expression. *African Music*. 1(1), pp. 16–27.

Rycroft, D. K. (1959) African music in Johannesburg: African and non-African features. *Journal of the International Folk Music Council*. 11, pp. 25–30.

Rycroft, D. K. (1961) The guitar improvisations of Mwenda Jean Bosco. *African Music*. 2(4), pp. 81–98.

Rycroft, D. K. (1962a) The guitar improvisations of Mwenda Jean Bosco (Part II). *African Music*. 3(1), pp. 86–102.

Rycroft, D. K. (1962b) Zulu and Xhosa praise-poetry and song. *African Music*. 3(1), pp. 79–85.

Rycroft, D. K. (1966) Friction chordophones in south-eastern Africa. *The Galpin Society Journal*. 19, pp. 84–100.

Rycroft, D. K. (1967) Nguni vocal polyphony. *Journal of the International Folk Music Council*. 19(1967), pp. 88–103.

Rycroft, D. K. (1971) Stylistic evidence in Nguni song. In Klaus Wachsmann (ed.) *Essays on Music and History in Africa*. Evanston, IL: Northwestern University Press. pp. 213–241.

Rycroft, D. K. (1975a) The Zulu bow songs of Princess Magogo. *African Music*. 5(4), pp. 41–97.

Rycroft, D. K. (1975b) A royal account of music in Zulu life with translation, annotation, and musical transcription. *Bulletin of the School of Oriental and African Studies, University of London*. 38(2), pp. 351–402.

Rycroft, D. K. (1991) Black South African urban music since the 1890's: some reminiscences of Alfred Assegai Kumalo (1879-1966). *African Music*. 7(1), pp. 5–32.

Sachs, C. (1943) *The rise of music in the ancient world, east and west*. W.W. Norton & Co.

Scase, M. M. *et al.* (2006) Time-dependent plumes and jets with decreasing source strengths. *Journal of Fluid Mechanics*. 563, pp. 443–461.

Scavone, G. P. (n.d.) *The RtAudio home page* [online]. Available from: http://www.music.mcgill.ca/~gary/rtaudio/ (Accessed 6 November 2014).

Schellenberg, E. (1997) Simplifying the implication-realization model of melodic expectancy. *Music Perception*. 14(3), pp. 295–318.

Schellenberg, E. G. *et al.* (2005) Children's implicit knowledge of harmony in Western music. *Developmental Science*. 8(6), pp. 551–566.

Schenker, H. (1954) *Harmony*. Chicago: University of Chicago Press.

Schenker, H. & American Musicological Society (1979) *Free composition*. New York: Longman published in cooperation with the American Musicological Society.

Schlaug, G. (2001) The brain of musicians: a model for functional and structural adaptation. *Annals of the New York Academy of Sciences*. 930(1), pp. 281–299.

Schmuckler, M. A. & Boltz, M. G. (1994) Harmonic and rhythmic influences on musical expectancy. *Perception & Psychophysics*. 56(3), pp. 313–325.

Schoenberg, A. (1969) *Structural functions of harmony*. Revised with corrections. London: Faber & Faber.

Schoenberg, A. (1983) *Theory of harmony*. London: Faber & Faber.

Schouten, J. *et al.* (1962) Pitch of the residue. *The Journal of the Acoustical Society of America*. 34(8.2), pp. 1418–1424.

Schrödinger, E. (1948) *What is life?: the physical aspect of the living cell*. Cambridge, UK: Cambridge University Press.

Schwartz, D. A. *et al.* (2003) The statistical structure of human speech sounds predicts musical universals. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*. 23(18), pp. 7160–7168.

Sethares, W. (1993) Local consonance and the relationship between timbre and scale. *The Journal of the Acoustical Society of America*. 94(3), pp. 1218–1228.

Sethares, W. (1998) Consonance-based spectral mappings. *Computer Music Journal*. 22(1), pp. 56–72.

Sethares, W. A. (2005) *Tuning, timbre, spectrum, scale*. London, UK: Springer-Verlag.

Seung, H. S. *et al.* (1998) Minimax and Hamiltonian dynamics of excitatory-inhibitory networks. In *Advances in Neural Information Processing Systems 10*. MIT Press. pp. 329–335.

Shatz, C. (1992) The developing brain. *Scientific American*. 267(3), pp. 60–67.

Shepard, R. N. (1982) Geometrical approximations to the structure of musical pitch. *Psychological Review*. 89(4), pp. 305–333.

Sloboda, J. A. (1985) *The musical mind: The cognitive psychology of music*. Oxford, UK: Oxford University Press.

Smith, L. R. (1967) *An investigation of the inherent qualities of musical intervals as definers of melodic tonality*. Ann Arbor, MI: University Microfilms.

Sorrell, N. (1996) Commentary. *Psychology of Music*. 24(2), p. 124.

Sperber, D. (1996) *Explaining culture: a naturalistic approach*. Oxford: Blackwell.

Stevens, C. (2004) Cross-cultural studies of musical pitch and time. *Acoustical Science and Technology*. 25(6), pp. 433–438.

Stevens, C. & Wiles, J. (1994) Tonal music as a componential code: learning temporal relationships between and within pitch and timing components. In J. D. Cowan *et al.* (eds.) *Advances in Neural Information Processing Systems*. pp. 1085–1092.

Stock, J. (1993) The application of Schenkerian analysis to ethnomusicology: problems and possibilities. *Music Analysis*. 12(2), pp. 215–240.

Stone, R. M. (1985) In search of time in African music. *Music Theory Spectrum*. 7(Spring), pp. 139–148.

Strogatz, S. H. (1994) *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Boulder, CO: Westview Press.

Susuki, Y. *et al.* (2007) Energy-based analysis of frequency entrainment described by van der Pol and phase-locked loop equations. *CHAOS: An Interdisciplinary Journal of Nonlinear Science*. 17(2), pp. 0231081–0231088.

Tekman, H. G. & Bharucha, J. J. (1992) Time course of chord priming. *Perception & Psychophysics*. 51(1), pp. 33–39.

Tekman, H. G. & Bharucha, J. J. (1998) Implicit knowledge versus psychoacoustic similarity in priming of chords. *Journal of Experimental Psychology: Human Perception and Performance*. 24(1), pp. 252–260.

Temperley, D. (2000) Meter and grouping in African music: a view from music theory. *Ethnomusicology*. 44(1), pp. 65–96.

Terhardt, E. *et al.* (1982a) Pitch of complex signals according to virtual pitch theory: tests, examples, and predictions. *The Journal of the Acoustical Society of America*. 71(3), pp. 671–678.

Terhardt, E. *et al.* (1982b) Algorithm for extraction of pitch and pitch salience from complex tonal signals. *The Journal of the Acoustical Society of America*. 71(3), pp. 679–688.

Terhardt, E. (1984) The concept of musical consonance: a link between music and psychoacoustics. *Music Perception*. 1(3), pp. 276–295.

Thompson, R. F. (1984) *Flash of the spirit: African and Afro-American art and philosophy*. New York, NY: Vintage Books.

Thomson, W. (1993) The harmonic root: a fragile marriage of concept and percept. *Music Perception*. 10(4), pp. 385–415.

Tillmann, B. *et al.* (1998) Effects of global and local contexts on harmonic expectancy. *Music Perception*. 16(1), pp. 99–117.

Tillmann, B. *et al.* (2000) Implicit learning of tonality: a self-organizing approach. *Psychological Review*. 107(4), pp. 885–913.

Tillmann, B. *et al.* (2006) The influence of musical relatedness on timbre discrimination. *European Journal of Cognitive Psychology*. 18(3), pp. 343–358.

Tillmann, B. *et al.* (2008) Tonal centers and expectancy: facilitation or inhibition of chords at the top of the harmonic hierarchy? *Journal of Experimental Psychology: Human Perception and Performance*. 34(4), pp. 1031–1043.

Tillmann, B. & Bharucha, J. J. (2002) Effect of harmonic relatedness on the detection of temporal asynchronies. *Perception & Psychophysics*. 64(4), pp. 640–649.

Tillmann, B. & Bigand, E. (2001) Global context effect in normal and scrambled musical sequences. *Journal of Experimental Psychology: Human Perception and Performance*. 27(5), pp. 1185–1196.

Tonietti, T. M. (2002) Is music relevant for the history of science? In P. Cerrai *et al.* (eds.) *The Application of Mathematics to the Sciences of Nature: Critical Moments and Aspects*. Springer. pp. 281–291.

Toussaint, G. (2003) Classification and phylogenetic analysis of african ternary rhythm timelines. In *Proceedings of BRIDGES: Mathematical Connections in Art, Music and Science*. pp. 1–18.

Toussaint, G. (2005) Mathematical features for recognizing preference in sub-Saharan African traditional rhythm timelines. In *Pattern Recognition and Data Mining*. Berlin, Heidelberg: Springer. pp. 18–27.

Toussaint, G. (2010) Computational geometric aspects of rhythm, melody, and voice-leading. *Computational Geometry*. 43(1), pp. 2–22.

Tovey, D. F. (1935) *Essays in musical analysis, Volume III: Concertos*. London: Oxford University Press.

Tracey, A. (1970) The Matepe mbira music of Rhodesia. *African Music*. 4(4), pp. 37–61.

Tracey, A. (1972) The original African mbira? *African Music*. 5(2), pp. 85–104.

Tramo, M. J. *et al.* (2001) Neurobiological foundations for the theory of harmony in Western tonal music. *Annals of the New York Academy of Sciences*. 930, pp. 92–116.

Tzanetakis, G. *et al.* (2007) Computational ethnomusicology. *Journal of Interdisciplinary Music Studies*. 1(2), pp. 1–24.

Tzanetakis, G. (n.d.) *MARSYAS* [online]. Available from: http://marsyas.info/ (Accessed 2 February 2014).

Tzanetakis, G. & Cook, P. (2000) MARSYAS: a framework for audio analysis. *Organised Sound*. 4(3), pp. 169–175.

Veenstra, A. J. F. (1958) The Begu Zulu vertical flute. *African Music*. 2(1), pp. 40–45.

Velasco, M. J. & Large, E. W. (2011) Pulse detection in syncopated rhythms using neural oscillators. In *Proceedings of the 12th International Society for Music Information Retrieval Conference, Miami*. pp. 185–190.

Velleman, P. (2008) Truth, damn truth, and statistics. *Journal of Statistics Education*. 16(2), pp. 1–14.

Waldschmidt, M. (2009) Report on some recent advances in Diophantine approximation. *arXiv preprint arXiv:0908.3973*. [online]. Available from: http://arxiv.org/abs/0908.3973 (Accessed 5 May 2013).

Walker, R. (1996a) Open peer commentary: can we understand the music of another culture? *Psychology of Music*. 24(2), pp. 103–114.

Walker, R. (1996b) A reply to the commentaries. *Psychology of Music*. 24(2), pp. 125–130.

Ward, W. E. (1932a) Music of the Gold Coast. *The Musical Times*. 73(1074), pp. 707–710.

Ward, W. E. (1932b) The Music of the Gold Coast (Continued). *The Musical Times*. 73(1075), pp. 797–799.

Ward, W. E. (1932c) The Music of the Gold Coast (Concluded). *The Musical Times*. 73(1076), pp. 901–902.

Warren, R. M. *et al.* (1972) Auditory induction: perceptual synthesis of absent sounds. *Science*. 176(4039), pp. 1149–1151.

Waterman, R. A. (1948) ‘Hot’ rhythm in Negro music. *Journal of the American Musicological Society*. 1(1), pp. 24–37.

Waterman, R. A. (1952) *African influence on the music of the Americas*. Sol Tax (ed.). Chicago, IL: University of Chicago Press.

Wessel, D. L. (1979) Timbre space as a musical control structure. *Computer Music Journal*. 3(2), pp. 45–52.

Whittle, P. (2010) *Neural nets and chaotic carriers*. London, UK: Imperial College Press.

Widrow, B. & Hoff, M. E. (1960) Adaptive switching circuits. *IRE WESCON Convention Record*. pp. 96–104.

Wiggins, S. (2003) *Introduction to applied nonlinear dynamical systems and chaos*. New York, NY: Springer-Verlag.

Wilson, H. R. & Cowan, J. D. (1972) Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*. 12(1), pp. 1–24.

Witten, I. H. (2011) *Data mining: practical machine learning tools and techniques*. 3rd ed. Amsterdam, Netherlands: Elsevier/Morgan Kaufmann.

Wolf, D. J. (2003) Alternative tunings, alternative tonalities. *Contemporary Music Review*. 22(1-2), pp. 3–14.

Wolfram, S. (2002) *A new kind of science*. Champaign, IL: Wolfram Media.

Wright, J. K. & Bregman, A. S. (1987) Auditory stream segregation and the control of dissonance in polyphonic music. *Contemporary Music Review*. 2(1), pp. 63–92.

Yang, H. H. (1995) Some results on the oscillation of neural networks. In *Proceedings of Nonlinear Theory and Its Applications, Las Vegas*. pp. 239–242.

Yaroslavsky, L. (2007) Fast discrete sinc-interpolation: a gold standard for image resampling. *Advances in Signal Transforms: Theory and Applications*. 7, p. 337.

Yuan, T. (n.d.) *A C++ producer-consumer concurrency template library* [online]. Available from: http://www.drdobbs.com/a-c-producer-consumer-concurrency-templa/184401751 (Accessed 22 November 2014).

Zwicker, E. (1961) Subdivision of the audible frequency range into critical bands (frequenzgruppen). *The Journal of the Acoustical Society of America*. 33(2), p. 248.

Appendix A

Theory and Perception in Western Music

In this section we eclectically present particular theoretical and psychoacoustic aspects of Western tonal music. This material is supplementary, and may assist the reader who is overwhelmed by the references made to music theoretical concepts in the course of this treatise. Particular psychoacoustic phenomena which would seem to participate in musical perception are also reviewed.

A.1 Western Music Theory

The following sections will present a somewhat skewed account of the principles of Western music theory as they pertain to the research topic. Much will necessarily be omitted, as the emphasis will lie on those topics which contextualise the approach taken here, specifically those topics which might be taken for granted by some readers.

A.1.1 Notes in Western music theory

The foundational elements of Western music theory are typically illustrated with reference to the keys of the piano. It can be seen that every group of seven adjacent white keys is accompanied by a particular configuration of five black keys. This arrangement is repeated across the length of the keyboard. The seven white keys in each group are each uniquely named using the letters A through G, and these names similarly apply to every similarly placed white key within every repeated group throughout the keyboard layout. Thus, one may select any white key, count seven white keys up or down, and land on a new key with the same name. The black keys are named with reference to the white keys adjacent to them. The black key between A and B, for example, may be called A-sharp or B-flat. Black keys also maintain their names within the group across the length of the keyboard. Close to the middle of a typical piano lies "middle C", easily identifiable by the location of two (rather than three) black keys immediately to the right. This is a common reference point to which we shall return frequently.

If we play middle C, followed in turn by each of the seven white keys on its right (thus ending on the next C), we hear the familiar sound of a *major* scale. If we similarly play middle C, followed by every adjacent key to its right (both white and black) up to the same C as before (that is thirteen keys in all), we hear the *chromatic* scale. In both cases, we have traversed the range of one *octave* (the distance from any given key to the next similarly named key in either direction). The identical naming of different keys reflects a basic tenet of Western music theory, termed *octave equivalence*, whereby notes that are any number of octaves apart are perceived to share a common *pitch class* (or *chroma*), being perceived as more similar to each other than to any other keys on the keyboard. They also share a particular acoustic relationship, being that the fundamental frequencies produced by two such notes are related as $1{:}2^n$, with n any positive or negative integer indicating the number of octaves between them ($|n|$, of course). Various notational schemes allow one to distinguish between different notes of the same pitch class. In this document we have adopted the following common convention: the note middle C, along with all notes in the octave to the right of it, up to but excluding the next C, carry the ordinal suffix "4". From the excluded C, the notes of the next octave are suffixed as "5", then "6", and so on. Octaves to the left of middle C are thus suffixed with "3", "2", "1", "0", "-1", and so on. The standard 88-key piano spans the range from A_0 to C_8. This does raise one small inconsistency which requires clarification. As we play our *major* scale, as above from C_4 (middle C) to C_5 (the next C to the right of middle C), we encounter C_4, D_4, E_4, F_4, G_4, A_4, B_4 and finally C_5. In other words, though letter naming wraps around at G, octave numbering is always relative to C.

The notes of the chromatic scale are perceived as "equally spaced" from each other, and their fundamental frequencies are in fact related as a geometric series with coefficient $2^{1/12}$. This division of every octave into twelve, equally spaced, discrete pitches (inclusive of the upper and lower bound) is termed *twelve tone equal temperament* (12tet). Each pitch is separated from the next by a *semitone*, the smallest *interval* employed in modern Western music theory. The notes of the major scale, on the other hand, can be seen to be omitting certain of the chromatic notes (all the black keys in the above example), and thus proceeds by a particular asymmetrical combination of *tones* (also called *whole steps*) and *semitones* (also called *half-steps*). The interval of a tone is thus the product of two semitones in terms of its intervallic ratio ($2^{1/12} \times 2^{1/12} = 2^{1/6}$), or the "sum" of two semitones in terms of its perceived intervallic distance. A major scale starts on a particular note, then proceeds upwards by tone, tone, semitone, tone, tone, tone, semitone, thus arriving at the note an octave higher (the pattern can of course be reversed for a descending scale). It is to this asymmetry that we attribute the perceptual hierarchy perceived within the major scale, which by convention labels one pitch as most characteristic of the entire scale. For a major scale consisting of only white keys, this is C, and in general the term *tonic*, *doh*, or *first degree* might be used to denote this pitch. The major scale may be built on any starting pitch by applying the same pattern of tones and semitones to that pitch. Thus a D major scale will replace two of the white notes (F and C) with black notes immediately to their right (F-sharp and C-sharp) in order to maintain the same pattern. We say that F and C are *sharpened*. Similarly, the major scale built on F requires that we *flatten* the B (a B-flat, rather than a B).

In all this produces twelve perceptually distinct major *keys*, though there are theoretically more. The latter use of the term "key" is distinct from the physical keys of the keyboard, referring as it does to the pattern of tones and semitones, though keyboardists tend to think of both in terms of a physical location or pattern of locations on the keyboard. Modern Western musical practise, for the most part, favours adherence to a single key for a substantial period of time. In other words, a given piece of music uses predominantly the notes of that key, with exceptions, called *accidentals*, typically understood to suggest a temporary change of key. When such a change of key spans a substantial section of the music, we speak of *modulation* to that key. Accidentals further play a more cardinal role in *minor* keys (which have a different arrangement of tones and semitones) in that they routinely occur on the sixth and seventh degrees of the scale as characteristic of certain melodic and harmonic contexts.

The fundamental frequency of A_4 is by convention fixed at 440 Hz. Thus we can determine the fundamental frequency of any other note by counting the number of semitones down to A_4 (a negative result if the note in question is lower than A_4), raising the semitone ratio to this power, and multiplying by 440 Hz.

$$f = 440\,Hz \cdot \left(2^{\frac{1}{12}}\right)^{n}$$

Thus the frequency of middle C, for example, is fixed at 261.63 Hz, or that of C_5 at 523.25 Hz. In this case, rounding obfuscates the fact that one of these frequencies should, in theory, be precisely twice the other. However, we should also note at this stage that various real-world constraints do necessitate deviations from the ideal. The piano, for instance, exhibits an acoustic peculiarity termed *inharmonicity* whereby certain strings, being under tremendous tension, tend to behave more like stiff bars than ideal strings, and the resultant deviations from a perfectly harmonic overtone series require that octaves be stretched in higher registers to avoid "beating" partials. Nonetheless, the ideal of 2^n octave relations is maintained more or less throughout the most commonly employed middle register, as are the equal geometric division of each octave into twelve semitones.

Given the key in which a piece of music is based, it is common to refer to elements of the music in relativistic terms. The notes of the major scale, from lowest to highest, are termed *tonic* (or *do* or first degree), *supertonic* (or *re* or second degree), *mediant* (or *mi* or third degree), *subdominant* (or *fa* or fourth degree), *dominant* (or *so* or fifth degree), *submediant* (or *la* or sixth degree) and *leading-note* (or *ti* or seventh degree).[86] The eight note, of course, is the tonic again.

---

[86] Note that the *Solfège* or *tonic solfa* terms (do, re, mi, etc.) are different for minor keys, and are sometimes used absolutely rather than relatively. There are numerous variations on the basic system (*so* or *sol*, *ti* or *si*?). We simply include these here for comparison.

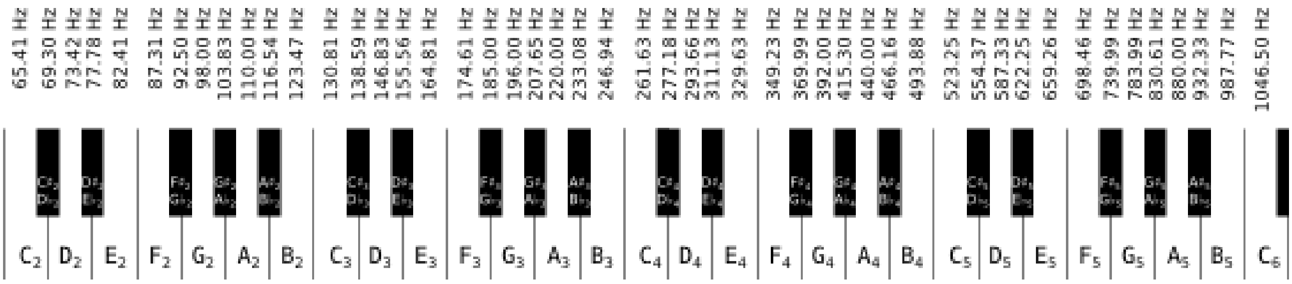


Figure A.1: Equal Temperament frequencies across four octaves, centred on "middle C"

A.1.2 Harmony in Western music theory

The term *harmony* is multi-faceted, but we will begin by approaching it as the sounding together of a number of notes. When two different notes are sounded together, we call this a dyad. Such a dyad is characterised by the *interval* between the component pitches, counted in scale steps. Thus, sounding any two adjacent notes of a scale comprises the dyad of a *second*. Selecting any two notes separated by only one other note gives the interval of a *third*, and so on (… ninth, thirteenth, etc.) Due to the asymmetry of the scale, however, such intervals do not always span the same number of semitones. Consider the the scale of C major (all the white keys). The third between C and E spans four semitones, while that between D and F spans only three. In fact, all *diatonic* thirds in the major scale are either *major thirds* (four semitones) or *minor thirds* (three semitones). Similar distinctions exist for other intervals, ranging through *minor*, *major*, *diminished*, *augmented* or *perfect* qualifiers. Our interest here lies specifically with the major and minor thirds.

A *triad* is the sounding together of three notes, most typically separated by thirds. The triad on C (in the C major scale) is termed the *tonic triad*, since it originates on the tonic note, and contains the notes C, E and G. The distinguishing features of this triad are its major third (between the lower notes C and E), minor third (between the upper notes E and G) and resultant *perfect fifth* (between C and G, the outer notes). The same structure is found in the *subdominant triad* (F, A and C) and *dominant triad* (G, B and D). All three are termed *major triads*. Conversely, the *supertonic triad* (D, F and A), *mediant triad* (E, G and B) and *submediant triad* (A, C and E) each contain a minor third between the lower notes and a major third between the upper notes. The outer interval remains a perfect fifth. These are *minor triads*. The remaining scale degree is the *leading note*, on which is produced the *diminished triad* (B, D and F) consisting of two minor thirds, spanning an outer interval of a *diminished fifth*. Any combination of simultaneously sounding notes may be referred to as a *chord*, but it is a particular trait of traditional Western tonal music theory to interpret such a chord in terms of a "stacking" of major and minor thirds on top of some *root* note. By extending this principle we obtain quartads (three stacked thirds or four notes, e.g. C, E, G and B) and quintads (add a D), sextads (plus F) or septads (plus A). Further extension simply repeats the same series, and sextads and septads tend to be too crowded for typical musical use. However, considering only the possible chords within the major scale, this *tertian* (based on thirds) approach to harmony already yields three different kinds of triad, four kinds of quartad and five kinds of quintad. Even more of the combinatorial options are exploited in *chromaticism*, an approach to harmony which more freely makes use of accidentals, thus frequently moving *outside of the key* (as in Jazz, for instance).[87] We might also acknowledge the possibility of constructing chords by the stacking of some interval other than the third. The central notion here, however, remains most common: tertian harmony as the stacking of major and minor thirds on some scale degree to obtain a chord.

The notes theoretically required of a particular chord may not all be present, nor will they necessarily be ordered as above. Musical context and convention may dictate the omission of certain pitches under defined circumstances, and the principle of octave equivalence allows any pitch to be transposed up or down by an arbitrary number of octaves. In particular, when the lowest sounding note of a chord is not the *root* of the chord, then we say that the chord has been *inverted*. Thus the lowest note of a chord might, in fact, be its *third* (the note a third above its root), its *fifth*, *seventh* or so on. Various conventions govern the suitability of such an inversion in a particular context.

In the theory of "chord functions", the three major triads (tonic, subdominant and dominant) are each perceived to represent a distinct category of chords, each containing a relative minor (submediant, supertonic and mediant, respectively) plus their various tertian extensions. This association is by virtue of the two notes shared between a major triad and its associated minor triad. The leading note triad, containing as it does the upper notes of the dominant quartad, is associated with the dominant function. Chord function emphasises the interpretation of a given chord in relation to its global context, namely the key. Shifting between chord functions lies at the heart of what is termed *harmonic progression*. Specifically, the movement from a dominant to a tonic chord (often notated as V-I), including various functional derivatives thereof, is regarded as the prime mover in establishing "motion towards a goal" in Western tonal music, being strongly associated with the perceived characteristics of tension and resolution. However, we should immediately note that the term "goal" might also be regarded more in the sense of an outcome than a destination. In such a case, the goal is the unambiguous establishment of a tonality. A traversal of tonic, subdominant and dominant functions achieves just that, having employed all of the notes of the scale.

A.1.3 Harmonic series in relation to the "evolution" of harmony

Pythagoras is generally credited with having "discovered" (or at least documented) various relationships between lengths of plucked string or blown pipes and the acoustic frequencies produced by these.[88] In particular, he demonstrated that the pitch produced by plucking a given string under tension stood in a

---

[87] It offends our own sensibilities to be quite so willing to generalise on this, but the example given is a readily acknowledged one. Both chromatic and diatonic tendencies are to be found to some degree within any "genre".

[88] Much of what is attributed to Pythagoras is likely better attributed to "The Pythagoreans", but we will not let this detract from the point here.

specific relationship to the pitch produced by halving the length of that same string while maintaining that tension. This relationship was, and still is, recognised as a measure of great similarity between two pitches, surpassed only by the greater similarity experienced between two notes having identical fundamental frequencies (said to be in unison). In most musical traditions of which we are aware, different frequencies which are 2^n multiples of each other (n being any positive or negative integer) are strongly associated as having the same fundamental identity, are thus given the same name, and are said to be the same note occupying different octaves. Thus a blown pipe (such as a reed flute) would sound a certain note, and if that pipe were cut to ⅛th (2^{-3}) of its initial length, the same note would be produced 3 octaves higher. The principle is carried further, in that the subdivision of the same pipe or string into thirds produces a pitch which, though certainly not the same note, is highly "consonant" with the first pitch. By this term is meant that the two pitches so produced are found to acoustically complement each other, and are thus strongly associated (though not as strongly as in the case of the unison or octave above). The association also holds between these two notes regardless of either or both being shifted up or down by any number of octaves, due to the strong identity relationship inherent in octave shifts (or octave *transpositions*). Taking two strings under tension which produce the same pitch, halving one and dividing the other at a point one third of the way along its length, produces two pitches which stand in the important relationship of a perfect fifth. The term "fifth" relates to the discussion of the modern major scale above, and we should now note that the fifth produced by these two strings differs from the fifth of equal temperament. Specifically, $3/2 \neq 2^{7/12}$.

We seldom relate musical pitches to string or pipe lengths, though these of course determine the wavelengths produced. Rather, it is the reciprocal parameter of frequency which is typically invoked in discussions of this nature, frequency (in Hz) being the quotient of the propagating acoustic wave's velocity (in $m.s^{-1}$) and its wavelength (in m). Thus the frequency produced by a given pipe or string doubles when that pipe or string is halved (since the wavelength is similarly halved), or triples when the same is shortened to a third of its original length. The octave described above is typically indicated as that interval (the perceived spectral distance between the components) formed by a given fundamental (f) and its second harmonic ($2f$). The perfect fifth is the interval formed between the second harmonic ($2f$) and the third harmonic ($3f$). The interval between the third harmonic ($3f$) and the fourth harmonic ($4f$) is called the perfect fourth, and is complementary to the perfect fifth in that they together form an octave (the interval between $2f$ and $4f$). Continuing the pattern, the octave between the fourth and eighth harmonics produces four more intervals, each perceived as progressively "smaller" (though the frequencies are of course in constant linear relationship to the fundamental of the series). These are the major third (ratio 5:4), the minor third (ratio 6:5), a smaller minor third (ratio 7:6), and a slightly large major second (ratio 8:7). The following octave produces eight intervals, of which only the first will be relevant to this discussion. This is the major second found between the eighth and ninth harmonics (ratio 9:8), significant by the realisation that this is the same interval produced by raising a given pitch by two perfect fifths (multiply f by 3/2 twice) and lowering the result by an octave (divide by 2), and simultaneously the difference between a perfect fifth (3/2) and a perfect fourth (4/3). The latter insight might have given some impetus to the idea that the cycle should ultimately close on itself, though this will shortly be demonstrated not to be the case.

The phenomenon described above relates to the fundamental acoustic frequency produced by the subdivision of a vibrating string or of a pipe containing a vibrating air column. The same principle, however, is at work even in single pitches produced by musical instruments. Any string under tension, when excited, vibrates not only at a specific fundamental frequency, but also at integer multiples of that frequency (similarly with blown pipes). It is this harmonic series (the concurrent sounding of a fundamental with various integer-multiple-valued frequencies having arbitrary amplitudes) which motivates the idea that, given two pitches, their degree of consonance is a function of the agreement between their respective harmonic series. From this it is clear that: the greatest consonance is achieved when both pitches are the same, or when they are separated by one or more octaves; a fairly high degree of consonance is apparent when they are separated by a fourth or fifth (as described above); somewhat less when they are separated by a major or minor third; and even less when separated by a major second. Ignoring various other intervals implicated along the way, the examples provided lay a basis for an "evolutionary" view of musical harmony, wherein Western music theoretic practice is seen to have "gradually crept up the harmonic series". This line of reasoning points out the almost exclusive use of the unison and octave in sacred music of the middle ages, followed by the introduction of parallel fourths and fifths in styles known as organum, later the gradual acceptance of major and minor thirds as consonant, later seconds (at first only in strictly controlled contexts, but later more liberally, e.g. Jazz), and ultimately half-tones (or semitones) in later harmonic practice (chromatic harmony).

The purpose of introducing such a view into this study is to point out that the Western system of musical harmony is the product of centuries of speculation and experimentation, and should never be regarded as having been "designed" in one fell swoop. As such, it has inherited axioms along the way which might deserve to be challenged.

A.1.4 Intervals with reference to equal temperament and its predecessors

We now pause to contemplate one of the central dilemmas of Western music theory: repeatedly raising a given pitch by octaves, and repeatedly raising the same pitch by perfect fifths will never converge to the same frequency. The compromise adopted by Western music, $(2/1)^7 \approx (3/2)^{12}$, creates a discrepancy whose obfuscation becomes a central theme in the development of music theory throughout the 13th – 18th centuries. The ratio between these two quantities is 531441/524288, or a 1.364% variance on the lower frequency, and is dubbed the "Pythagorean comma". We should then clarify why it would even be desirable that a geometric series of octaves should meet a geometric series of perfect fifths.

Prior to the modern notion of music embodying a particular key (the idea that a particular pitch, including all its octave transpositions, represents the auditory focal point of that music), it was nonetheless clear that an arbitrary number of pitches might be found within any octave. The desire to repeatedly perform composed sequences of such pitches required a generative principle from which those pitches might be consistently selected. Since the principle of octave equivalence is of no use in this regard (it merely establishes the boundaries within which this subdivision must take place), the next closest association (the perfect fifth) was recursively employed to calculate the string or pipe lengths required to populate the octave. This "Pythagorean tuning system" thus defined a number of discrete pitches per octave, each derived from the chosen reference pitch by recursively applying the ratio 3:2 to string or pipe length (remove a 3rd of the length), and normalising the results to a single octave by applying the ratio 1:2 as necessary (double the length). From the point of view of frequency, a given frequency within some octave is multiplied by 1.5 and, if the result falls outside of that octave, divided by 2 to bring it back within bounds. The results of applying this principle to an arbitrary frequency (pitch ordinal 0) are illustrated in the following table:[89]

<table>
<tr><th>pitch ordinal</th><td colspan="2">0</td><td colspan="2">1</td><td colspan="2">2</td><td colspan="2">3</td><td colspan="2">4</td><td colspan="2">5</td><td colspan="2">6</td><td colspan="2">7</td><td colspan="2">8</td><td colspan="2">9</td><td colspan="2">10</td><td colspan="2">11</td><td colspan="2">12</td></tr>
<tr><th>ratio</th><td colspan="2">$\frac{1}{1}$</td><td colspan="2">$\frac{256}{243}$</td><td colspan="2">$\frac{9}{8}$</td><td colspan="2">$\frac{32}{37}$</td><td colspan="2">$\frac{81}{64}$</td><td colspan="2">$\frac{4}{3}$</td><td colspan="2">$\frac{729}{512}$</td><td colspan="2">$\frac{3}{2}$</td><td colspan="2">$\frac{128}{81}$</td><td colspan="2">$\frac{27}{16}$</td><td colspan="2">$\frac{16}{9}$</td><td colspan="2">$\frac{243}{128}$</td><td colspan="2">$\frac{2}{1}$</td></tr>
<tr><th>chromatic interval</th><td rowspan="2"></td><td colspan="2">$\frac{256}{243}$</td><td colspan="2">$\frac{2187}{2048}$</td><td colspan="2">$\frac{256}{243}$</td><td colspan="2">$\frac{2187}{2048}$</td><td colspan="2">$\frac{256}{243}$</td><td colspan="2">$\frac{2187}{2048}$</td><td colspan="2">$\frac{256}{243}$</td><td colspan="2">$\frac{256}{243}$</td><td colspan="2">$\frac{2187}{2048}$</td><td colspan="2">$\frac{256}{243}$</td><td colspan="2">$\frac{2187}{2048}$</td><td colspan="2">$\frac{256}{243}$</td><td rowspan="2"></td></tr>
<tr><th>diatonic interval</th><td colspan="4">$\frac{9}{8}$</td><td colspan="4">$\frac{9}{8}$</td><td colspan="2">$\frac{256}{243}$</td><td colspan="4">$\frac{9}{8}$</td><td colspan="4">$\frac{9}{8}$</td><td colspan="4">$\frac{9}{8}$</td><td colspan="2">$\frac{256}{243}$</td></tr>
</table>

Table A.1: Pythagorean tuning

Implicit in this system is the idea that pitches are related to each other by the number of such multiplication (and possibly division) operations required to move from one to the other. The expectation, as noted above, is that the "cycle" would ultimately close, and that a particular number of these operations would eventually bring one back to the same reference pitch, but this is not the case. Notwithstanding this dilemma, Pythagorean tuning proved to be a highly effective principle in music which seldom strayed far from the reference pitch (in terms of the associative principles described above), medieval vocal chant providing notable examples hereof. As musicians and audiences gradually accepted thirds as consonances, it became desirable to exchange the Pythagorean thirds (ratio 81:64, obtained by raising a given pitch by four fifths and lowering by two octaves, i.e. multiply by 1.5^4 and divide by 2^2) for the more aesthetically pleasing (and presumable more closely related) third occurring between harmonics four and five of the natural harmonic series (5:4, or for comparison, 80:64). This difference, too, was deemed deserving of a special place in the music theory of the day and is now known as the "Syntonic comma", a variance of 1.25% above the lower frequency. Thus the Just Intonation system invoked three harmonic ratios in its construction,

---

[89] In order to foreground the argument presented here, enharmonics have been omitted from the tables presented.

namely the octave (2:1), the fifth (3:2) and the major third (5:4). Starting from any particular reference pitch, today termed the tonic, the 3:2 ratio and its octave normalised inverse, 4:3, were first used to mark out two important counterpoints, today termed the dominant and subdominant respectively. Over each of these three was added a Just major third (5:4) and a perfect fifth (3:2). The same idea was applied to various other pitches so obtained in order to fill in the more remote pitches. The resulting subdivision of an arbitrary octave is illustrated as follows:

pitch ordinal	0	1	2	3	4	5	6	7	8	9	10	11	12
ratio	$\frac{1}{1}$	$\frac{16}{15}$	$\frac{9}{8}$	$\frac{6}{5}$	$\frac{5}{4}$	$\frac{4}{3}$	$\frac{45}{32}$	$\frac{3}{2}$	$\frac{8}{5}$	$\frac{5}{3}$	$\frac{9}{5}$	$\frac{15}{8}$	$\frac{2}{1}$
chromatic interval		$\frac{16}{15}$	$\frac{135}{128}$	$\frac{16}{15}$	$\frac{25}{24}$	$\frac{16}{15}$	$\frac{135}{128}$	$\frac{16}{15}$	$\frac{16}{15}$	$\frac{25}{24}$	$\frac{27}{25}$	$\frac{25}{24}$	$\frac{16}{15}$
diatonic interval			$\frac{9}{8}$		$\frac{10}{9}$	$\frac{16}{15}$		$\frac{9}{8}$		$\frac{10}{9}$		$\frac{9}{8}$	$\frac{16}{15}$

Table A.2: Just intonation

It remains a matter of some speculation whether the composition of increasingly harmonically complex music drove the development of tuning systems, or whether it was the introduction of refined tuning systems that invited more harmonically complex compositions. We might reasonably assume that both factors were in play, particularly when considering the extended period of time in which these developments took place, and the fact that different geographical locations adopted such at different times. It also seems plausible that the Pythagorean major third (81:64) gave little offence until both pitches were sounded simultaneously. This would lend credence to the idea that musical harmony has a strong physical basis, in addition to whatever psychological basis might account for perceptions of consonance between notes sounding at different times. Certainly the introduction of Just intonation (and numerous variants thereof) paralleled the rise of a new approach to musical composition wherein the previously static "reference point" (what is generally termed the "key" of the music) was commonly caused to change during a particular piece of music. A familiar pattern, for example, might be that a composition would start out in one key (denoted the tonic), "modulate" (the term essentially describes the shift of key centre) to a closely "related" key (most notably the key built on the dominant, or in other words, the pitch having a ratio of 3:2 to the tonic key), and then return to the original key before ending.

Both of the above systems, and numerous variations on these, sought to characterise their intervals in terms of the natural harmonic series, and thus by harmonic ratios. Herein lies the importance of Mean Tone temperament: that it introduced geometric ratios. Specifically, it discarded the perceptibly irregular division of the Just major third (5:4) into a major tone (9:8) and a minor tone (10:9), differing from each other by a Syntonic comma (presented earlier), in favour of a single sized (in terms of human perception) tone having a ratio √5:2. It also “tempered” (slightly altered) the harmonic fifths (3:2) so as to cause a cycle of four fifths to equal the third occurring as fifth harmonic in the natural harmonic series. This gives the fifths ratios of ^{4}√5:1, slightly smaller than naturally harmonic fifths. A complete octave would be constructed thus:

pitch ordinal	0	1	2	3	4	5	6	7	8	9	10	11	12
ratio	$\frac{1}{1}$	$\frac{16}{15}$	$\frac{9}{8}$	$\frac{6}{5}$	$\frac{5}{4}$	$\frac{4}{3}$	$\frac{45}{32}$	$\frac{3}{2}$	$\frac{8}{5}$	$\frac{5}{3}$	$\frac{9}{5}$	$\frac{15}{8}$	$\frac{2}{1}$
chromatic interval		$\frac{(\sqrt[4]{5})^7}{4}$	$\frac{2}{(\sqrt[4]{5})^5}$	$\frac{8}{(\sqrt[4]{5})^5}$	$\frac{(\sqrt[4]{5})^7}{16}$	$\frac{8}{(\sqrt[4]{5})^5}$	$\frac{5}{16}$	$\frac{8}{\sqrt{5}}$	$\frac{(\sqrt[4]{5})^7}{16}$	$\frac{8}{(\sqrt[4]{5})^5}$	$\frac{8}{(\sqrt[4]{5})^5}$	$\frac{(\sqrt[4]{5})^7}{16}$	$\frac{8}{(\sqrt[4]{5})^5}$
diatonic interval		$\frac{\sqrt{5}}{2}$		$\frac{\sqrt{5}}{2}$		$\frac{8}{(\sqrt[4]{5})^5}$	$\frac{\sqrt{5}}{2}$		$\frac{\sqrt{5}}{2}$		$\frac{\sqrt{5}}{2}$		$\frac{8}{(\sqrt[4]{5})^5}$

Table A.3: Mean-tone temperament

Here, we suggest, it first becomes plainly apparent that practical concerns (in fact, technological limitations) have usurped the grip of “harmonic” in favour of “geometric”. Ultimately this opens the door on an almost complete abandonment of “harmonic” principles, with the entire octave being subdivided into twelve geometrically equal divisions. This is Equal Temperament, the *de facto* standard of our day. For comparison, an octave is presented below:

pitch ordinal	0	1	2	3	4	5	6	7	8	9	10	11	12
ratio	1	$(\sqrt[12]{2})^1$	$(\sqrt[12]{2})^2$	$(\sqrt[12]{2})^3$	$(\sqrt[12]{2})^4$	$(\sqrt[12]{2})^5$	$(\sqrt[12]{2})^6$	$(\sqrt[12]{2})^7$	$(\sqrt[12]{2})^8$	$(\sqrt[12]{2})^9$	$(\sqrt[12]{2})^{10}$	$(\sqrt[12]{2})^{11}$	2
chromatic interval		$\sqrt[12]{2}$	$\sqrt[12]{2}$	$\sqrt[12]{2}$	$\sqrt[12]{2}$	$\sqrt[12]{2}$	$\sqrt[12]{2}$	$\sqrt[12]{2}$	$\sqrt[12]{2}$	$\sqrt[12]{2}$	$\sqrt[12]{2}$	$\sqrt[12]{2}$	$\sqrt[12]{2}$
diatonic interval		$(\sqrt[12]{2})^2$		$(\sqrt[12]{2})^2$		$(\sqrt[12]{2})^1$	$(\sqrt[12]{2})^2$		$(\sqrt[12]{2})^2$		$(\sqrt[12]{2})^2$		$(\sqrt[12]{2})^1$

Table A.4: Equal temperament

The four systems presented above merely scratch the surface of centuries of musical practice and theory, but they provide enough of a context to make the following observations:

- Harmonic principles were valued by earlier musicians, but gave way to geometric compromises as the complexity of musical composition increased;
- By increased complexity, we mean a trend toward shifting key centres, as well as the simultaneous sounding of different pitches in various configurations;
- Eventually harmonic principles gave way, except in the sole case of the octave, and were replaced by equal geometric divisions.

Is the compromise a suitable one? For the moment we can simply state that the degrees of variance between the geometrically-derived equal temperament pitches and their harmonically-derived counterparts are deemed by most to be sufficiently small to justify their acceptability.

A.2 Psychoacoustics & Musical Perception

The following sections briefly review the most important psychoacoustic phenomena related to musical perception. Once again, an in-depth treatment is not offered here, though it hoped that this will suffice for the present purpose. Please consult the references for more, should this be needed.

A.2.1 Critical bandwidth

In the study of human audition, two ostensibly related meanings attach to the term "critical bandwidth". Within psychoacoustics, evidence suggests that pitched sounds are detected by a battery of band-pass filters, thus accounting for various observed masking effects (particular sounds not being perceived in the presence of other, nearby, slightly louder sounds). Physiological studies of the ear find evidence to suggest that stimulation of the Basilar membrane by a particular frequency results in a response over a similar bandwidth. [90] The characteristics of this "auditory filter" are found to depend both on frequency and on the intensity of the sound perceived. At low intensity, the filter is found to be relatively symmetrical, skewing toward high frequencies as that intensity increases. Bandwidth increases with linearly increasing frequency, but effectively decreases in relation to the exponentially increasing octave. This will shortly be seen to account for the relativity of what is deemed "consonant" in relation to the actual frequencies involved. Critical bandwidth is typically modelled in terms of a computationally simpler Equivalent Rectangular Bandwidth (ERB), proposed as that rectangular band-pass filter passing the same total power as the auditory filter. The formula for calculating ERB is: $24.7(4.37f_c + 1)$ Hz, where f_c is the filter centre frequency in kHz (valid for $100 \text{ Hz} < f_c < 10 \text{ kHz}$) (Glasberg & Moore 1990:114). The concept of critical bandwidth is typically invoked to account for a subject's inability to perceive a particular sound in the presence of certain other sounds (so-called "masking effect"). It also finds application in the "Place Theory" of human pitch perception, but herein it fails to account for the significantly finer resolution of pitch discrimination observed in humans (typically around 1/30th of critical bandwidth for any given centre frequency). Nonetheless, the principle serves well in addressing the question of why certain combinations of notes are deemed more consonant than others, as will be discussed in the next section.

A.2.2 Consonance & dissonance

The term "consonance" essentially indicates a degree of aesthetic accord between different sounds, whereas "dissonance" (in some contexts, "dischord") refers to a perceived disagreement or conflict between

---

[90] A comprehensive account of the physiology of the ear, at least for our purposes, is to be found in (Mazzola, 2002:1035-1046)

different sounds. The history of Western musical theory bears witness to the subjectivity of such a judgement, having described certain combinations as dissonant at one point, and consonant at another. Moreover, the trend might be described as "evolutionary", in that today's dissonance might well be tomorrow's consonance, though the reverse is less likely (the perfect fourth is an interesting exception). Since Rameau (Rameau 1971), the major and minor thirds have come to represent the standard for musical consonance, along with their inversions (in view of octave equivalence, either note can be transposed to lie on the opposite side of the other, thereby forming the complementary minor or major sixths, respectively). The intervals of the perfect fifth and octave are now seen as too consonant, and the fourth has even attracted the label of dissonance unless used in very specific ways. The interval of a diminished fifth, along with the smaller major and minor seconds (and their inversions, the minor and major sevenths, respectively) remain dissonant for the moment, even if exploited for this very characteristic in much modern music.

Accounting for perceptions of consonance and dissonance was, in the not too distant past, done largely in terms of the natural harmonic series and simple ratios. It was understood that a tensioned string having half the length of another similarly tensioned string would vibrate with twice the frequency of the other string when plucked. Similarly, a blown pipe having a third of the length of another pipe would sound a frequency three times that of the other pipe, all other factors being equal. With this reciprocity came the insight that a simpler proportional relationship (one which could be expressed by means of small integer values) would produce sounds that maintained a specific phase relationship. The slightest deviation would produce sounds which would gradually shift in and out of phase with one another. The result is expressed mathematically in the identity:

$$\sin A + \sin B = 2\sin\left(\frac{A+B}{2}\right)\cos\left(\frac{A-B}{2}\right) ,$$

or intuitively as a resultant tone at the average of the component frequencies (the sine portion), amplitude modulated by a cosine at half the difference between the component frequencies. Musicians refer to this slower modulation as "beats" or "beating". The phenomenon is widely exploited when tuning instruments to some reference, the goal being to tune so as to minimize "beating". However, instruments which are capable of producing more than one simultaneous sound (polyphonic instruments) manifest "beats" between their various notes as a result of the interaction between the harmonic series of each note. A violinist, for example, will produce notes on two adjacent strings that are expected to sound as a "beatless" fifth. This is achieved when the frequencies produced by the individual strings are in the ratio of 3:2, thereby causing the higher-pitched string's harmonics to sound in unison with the harmonic series of the lower-pitched string. By extension, any combination of fundamental tones will result in a similar interaction between harmonics of those fundamentals. It is in this way that the relative consonance or dissonance of chords (combinations of notes) might be quantified with reference to the amount of agreement between component harmonics. Such comparison might be conducted within the context of Plomp & Levelt's sensory dissonance curve (Plomp & Levelt 1965:556), which predicts perceived consonance or dissonance in terms of applicable critical bandwidth. Specifically, given a particular reference frequency, a particular critical bandwidth is implied. The relative consonance or dissonance of any other tone in relation to the reference tone might be predicted as follows: within 5% of applicable critical bandwidth is perceived as consonant; from 5% - 50% of applicable critical bandwidth is perceived as dissonant (with maximum dissonance at 25% of critical bandwidth); more than 50% of critical bandwidth is essentially consonant, increasingly so with increasing difference.

Appendix B

Priming Research in Music

We need some way in which to empirically validate the presence of directed motion, and moreover to pinpoint its "location" in the audio signal.[91] Traditional approaches would be content to analyse music in terms of chordal function within a prevailing musical key (Riemann 1896), finding the cause of directed motion in the theoretical progression away from or towards states of relative stability (Lerdahl & Jackendoff 1983; Schenker 1954). Such an approach will not serve our purposes here since it is inextricably bound to a theory of music which is increasingly acknowledged as being entwined in culture-specific percepts. Moreover, even if this were not the case, extant methods for the automated transcription of audio into common music notation (Klapuri & Davy 2007) only function well under tightly constrained circumstances, and are thus likely to encounter significant challenges when faced with non-Western music. Further, our position questions whether music notation necessarily provides the best platform for analysis in any case (McLean 1981). Rather, we intend to find the cause of the "directed motion" percept in the audio signal itself.

A number of issues immediately arise. Does "directed motion" exist outside of Western music? Is it possible, even in Western music, to definitively describe the percept we seek? Would we expect the same description to feature cross-culturally? For that matter, can we even expect any degree whatsoever of intra-cultural conformity in respect of such a description? The purported link between musical expectancy and emotional response (Meyer 1956) has led some to engage associated matters by way of emotional self-appraisals (Konečni 2008), but the strong dependence on language in such an approach is problematic for us here, as is the relative difficulty of extracting robust, numerically viable data from such reports. Our research has led us to an experimental paradigm which acknowledges these methodological difficulties in general, circumventing the need to articulate explicit concepts by instead tapping into implicit knowledge. We will now review, in some detail, the development of an experimental paradigm which has come to be known as *harmonic priming*, highlighting the various ways in which the experimental, analytical and interpretive methods have evolved in response to an emergent understanding of musical cognition.

In the broader field of psychological research, priming refers to a measured effect of prior experience on current performance, particularly as demonstrated by some quantitative measure in respect of a simple task.

---

[91] We must immediately acknowledge that the percept does not lie "in the signal" *per se*, but we hope the reader will grant us this latitude, metaphorically speaking. More precisely, we wish to identify the signal attributes which give rise to such a percept and, to whatever precision is feasible, their temporal location.

Effects have been observed in a range of domains, though most research, by far, has focused on linguistic and visual phenomena. A commonly invoked paradigm is *repetition priming*, wherein participants demonstrate various kinds of preference for previously encountered symbols or concepts over those that have not been encountered before, typically demonstrated in faster reaction times and lower error rates in recognition tasks, and better recall in memory tasks, for instance. *Semantic priming* demonstrates similarly assisted measures of performance in respect of semantically related concepts, typically symbols or concepts having some perceived logical association (e.g. “doctor” and “nurse”, but not “bread” (Tillmann *et al.* 1998:99)). Research in music cognition has steadily advanced the *harmonic priming* paradigm (referred to in earlier studies as *chord priming*), arguably a musical version of semantic priming, as evidence of domain-specific, cognition-assisted processing of music. In the main, ongoing debate has pitted two opposing accounts against each other: the spreading activation hypothesis (based on a modified semantic priming) and the overlapping spectra hypothesis (based on repetition priming), with validation of the former resting largely on invalidation of the latter. In this sense then, music is anomalous amongst human sensory experiences, with schematic expectations apparently trumping veridical expectations. The origins and precise characteristics of the schemata, however, remain elusive. Most accounts today favour a statistical model, advancing a view of tonal cognition as acquired by learning.

A simple demonstration of harmonic priming sounds a prime (usually a major triad) immediately preceding a target (generally another major triad). Some arbitrary aspect of some targets might be altered to some extent (such a modified target is termed a foil), and the participant is asked to make a perceptual judgement of the alteration (or lack thereof). Typical alterations include slight mistuning of one of the notes of the triad (participant is asked whether the target is “in-tune” or “out-of-tune”) or modification of the timbre of the target (participant is asked to discriminate bright sounds from dull sounds, for instance), though virtually any simple perceptual judgement may suffice. Whatever the task, participants are urged to respond as quickly and accurately as possible, and it is the response times and error rates which become the subjects of subsequent analysis. Broadly, results indicate a “processing advantage” which lowers response times and error rates in respect of pairs of prime and target which are more closely “harmonically related” in the sense of Western music theory. Failure of the repetition priming paradigm to account for these results has bolstered a view of harmonic priming as evidence of a largely cognitive, rather than perceptual process at work. We now review selected approaches and findings in the published literature.

It would appear that Jamshed Bharucha and Keiko Stoeckig were first to introduce the notion of “reaction time” as an indicator of harmonic priming (Bharucha & Stoeckig 1986).[92] Bharucha and Stoeckig contextualised their particular approach in the light of comments by Schenker and Meyer[93], as well as the

---

[92] In a subsequent paper, Bharucha pointed to precedents in the psycholinguistic study of semantic facilitation (Bharucha 1987:7).

[93] Specifically, Schenker's characterisation of goal-directed harmonic progression (Schenker 1954:219; Schenker &

then more recent theory posited by Mari Jones (Jones 1982), which viewed a musical context as a "... constellation of vectors, determined by previously internalized regularities" (Bharucha & Stoeckig 1986:403). Jones's vectors, in line with Schenker and Meyer, represent "ideal prototypes", a reference from which music derives its affective power (which she termed "psychological motion") by selectively deviating from the expectations implied by that reference. As to the origin of these "internalized regularities", Jones explicitly acknowledged the role of "enculturation" (Jones 1981:37), a view which arguably set the tone for Bharucha's and others' subsequent emphasis on statistical learning. Indeed, Bharucha promptly gave effect to the implications of enculturation by adapting the priming experimental design to a study of the Indian *rāg Bhairav* (Bharucha 1987:13–15). The key point to be made in this review is the extent to which these early studies established the methods employed by later studies. Notably, in their studies of priming in Western music, Bharucha and Stoeckig employed a single chord as prime throughout their three experiments.[94] Later studies have introduced expanded contexts consisting of longer musical progressions. In their first experiment, they set a major/minor discrimination task (1986:405–407), introducing an in-tune/out-of-tune discrimination task in two subsequent experiments seeking to disambiguate the results obtained in respect of major and minor contexts (1986:407–409).[95] Subsequent priming studies would seem to have avoided such crossing of major and minor contexts, and major/minor discrimination tasks have also been noticeably absent since. However, the major/minor interaction of the first experiment did yield a particular insight which does not seem to have received further attention, specifically that prime-target pairs consisting of a major and minor chord evoked different responses as a function of temporal order: "[w]hen related, responses were faster for major targets; when unrelated, responses were faster for minor targets." (1986:407) Though apparently similar asymmetries have continued to feature in later results, these have typically been in respect of mistuned foils or explicitly dissonant chords, rather than in terms of properly constructed minor chords. At the time, this observation was offered in support of the notion of minor chords as inherently less stable than major chords, and thus of the inherent bias in the related/unrelated condition,[96] and ultimately also as an explanation of the ambiguity observed in the first experiment. They did observe, in their second and third experiments (and as reported by others since), the "... strong bias to judge targets to be in tune when related and out of tune when unrelated." (1986:408,409) Finally, Bharucha and Stoeckig reported "... no significant correlation between priming and musical training", suggesting that "... a decision task such as in-tune/out-of-tune can fruitfully tap the underlying processes of listeners of all levels of training." (1986:410)

Bharucha's investigation of facilitated processing in Indian music (a component of (Bharucha 1987))

American Musicological Society 1979:5) and Meyer's linking of expectation and affect (Meyer 1956).

[94] Examples of the prime-target pairs serving the crossing of the various conditions can be seen in Table 1 of (Bharucha & Stoeckig 1986:405).

[95] In the first experiment, priming did not appear to have an effect in the minor context.

[96] In terms of Bharucha's own "spreading activation model", "... major target chords cause greater activation of their nodes than do minor target chords, independent of prior context, because of greater reverberatory activation from their parent keys. If the activation level of the target chord node is used to help guess if it is major or minor, prior activation (from a related prime) will bias the decision in favor of major." (Bharucha & Stoeckig 1986:407)

required the adaptation of a priming task to the "nonharmonic character" of that music. The *rāg Bhairav* was employed in such a way as to set up a strong schematic expectation for the target "*Pa*", primed as it was by the preceding downward glide to "*Dha*-". One of two targets were offered, either the highly expected "*Pa*", or an ostensibly less expected "*Dha*". In either case, participants were to indicate whether the targets were properly in tune, foils being a flattened "*Pa*" and sharpened "*Dha*", respectively. Preempting many later studies, Bharucha recognised the need to control for the possible effects of repetition priming by eliminating "*Pa*" from the context preceding the prime-target pair. Also, "in-tune" responses and "out-of-tune" responses were constrained to the right and left hand, respectively, controlling for a further potential confound. The characteristic drone accompaniment on "*Sa*" was included in the first experiment, but excluded in a subsequent experiment to eliminate the possibility of perceptual facilitation due to overlapping harmonic spectra. The results in both cases indicated that "[t]he expected tone was discriminated from its foil more quickly and more accurately than the control tone was discriminated from its foil" (Bharucha 1987:14).

More generally, Bharucha's (1987) discussion of the various issues informing his connectionist framework (what would later become the MUSACT model) engaged most, though not all, of the themes which have continued to underpin research employing the harmonic priming paradigm. These include the role of expectancy in affording music's dynamism and affective character, the interplay of schematic and veridical expectations (including the role of sequential memory traces), the automaticity of schematic representation and its effect on memory and perceptual judgement, the interaction between bottom-up (acoustically driven) and top-down (cognitively driven) processes (including the question of what role, if any, is played by overlapping harmonic spectra), and the respective roles of implicit (passive) learning and musical training. Brief mention was also made of an "in preparation" manuscript addressing the possibility of a continuum of facilitation, presumably a function of harmonic relatedness along the cycle of fifths, but tantalisingly, this does not appear to have been published.[97]

A matter which remains problematic for our own research is Bharucha's choice of representational unit. He acknowledged that his approach was "... atypical within connectionist modeling", but nevertheless insisted that "[i]n the absence of additional evidence, the choice of representational units for a psychological theory should be driven by our intuitions about what aspects of music are perceived to be unitary" (Bharucha 1987:23) Thus his model explicitly posited layers at the level of the note, chord and key (1987:16, Fig. 7), thereby playing directly into McLean's "extension trap" (McLean 1981). Certainly, it would be far more desirable for the units of representation to emerge from the data rather than being specified *a priori*, yet it is equally easy to share Bharucha's scepticism about the existence of "key detectors" in the auditory system (Bharucha 1987:4) and thus the need to have that particular percept built on lower levels of perception.

---

[97] This was cited as: Bharucha, J. J., & Stoeckig, K. *Graded activation of representational units for music: Evidence from priming*. Manuscript in preparation.

Working backwards then, the acknowledgement of key might reasonably seem to require chords and notes. Still, this remains a prime example of how decisions about systems architecture inevitably introduce cultural bias into any resulting analysis. Somewhat more heartening is Bharucha's observation that connectionist models, in general, afford retrieval by content rather than by index (1987:26), though we are inclined to question whether this is given full effect in a model built on *a priori* representational units, arguably an imposed index anyway.[98] The hypothesis of expectancy violation producing an "error signal", potentially a trigger to emotional response, and incrementally minimised to bring expectation in line with the environment (1987:26), plausibly addresses much of the empirical evidence pertaining to the perception of music at an abstract level,[99] and is moreover highly suggestive of machine learning architectures in general. However, Bharucha did point out, citing Fodor, that "... understanding a program requires no knowledge of the underlying hardware, since the same program can be instantiated by radically different hardware." (1987:27) Thus he did not regard neural plausibility as a criterion for evaluating a theory of cognition.

Turning to a matter not engaged above, further scrutiny was subsequently brought to bear on the effects of time-domain parameters such as stimulus onset asynchrony (SOA) and prime duration (Tekman & Bharucha 1992). A series of four experiments found priming effects to be active in respect of major chords with SOAs ranging from 50 ms to 2,500 ms, whether prime and target (or mistuned foil) were adjoined, separated by up to 2,450 ms of silence, or separated by a noise mask. Full crossing was implemented with regard to the different SOA and duration conditions, related and unrelated conditions (either perfect fifth or tritone interval between prime and target, respectively), and the in-tune or out-of-tune conditions (fifths were flattened slightly). The results largely underscored earlier findings, but are significant in that they revealed the speed with which harmonic priming is engaged (less than 50 ms), and the persistence of the effect (2,450 ms or more, despite the presence of masking noise) (1992:38). In view of effects observed in response to a 50 ms prime, followed by masking noise, Tekman & Bharucha suggested that "...processing of harmonic relationships continues independent of any involvement of an auditory sensory store, which would support the idea of a cognitive, rather than a sensory, basis for the priming effect." (1992:38) In other words, whatever contribution is of a sensory nature, if any at all, is probably made within the first 50 ms. Their results also revealed the non-linear fluctuation of priming effects in respect of the various time intervals. At this stage, they deferred discussion of the potential contribution of overlapping spectra to Bharucha & Stoeckig's earlier results, though that matter was to come under closer scrutiny in a later paper (see (Tekman & Bharucha 1998), reviewed below).

---

[98] We provocatively raise this contention only to highlight the extent to which Bharucha's MUSACT model, though conceptually promising for our own endeavour, remains rooted in precisely the kind of assumptions we are seeking to escape. Bharucha's "connections" are, after all, between notes, chords and keys, and thus between addressable indexes to the physical reality, though he seems to regard these as content. If representational units were truly to emerge from the data, however, we might properly be able to conceive of a "culture-free", data-driven theoretical framework.

[99] Schmuckler & Boltz mention, for example, "measures of perceived relatedness", "memory confusions" and "response times to musical events" (Schmuckler & Boltz 1994:313)

A significant avenue for expansion of priming research in music was proposed by Mark Schmuckler and Marylin Boltz, who raised the matter of whether harmonic priming effects were wholly independent of rhythm (Schmuckler & Boltz 1994). Their two experiments essentially drew on the various methods employed by Krumhansl, Shepard, Kessler *et al.* (expectancy/completion ratings) on the one hand, and Bharucha, Stoeckig, Tekman, *et al.* (speeded reaction time judgements) on the other, thus triangulating their findings. In both cases, their dataset consisted of four-bar extracts from twelve simple melodies in common time, each accompanied by one of three common four-chord progressions.[100] The immediate significance lies in the application of priming methods to stimuli consisting of more than two chords. These were then manipulated along three dimensions: harmonic identity of the final chord (high, medium or low expectancy); rhythmic placement of chord onsets for the first three chords (always on first beat of bar, or variously placed on the second, third or fourth beats); and rhythmic placement of the final chord (first beat of the bar, early or late by a beat). Their Experiment 1 required participants to assign expectancy ratings on a 7-point scale in respect of the final chord, while Experiment 2 required a speeded yes/no judgement of the appropriateness of the same. The results of both experiments were shown to be consistent, and moreover to support their hypothesis that harmonic expectancy is modulated to some degree, at least, by rhythmic factors. The temporal manipulations employed, however, hardly scratch the surface of what is found in real music, and Schmuckler & Boltz limit their comments to relatively superficial aspects of rhythmic anticipation or delay (1994:322). Nonetheless, they do raise awareness that models of musical expectancy which disregard the temporal dimension entirely are likely to miss the mark (1994:323).

Emmanuel Bigand and Marion Pineau recognised the opportunity to explore harmonic priming within a global, rather than local context (Bigand & Pineau 1997). Having noted the earlier prime-target stimuli consisting of only two chords (Bharucha & Stoeckig 1986; Tekman & Bharucha 1992), as well as the slightly longer sequences employed in (Schmuckler & Boltz 1994), they nevertheless felt the latter to have been confounded by failing to "... disentangle the potential influence of global and local structures." (Bigand & Pineau 1997:1099) Their remedy was to employ four eight-chord sequences each ending on V → I (a perfect cadence), with each sequence minimally modified to produce an alternate sequence, ending with precisely the same chords, but this time functioning as I → IV (a relatively unexpected ending).[101] They too employed both expectancy ratings and yes/no judgements of appropriateness in two separate experiments, adding a consonance/dissonance judgement task as a third. Since the final two chords did not vary acoustically between alternate presentations, though their context did, these stimuli invited disparate interpretations of the roles of local and global contexts in harmonic priming. Their results strongly supported the dominance of the global effect. Given different contexts, identical prime-target pairs (in the two-chord, adjacent sense)

---

[100] These were evenly distributed amonst: "authentic cadence" (vi-IV-V-I), "half cadence" (iii-vi-ii-V), and "deceptive cadence" (IV-I-V-vi). Small modifications were made to the melodies as necessary. See (Schmuckler & Boltz 1994:316) for details.

[101] All of the sequences employed may be seen in (Bigand & Pineau 1997:1101).

elicited markedly different priming responses, a result which they argued to be at odds with extant models.[102] Engaging the question of whether such facilitation could come about as a result of sensory effects, Bigand & Pineau had earlier noted evidence that harmonic priming occurred at a cognitive level, but cautioned that such a conclusion be moderated by considering: failure to consider the difference in melodic arrangement between the tones of the prime and target; artificiality of the timbres employed in such experiments; and extreme distinctions between levels of harmonic relatedness (Bigand & Pineau 1997:1100). Their analysis, however, found no statistical evidence of global sensory priming, ultimately leading them to concur with the recommendation of (Schmuckler & Boltz 1994) that Bharucha's model be modified to incorporate temporal weighting factors.

Tekman and Bharucha later examined the relative merits of perceptual and cognitive effects on priming, reporting a significant effect in respect of psychoacoustic similarity at stimulus onset asynchronies (SOAs)[103] of 50 ms, while harmonic relatedness dominated at SOAs of 500 ms or more (Tekman & Bharucha 1998). The stimuli in their first experiment were pairs of 12tet major triads separated by either two or four semitones (closely related, and distantly related, respectively), so selected to maximise the perceptual similarity between the more distantly related prime and target. Their task was an in-tune / out-of-tune judgement, with foils having their fifths detuned by 37.5 cents. Noting the overlaps between psychoacoustic similarity and harmonic theory, they acknowledged: "[p]arsimony dictates that if psychoacoustic similarity is sufficient to account for relationships in musical harmony, the postulation of implicit knowledge is unnecessary." (1998:252) They concluded that the effects of psychoacoustic similarity on priming were short-lived, ultimately overshadowed by "conventional relatedness". Their conclusions were also shown to correlate with the predictions of Bharucha's own MUSACT model, a layered self-organising neural net trained to map pitches to chords and chords to keys. This model similarly favoured shared spectral features during the initial activation phases until it had settled down, at which point "conventional relatedness" (in the musical sense) emerged as the most stable configuration. This has come to represent the epitome of the spreading activation hypothesis. A second experiment further reduced the potential effects of spectral overlap by replacing the target at four semitones distance from the prime with one at six semitones distance (an interval known variously as the tritone, diminished fifth or augmented fourth) and produced comparable results. It was noted that the responses to the stimuli in the latter experiment could equally have been modelled in terms of psychoacoustic similarity as by statistically acquired implicit knowledge. However, priming was no longer observed at the shorter 50 ms SOA in this case ("...no crossover as a function of time." (1998:258)). Again, this correlated with the predictions of the MUSACT model, since the lower degree of spectral overlap gave little advantage to either target in the early part of the neural network's training.

---

[102] Specifically, Bharucha's MUSACT model. This interpretation is contradicted by (Tillmann *et al.* 1998:104), and empirically refuted by (Bigand *et al.* 1999:195).

[103] That is to say, the time interval between the onset of the prime and that of the target or foil.

Typically, priming research has discarded the results obtained in respect of the foils. Tekman and Bharucha, however, reported an interesting reversal (also encountered in earlier work) which they attributed to perceptual bias: “Responses were faster to close targets than to distant targets, and there was a tendency in the opposite direction for the foils.” (1998:257) Though little more is made of this in the published literature, we suggest that this might be symptomatic of the strain brought to bear on harmonic perception by the approximations of Equal Temperament, where more distantly related sonorities are increasingly “out-of-tune” relative to any particular key, at least in the sense of Just Intonation.

Barbara Tillman joined Bigand & Pineau in re-examining the relative contributions of local and global contexts to harmonic priming effects, this time employing an experimental design specifically tailored to evaluate their combined influence on expectancy formation (Tillmann *et al.* 1998). Once again the sequences consisted of simple diatonic progressions of eight chords (twenty in all), each ending on a perfect cadence, this time crossed to create four conditions: globally and locally related (GRLR); globally related, locally unrelated (GRLU); globally unrelated, locally related (GULR); and globally unrelated, locally unrelated (GULU). Unrelated conditions in each case were effected by transposing the pertinent chords up by a semitone (p. 102).[104] A secondary goal of this study sought to compare empirically obtained data against that obtained by means of Bharucha's MUSACT model. Further, stimuli were presented at either a slow or fast tempo, under the proposition that such would be seen to influence the relative contributions made. The task was a consonance/dissonance judgement, dissonant targets being effected by adding an slightly softer augmented octave to the root (1998:106). Twenty-four participants represented musicians, along with twenty-four non-musicians. Separate analyses of response accuracy and response times yielded various main and interaction effects, from which Tillmann *et al.* concluded that “... both local and global contexts influence the formation of harmonic expectancy.” (1998:111) Specifically, “[t]he processing of a consonant target chord was faster and more accurate when the target was harmonically related to the immediately preceding chord, regardless of the global context.” (1998:111) In finding evidence that “... one chord is sufficient to generate expectancy that related chords will follow”, they underscored the results of many earlier harmonic priming studies. On the other hand, their results also confirmed the global effects reported by (Bigand & Pineau 1997). Their particular contribution showed how “... global context effects persist even after the intervention of a chord that was strongly unrelated to the target.” (Tillmann *et al.* 1998:111) From this emerges a hierarchical view of expectancy generation in music, a conclusion also reported in (Bigand *et al.* 1999), whereby the most potent facilitation occurs when expectations are met at multiple levels. Further it becomes evident that “... the effects of local and global context in music are unlikely to be independent”, and moreover that effects at various levels are not simply summed (1998:112). As predicted by the spreading activation model, the faster presentation tempo was found to increase the differential between response times in respect of globally related and globally unrelated targets (1998:110).

104By the authors' own admission, a crude strategy (Tillmann *et al.* 1998:102), but nonetheless effective.

Musical expertise was once again found to contribute little to the measured effects (1998:114).

Bigand & Pineau's (1997) enquiry provoked a further, somewhat more refined approach, this time incorporating observations by Schmuckler & Boltz (1994) and once again seeking to address, in a series of three experiments, unresolved issues regarding the relative contributions of various structural levels to the overall effects of priming in music (Bigand *et al.* 1999). First amongst these was the concern that the minor modifications applied to the longer prime chord sequences might have nonetheless rendered those progressions "less musically fluent", even "awkward", and thus introduced a potential confound in respect of reaction times and error rates measured in the "unexpected" conditions. In response, the authors introduced a "self-paced listening method", allowing participants to individually trigger the transition to each chord of the progression, recording the resulting inter-onset-intervals (IOIs) as evidence of any greater processing difficulty encountered by the participants (1999:186). Besides the expected priming effects, analyses of these IOI profiles revealed no main effect of context on IOI values, indicating that "... the chords defining the expected and unexpected conditions were perceived with the same fluency." (1999:188) The second experiment looked more closely at the length of the prime chord sequence, here varying from seven, to four, to two, to only one chord preceding the target. Interrogating Bharucha's earlier claim that the MUSACT model required only one chord from which to generate a precise harmonic expectancy,[105] the authors now set out to establish "... the critical amount of information necessary to provoke the priming effect." (1999:189) Generally, their findings showed the effect of global context to decline and ultimately disappear with diminishing prime length. Though there was some variance between the results of musicians and those of non-musicians, two-chord primes consistently yielded the lowest response times.[106] This was found to be problematic for Bharucha's model, and only partial improvement was effected by modifying the decay parameter (1999:190). The third experiment introduced two further priming conditions to the duo of expected and unexpected sequences. All sequences consisted of two adjoining phrases, themselves consisting of seven chords each, except for the "asymmetrical temporal structure" described below. All four versions of a given sequence ended on the same two chords. As before, the expected condition contextualised these final two chords as a V → I progression (a perfect cadence), while the modifications to the unexpected condition implied a I → IV function for the same chords. A "middle expected" condition introduced an additional contextual layer by way of a temporary modulation to the dominant key, rendering the same chords highly expected at a global level and unexpected at a local level. Finally, an "asymmetrical temporal structure" condition eliminated one chord from the first phrase, this on the hypothesis that attention

---

[105] A claim which we do not find to be made explicitly in the cited reference (Bharucha 1987), but which might nonetheless be reasonably inferred from the discussion of the network's attainment of equilibrium, and response to subsequent activation (1987:17).

[106] We can't tell for sure, since the authors provide only one example of their four chord sequences (Bigand *et al.* 1999:185), but in the light of available evidence it is not altogether surprising that two-chord primes should feature prominently in these results. In the given example, the final three chords in both expected and unexpected conditions sketch out the tonic, dominant and subdominant functions, thus minimally, yet definitively confirming the sense of key.

and expectancy should have, as a result, focused on the second-to-last chord overall, rather than the last. The latter condition acknowledged issues raised somewhat by Schmuckler & Boltz (Schmuckler & Boltz 1994), but more extensively by Jones's theory of dynamic attending (Jones 1987; Jones & Boltz 1989), specifically that "... the expectation about the *what* of the incoming events interacts with the expectation about the *when* of these events." (Bigand *et al.* 1999:191) Amongst the conclusions drawn from the various analyses of these results were: that there was a differentiated response to unexpected and middle-expected conditions, thereby challenging a simple dual view of global and local contexts; that the strength of priming effects depended cumulatively on the number of levels at which expectation was garnered; and that temporal organisation interacted with harmonic structure, despite the absence of a statistically significant main effect of temporal organisation (1999:193–194). Interestingly, the latter effect was most evident amongst musicians, and was furthermore the only effect observed in this experiment not simulated by Bharucha's model. Upon reviewing their findings, Bigand *et al.* considered two frameworks typically invoked to account for the effect of global harmonic context, namely *tonal hierarchies* (e.g. Bharucha's or Krumhansl's) and *event hierarchies* (*a la* Lerdahl & Jackendoff), concluding that the latter might be superfluous to the matter at hand.[107]

Tillman, Bharucha and Bigand extensively evaluated the MUSACT model for its ability to simulate empirical data drawn from studies of a wide range of musical phenomena, including the reaction times and error rates observed in harmonic priming studies, but also extending to: subjective similarity ratings between pairs of chords and the ability to make same/different comparisons between longer sequences of chords; perceived distance between keys, perception of modulation (change of key) and its perceived asymmetry; and perceived stability of tones in a tonal hierarchy (probe tone ratings), including the latter's apparent effects on melodic memory (Tillmann *et al.* 2000). With specific reference to musical expectancy, the authors reviewed earlier single-chord-prime studies by Bharucha & Stoeckig and Tekman & Bharucha, priming by way of longer sequences in Bigand & Pineau, as well as Patel *et al.*'s findings in respect of event related potential (ERP) measurements in response to musical sequences (Patel *et al.* 1998). In all cases, compelling evidence was served up in support of the spreading activation model of musical cognition, and particularly of the pivotal role of implicit learning processes in such. The term "tonal acculturation"[108] was offered as a descriptive label of such passive learning, "... the acquisition of highly complex information without awareness." (Tillmann *et al.* 2000:906).

Tillman *et al.* gave some consideration to the matter of cultural specificity. By their account, any musical

---

[107] "The present study reveals that global context effects on chord processing may results from the activation of tonal hierarchies alone. This means that it may not be imperative to consider a secondary integrative stage of processing." (Bigand *et al.*, 1999:195)

[108] One might be tempted to distinguish here between "acculturation" and "enculturation", the former term having recently adopted a more specific meaning within the context of inter-cultural exchange. Thus, in respect of one's own culture, the term "enculturation" is more apt, while "acculturation" refers to the learning of another (initially foreign) culture's schema. The distinction, however, is not uniformly applied in the literature.

system contains both general (universal) and specific (cultural) constraints. They identified the presence of fixed scales and octave equivalence as general constraints, "... grounded in fundamental processes in neurophysiology: frequency-tuned units in [the] auditory cortex, a layered architecture, plasticity in [the] auditory cortex, and hebbian learning." Further: "[w]ith the added constraint that hebbian learning is restricted to the most active output and its neighbours, self-organization is obligatory and automatic." Specific constraints (cultural), on the other hand, "... pertain to the way in which these discrete units are combined simultaneously or sequentially." Such constraints only emerged as a product of learning, "... guided by bottom-up information only, and there is no external teacher." Furthermore, "... no explicit rules or concepts [are] stored in the model." Thus, their claim was that the model had sufficient generality to model "... different kinds of musical systems" by exposure to appropriate stimuli (Tillmann *et al.* 2000:907). Given the context of our own research, it should not be too hard to spot potential confounds to such a claim. Foremost amongst these, the embedded notions of pitch, chord and key demand scrutiny, as do the relative emphases placed on matters of pitch and rhythm, respectively.[109] The former would seem to require extensive reconfiguration of the model architecture in response to the theoretical tenets of any music not based on the 12tet chromatic pitch scale, triadic harmony and major/minor system of keys, a particularly problematic prospect where such knowledge is not formally codified. The latter, the authors propose, might be addressed by either of two strategies. A single-component "interactive-integration model" might simultaneously encode inputs with respect to both pitch and metre (after (Berger & Gang 1997, 1998) or (Stevens & Wiles 1994)), or the processing of the two domains might proceed independently, pitch being the target of the spreading activation model and rhythm being subjected to parallel analysis by a model such as that of (Large & Kolen 1994), followed by an integration phase. The single-component approach leads on from the work of (Jones & Boltz 1989), specifically the theory of dynamic attending, while a two-component approach is suggested by evidence of independent processing of temporal and non-temporal information in (Peretz & Kolinsky 1993), as well as by neurophysiological reports of double dissociation: specifically of amelodia without arrhythmia and *vice versa* (Tillmann *et al.* 2000:909). Still, the authors are less concerned about the potential impact of such interaction since "... we know of no findings pointing to interactions wherein temporal structure changes qualitatively the hierarchies of tonality established by pitch structures." (2000:909)

A subsequent study investigated the extent to which the schematic expectations implied by harmonic priming could be modified by veridical expectations, thus addressing the "... paradox in the cognitive psychology of music [of] how a familiar piece of music can contain surprises." (Justus & Bharucha 2001) Invoking Fodor's taxonomy of fundamental psychological processes, they characterised the harmonic schemata as a modular input system, generating quick, unconscious expectations independent of competing

---

[109] Nonetheless, Tillman *et al.* do characterise their model as "quasitemporal", rather than "atemporal" (Tillmann *et al.* 2000:909).

information (2001:1001). In a series of three experiments, they first set up specific veridical expectation by presenting a preview of the prime and target major triads to follow (either two semitones apart/close or four semitones apart/distant, both "in-tune"). The subsequent presentation of the same two chords, with the target possibly mistuned, demonstrated the effect of the fulfilment or violation of schematic expectations on the tuning judgement, while consistently satisfying veridical expectations. In their second experiment, they replaced the preview with an accumulative veridical bias toward the "distantly related" condition (75% of the targets were to distant targets), specifically designed to counteract the schematic expectation. The final experiment crossed both conditions by reintroducing a preview (veridical expectation), violated in 25% of the trials. They conclude from the results of various analyses of variance (ANOVAs) that "... the fulfilment and violation of these two types of expectation work independently." (2001:1009) Further: "Musical expectation seems to be automatic, unconscious, fast and informationally encapsulated, ..." (2001:1010) However, they refrain from suggesting adaptationism or domain-specificity.

This very issue (domain specificity) partly motivated a study of harmonic priming effects in scrambled musical sequences (Tillmann & Bigand 2001). Noting evidence from psycholinguistics citing the suppression of semantic priming in scrambled sentences, Tillman & Bigand redeployed the twenty eight-chord sequences of (Bigand & Pineau 1997), together with their "unrelated" versions, further crossing these with a "scrambled" condition. To the latter, their first experiment invoked a simple switching of the order of each pair of consecutive chords, minus the final V → I or I → IV ending (related or unrelated conditions, respectively – producing the order 2-1-4-3-6-5-7-8), while the second experiment permuted chords four by four (i.e. producing the order 4-1-5-2-6-3-7-8). These same sequences were employed in the third experiment, extended by a four-chord excerpt which either did or did not come from the early part of the sequence itself. The goal of the latter was to focus participants' attention on the global context by performing a recognition task, in addition to the typical consonance/dissonance judgement required throughout all three experiments. The results proved to be somewhat counter-intuitive. Though both musicians and nonmusicians readily identified the scrambled sequences as less coherent, perceptual facilitation nonetheless prevailed, as evidenced by consistently lower error rates and response times in respect of highly expected targets, and a bias to perceive related targets as consonant. In particular, no statistically significant main effect on responses was observed in respect of the scrambling of chord sequences. Some interesting interaction effects were noted, such as a significant three-way interaction between Version (normal or scrambled) × Harmonic Context (related or unrelated) × Musical Expertise (musician or nonmusician). Specifically, nonmusicians appeared to be even more sensitive to harmonic context in the scrambled than in the normal conditions, despite the differential of their coherence judgements having been lower than that of musicians (Tillmann & Bigand 2001:1194). Significantly, this particular study further expanded the range of analytical techniques typically employed by introducing ANOVAs on sensitivity (d') and bias (C) parameters, as defined in Signal Detection Theory (Green & Swets 1966). Overall, the results were once again shown to concur with the predictions of the MUSACT model, arguably a vindication of its relatively low dependence on temporal factors.

A more explicit crossing with psycholinguistics examined the extent to which harmonic priming might influence a non-musical task, in this case the identification of sung phonemes (Bigand *et al.* 2001). Citing conflicting evidence supporting both integrated and independent accounts of linguistic and musical processing, Bigand *et al.* set out to see whether the canonical eight-chord sequences employed in previous studies would influence participants' ability to speedily and accurately discriminate between the phonemes /i/ and /u/, as sung by the four voices sounding the final chord. This particular task further had the potential to eliminate much of the observed difference in response time between musicians and nonmusicians. The results largely met with expectations, with responses being faster for related than for unrelated conditions, and with far less difference between the average response times of musicians and nonmusicians (2001:B16). These results were obtained despite participants not being asked to pay any attention to the chord progressions. Nonetheless, such were apparently processed in any case. Bigand *et al.* suggested that their findings supported a more interactive account of linguistic and musical processing than had hitherto been considered.

A further exploration of the extent of harmonic priming effects within musical perception examined participants' ability to detect temporal asynchrony in target chords (Tillmann & Bharucha 2002). For comparison, this study included a parallel consonance/dissonance discrimination task, though the primary focus of the study was a task which introduced foils consisting of chords having their uppermost tone delayed by 50 ms. All stimuli consisted of only two chords (prime-target pairs), forming a perfect cadence in the related condition and a tritone harmonic leap in the unrelated condition. The obligatory 2 × 2 ANOVAs, with harmonic context and target type as within-participant factors, confirmed the effects of harmonic priming on judgements of temporal asynchrony in respect of both error rates and response times, once again uncorrelated with years of musical training. However, the higher sensitivity (d') for unrelated targets seemed inconsistent with earlier results,[110] prompting a side-by-side comparison of synchronous/asynchronous and consonant/dissonant tasks in a within-participants design. The results hereof confirmed that "... d' for the intonation task was higher for related than for unrelated chords...", while "... d' for the asynchrony task was higher for unrelated than for related chords" (Tillmann & Bharucha 2002:645), leading the authors to carefully reconsider their interpretations of spreading activation levels as absolute. Rather, they hypothesised, it might have been the degree of activation change which rendered tone detection more or less difficult (2002:647). By such an account, a delayed target would have already been highly activated in the related condition, relative to the unrelated condition, thus resulting in a smaller activation change. On the other hand, the particular tone employed to effect dissonant targets would have received greater activation

---

[110] Tillmann & Bigand found higher sensitivity (d') for related than for unrelated contexts in a consonance/dissonance task employing a long prime context (Tillmann & Bigand 2001:1188).

from the unrelated condition, and thus subject to a smaller activation change in that case. This interpretation is encouraging, not only for its better fit to the observed data, but for the renewed emphasis it seems to place on the perception of dynamism as a component of musical cognition. Many interesting research questions would seem to arise as a result of such a subtle shift in emphasis. Tillman & Bigand suggest that, for example, "... the question arises whether listeners' sensitivity to accurate timing is smaller for strongly related events than for unrelated ones, or whether a deviant tone would be more easily detected when it is a nondiatonic event – or even a less related diatonic event – in the key context." (Tillmann & Bharucha 2002:648).

Bigand *et al*. re-examined the contributions of sensory and cognitive components to harmonic priming within the longer priming context in two experiments (Bigand *et al.* 2003). In the first, a consonance/dissonance judgement, participants heard a conventional sequence of six chords, followed by a prime-target pair of chords forming the progression I → IV or V → I. The foils were created by transposing the root or fifth of the target a semitone up or down, respectively. The six-chord preludes were then designed to either include or exclude the subdominant (IV) target chord. Once again, these conditions were designed to elicit contrasting results in respect of either sensory or cognitive accounts of the observed priming effects. Significantly, the second experiment introduced different presentation tempi (300, 150, and 75 ms per chord), resulting in twelve experimental conditions. This latter experiment once again served up evidence of sensory components (i.e. spectral overlap) determining the outcome at the highest presentation rate, but only in respect of those participants who had not previously encountered the same chord sequence at a lower tempo. At the slower tempi, the expected effects of schematic expectations returned. Their study also highlighted the relatively insignificant contribution of musical training to the harmonic priming effect. Overall, their conclusions "... provide converging evidence that anticipatory processes intervening during musical listening essentially occur at a cognitive level of representation" (2003:167), yet they do not dismiss the role of sensory components outright. Acknowledging centuries of speculation seeking to bind musical perception to the acoustical structure of musical sound, they suggested that: "... [i]t seems more parsimonious to represent mentally only the final state of this historical evolution of the musical system." (2003:169) Rather, they emphasised the "... relative autonomy of the cognitive component involved in music perception once learning has occurred." (2003:169)

Continuing the earlier interest in cross-domain priming, specifically between music and language (Bigand *et al.* 2001), Poulin-Charronnat *et al.* crossed harmonic and semantic priming in a task "... investigating how musical harmony may potentially interfere with the processing of words in vocal music." (Poulin-Charronnat *et al.* 2005:B67). In the main, this paper sought clarity as regards the compelling, yet contradictory evidence supporting both integrated and independent accounts of musical and linguistic processing (see (2005:B68) for a review). The authors noted, however, a marked preference for melodic stimuli in research favouring an independent model, while those concluding for an integrated model employed harmonic chord sequences. Their task presented participants with 96 sung sentences, consisting of 12 sentences ending on a word semantically consistent with the context of the sentence as a whole, accompanied by either a related or unrelated harmonic context (thus ending on either the tonic or subdominant, for a total of 24 renditions), together with 12 sentences having semantically unrelated endings, again either in a related or unrelated context (another 24 renditions), together with a similar treatment afforded a further 24 sentences, this time ending in non-words (producing the balance of 48 renditions). Thus the data set completely crossed the 12 seed sentences with the conditions of semantic relatedness, harmonic relatedness and lexicality. The task required participants to simply indicate, as quickly and accurately as possible, whether the last word was, in fact, a real word. Results indicated significantly more correct responses for semantically related words, with an enhanced effect in respect of tonic relative to subdominant targets, this effect duly mirrored in the response rates. Further to the conclusion that musical structure, "... processed in an automatic and irrepressible way", interfered with linguistic processing at some level, the authors underscored the relatively subtle musical modifications employed and the complete avoidance of local harmonic priming effects, thereby indicating how both trained and untrained musical listeners must have been automatically processing subtle musical cues at a global level (2005:B73–B74). The authors considered extant evidence supporting dissociated processing in respect of music and language, but noted the reliance on explicit tasks in that research, as opposed to the implicit tasks employed in their own work. Still, this left open questions as to how the effects of semantic and harmonic priming combined, since their data contradicted a simple additive model.[111] Having critically engaged and dismissed various interpretations, they found the most promise in Mari Jones' dynamic attention theory (Jones 1987; Jones & Boltz 1989). By that account, the most referential musical moments (in this case, the tonic endings) served as "attentional markers" and "capture[d] more attentional resources", with the result that "... the amount of attentional resources available for the linguistic processing would be greater on tonic than on subdominant chords, resulting in different sizes of the semantic priming effect." (Poulin-Charronnat *et al.* 2005:B75). In other words, harmonic priming effects were hereby accorded a pre-attentive status, prior to linguistic processing, an account found to be congruent with the results of several ERP studies.[112] They summarise as follows: "... music could modulate semantic priming in vocal music, by modifying the allocation of attentional resource necessary for linguistic computation." (2005:B76)

Increasing interest in cross-domain priming might have reinvigorated interest in the failure of anything like repetition priming to feature in music cognition. Bigand *et al.* considered reports of repetition priming in diverse media, including letters, words (written and spoken), drawings and pictures, environmental

---

[111] Specifically, "[t]he difference between the semantically related and unrelated conditions was larger for targets sung on tonic chords than those sung on subdominant chords." (Poulin-Charronnat *et al.* 2005:B74–B75)

[112] Semantic incongruities have typically been associated with the N400 component, while harmonic priming studies have begun to reveal even earlier effects.

sounds, and even sinusoidal sounds, as evidenced in various experimental paradigms, such as recognition thresholds, perceptual identification, completion of degraded target events, lexical decisions and naming (Bigand *et al.* 2005:1348), and pondered why comparable results had not emerged in respect of musical stimuli. They also considered a view of harmonic priming as more closely related to repetition priming, by virtue of overlapping harmonic spectra, than to semantic priming, though this had previously been challenged by way of stimuli lacking either common tones or overtones (Bharucha & Stoeckig 1986), and moreover by the separation of prime and target by a two second noise mask (Tekman & Bharucha 1998). Bigand *et al.*'s study emerged from a pilot experiment which demonstrated comparable facilitation in respect of both harmonically related and repeated chords, relative to unrelated chord pairs. In that experiment they employed a consonant/dissonant judgement on chord pairs separated by 0, 1, 3, 4, 5 or 6 steps along the cycle of fifths, failing to reveal any significant difference in terms of error rate or response time between separations 0 (repetition) and 1 (harmonically related). Straight-forward repetition differs from harmonic relatedness in that some voice(s) must necessarily "move" in order to effect the necessary change of harmony in the latter case. Thus, in their Experiment 1 (which examined only the repeated and related conditions), an additional condition was introduced in which all voices were required to move to a different pitch, even in the repetition condition. A further modification was introduced by conducting the experiment both with and without an interstimulus interval (ISI) of 50 ms. All trials were separated by a random burst of twelve sine tones so as to eliminate potential traces of previous trials in memory. Analyses revealed no difference in the effect on correct response time in respect of the two forms of priming, while accuracy was generally higher in the related condition. Results were also better in respect of the smoother voice leading and the 50 ms ISI. The contradictions are plainly apparent, since repetition effectively constitutes the smoothest possible voice leading, particularly at zero ISI. Three explanations were considered (bias in the design, dependence on stimulus onset asynchrony (SOA), and forward masking effects), informing the design of Experiment 2. Results consistent with the previous experiment were again obtained: "... chord repetition did not result in facilitation over harmonic relatedness." (Bigand *et al.* 2005:1359). In fact, faster processing (and more accurate processing for nonmusicians) was observed in the related condition than in the repeated condition. Reducing a potential "forward masking" effect (by increasing ISI) similarly failed to influence priming strength, nor did a decrease in SOA favour repetition (as suggested by the "reversal" noted in respect of foils in (Tekman & Bharucha 1998)). Still, the interpretation of harmonic relatedness in terms of chord function (dominant to tonic, rather than tonic to subdominant) could not have been unambiguously inferred from the simple two-chord prime target pairs. Experiment 3 thus employed twenty longer chord sequences, once again crossing harmonic relatedness and repetition, leading to four conditions: non-repeated tonic targets, non-repeated non-tonic targets, repeated non-tonic targets, and repeated tonic targets (Bigand *et al.* 2005:1361). This time it was quite clear that tonic targets were consistently more strongly facilitated, regardless of repetition. In fact, chord repetition slowed responses to tonic targets to a degree, while speeding responses to dominant (repeated non-tonic) in comparison to subdominant (non-repeated non-tonic) targets. However, this observation touches on another matter entirely, namely the question of whether dominant and subdominant musical functions are, in fact, hierarchically equivalent. Were such to be the case (as is suggested by the symmetry employed in numerous music-theoretical accounts), then Bigand *et al.*'s results might suggest the presence of repetition priming. On the contrary, these same results, taken together with other (often incidental) observations,[113] might be symptomatic of a lower hierarchical status for the subdominant function, relative to the dominant function. Bigand *et al.* noted that their results contradicted the predictions of sensory models (such as that of (Parncutt 1989)), which would have afforded greater effect to repetition priming. However, cognitive models (such as that of (Bharucha 1987) and (Lerdahl 2001)) also provided less than perfect fits to the data. In particular, Bharucha's MUSACT model lacked the representation of pitch height, since that model merely reflected pitch chroma, and was thus unable to account for the impact of voice leading observed in Experiment 1. Arguably more problematic was the fact that MUSACT was strongly influenced by the most recent chord, thus predicting the effects of repetition priming in simple two-chord cases. In the case of an expanded musical context: "... the model simulates correctly the effect of musical function, but overestimates the influence of repetition and fails to account for reduced facilitation for repeated tonic over nonrepeated tonic and increased facilitation for repeated nontonic over nonrepeated nontonic." (Bigand *et al.* 2005:1368) Bigand *et al.* therefore considered Fred Lerdahl's Tonal Pitch Space Theory (Lerdahl 2001) in search of a better fit to their data.[114] At first glance, that model seemed to similarly suggest greater facilitation in repeated than in related conditions, since the only TPS distance smaller than the distance between dominant and tonic was the distance between any chord and itself. This was addressed by extending the TPS theory to include a parameter expressing the distance between the final two chords, with positive values expressing increasing tension and negative values expressing resolution. Summing this value with the tonal stability value of the target chord (in terms of Lerdahl's TPS theory) produced values congruent with the observations in Bigand *et al*. What stands out here is the subtle shift in emphasis, once again, from static values obtained in respect of particular chords, to an increased concern for the degree of change (cf. (Tillmann & Bharucha 2002), reviewed earlier). Having thus established compelling evidence of music's incongruity as regards the ubiquity of repetition priming in other domains (including the auditory domains of speech and environmental sounds), Bigand *et al.* suggested that the cause might lie with the balance of syntactic and semantic features characteristic of each domain. If music is wholly defined by syntax, lacking any semblance of semantic meaning (a hotly debated issue throughout history, to be sure), then perhaps it is this very characteristic which determines the enigmatic behaviour observed here. Clearly, this is no more than a provocative statement at this point, and further research on the matter will have to follow. Bigand *et al.* concluded that Western listeners were likely to

---

[113] The matter is explicitly addressed in (Tillmann *et al.* 2008).

[114] Though Lerdahl's model had seldom been referenced in priming research before this point, Bigand had established the potential thereof to contribute in Bigand, E. (2003). Travelling through Lerdahl's Tonal Pitch Theory: A psychological perspective. *Musicae Scientiae*, 7(): pp. 121-140.

strongly expect change rather than stasis, and final resolution at endings. Thus, lacking the necessary expectation for repetition, there was no means to facilitation, and thus no appreciable priming effect.

In a rather broad sweep, Schellenberg *et al.* applied the methods of selected priming studies to "Western" children from three continents (Schellenberg *et al.* 2005). Experiment 1, based on (Bigand *et al.* 2001), required French children aged 11 and 6 (with and without formal musical training, respectively) to discriminate between the sung vowels /i/ and /u/ following a priming sequence of seven chords. Experiment 2, employing Australian 8- and 11-year-olds, adopted two-chord primes preceding a tonic target (specifically, IV → V → I and ♭VI → ♭III → I). In this case, musical training (and absence thereof) was equally represented in each group, providing the means to disentangle the effects of formal and incidental exposure to musical knowledge. Primes were composed of Shepard tones, while targets employed either piano or trumpet timbres in a timbre discrimination task. Experiment 3 largely replicated the methods of (Bigand *et al.* 2003), this time with Canadian 8- and 11-year-olds. In this, a more conservative criterion guided the selection of musically trained and untrained participants in each age group in a bid to highlight any response differential due to such training. The various eight-chord sequences were presented with a piano timbre, with targets modified for a consonance/dissonance judgement by the addition of an augmented fifth or augmented octave as required. All three experiments served up evidence of the usual facilitation (greater accuracy and lower response times) in respect of tonic targets. Furthermore, neither formal training nor age proved to be reliable predictors of any enhancement of priming effects, suggesting a relatively complete implicit knowledge of basic harmonic functions in musically untrained children as young as 6 years of age. Engaging evidence to the contrary,[115] they highlighted the dependence in such contradictory studies on explicit articulation of musical knowledge. Implicit knowledge, on the other hand, "... is often inaccessible to deductive reasoning", "... remembered for longer periods of time", "... relatively insensitive to individual differences in age and IQ", and "... more resistant to cognitive and neurological disorders" (Schellenberg *et al.* 2005:561). Perhaps most tellingly, they reflected on their own results (Experiment 3), wherein a consonance/dissonance task (the most explicitly musical of the three) provided the best (though weak) evidence of the effects of exposure on response times (2005:561). Approaching from a slightly different angle, they engaged the tenets of the "musical universals" debate. Perhaps, they suggested, "... Western harmony is based on processing biases that make it relatively easy to learn." (2005:562) Noting the ubiquity, across various musical cultures, of features appearing to have stemmed from "... processing predispositions and constraints rather than mere coincidence" (2005:562),[116] and in the light of various studies demonstrating such biases in infants (including sensitivity to small-integer frequency ratios), they proposed that "... Western harmony and tonality – although specific to Western music – are not constructed arbitrarily with respect to

---

[115] Schellenberg *et al.* reviewed studies suggesting a protracted learning process, citing Krumhansl & Shepard (1979), Krumhansl & Keil (1982), Sloboda (1985) and others. See (Schellenberg *et al.* 2005:552–553).

[116] These include: musical scales with unequal steps; five to seven discrete pitches to the octave; melodies proceeding by small intervals – all very much in line with the observations in, for instance, (Nettl 2001).

perceptual predispositions." (2005:562) An "interactionist perspective, with contributions from both nature and nurture", they proposed, might best serve a better understanding of the acquisition of musical knowledge (2005:562). From this perspective they challenged Pinker's "auditory cheesecake" view of music as having no adaptive function, pointing to the role of music in social bonding (especially between infants and caregivers), emerging neurophysiological evidence of brain specialisation in respect of music as well as the identification of areas subserving both music and language processing, and broader speculation as regards the origins of music. In sum, Schellenberg *et al.* show quite suggestively how the results of musical priming research might speak to the matters pursued in this research.

Tillmann *et al.* noted the emergence, in semantic priming research, of a distinction between explanatory theoretical frameworks based on either prelexical or postlexical influences (Tillmann *et al.* 2006:344). In other words, while some accounts remained focused on stored semantic knowledge, others began to look to participants' problem-solving strategies under experimental conditions. In particular, the response biases that had been encountered in harmonic priming tasks invoking tuning tasks, such as in (Bharucha & Stoeckig 1986) or (Tekman & Bharucha 1998),[117] might have been most parsimoniously interpreted in terms of congruency effects. By such an interpretation, "related" and "in-tune" would have been perceived as normals, with "unrelated" and "out-of-tune" as deviant. The combination of normal and deviant (e.g. "unrelated" and "in-tune") might have triggered an incongruency effect, resulting in delayed processing, ultimately leading to a confound and mistakenly interpreted as response bias, as evidenced by the lower response times obtained in respect of out-of-tune targets in the unrelated condition. The timbre discrimination task employed in Tillman *et al.* sought to circumvent this possibility on the understanding that "... when the prime context varies on the same dimension as the dimension of the experimental task, the prime influences the target judgement." (Tillmann *et al.* 2006:345) Clearly, Tillmann *et al.* were here led by a firm conviction that tuning and harmony shared a dimension while timbre represented an independent dimension, all axioms that we might have to be considerably more sceptical about in a non-Western context.[118] In their Experiment 1, the materials of (Bigand *et al.* 2003), played on a piano timbre and ending on either tonic or subdominant, were further crossed with either a piano or harp timbre ending. With respect to the response time data, a 2 × 2 ANOVA, with relatedness and timbre as within-participant factors, produced a significant main effect of relatedness and an interaction effect with timbre. This was confirmed by an ANOVA with chord sequences as random variable. "TimbreB" consistently drew shorter response times, only marginally affected by the relatedness condition, and suggesting the influence of processes other than the automatic activation of listeners' tonal knowledge. Experiment 2a addressed this by varying the target between two timbres (electric piano or harp), neither of which were employed by the prime (which

---

[117] An ostensibly similar bias favours major over minor or minor over major, depending on the context. See (Bharucha & Stoeckig 1986:407), mentioned earlier.

[118] Such distinctions are less clear in the light of, for instance, some of William Sethares's enquiries (Sethares 1993, 2005).

remained on piano). Thus it was less likely that participants would interpret the timbre task in "same/different" terms. The results confirmed the authors' suspicions, serving up a significant main effect of relatedness on response time, with a slight effect of timbre which did not reach significance. Experiment 2b extended the design by introducing additional timbral complexity in the prime sequences. Each of the prime chords was now played by two timbres, one forming the attack, another the sustain portion. The targets remained as for Experiment 2a (this time an electric piano or a harpsichord), and those timbres never occurred in the prime sequences. Results once again served up main effects both of relatedness and timbre, but this time with no statistical interaction. Tillman *et al.* suggested that "... faster response times for TimbreB might be linked to the slightly sharper attack of the harpsichord timbre", noting that "[t]his observation shows the sensitivity of response times to acoustic features of target events" (Tillmann *et al.* 2006:352).[119] Moreover, considering the effects of timbre on auditory scene analysis (Bregman 1994), their Experiment 2b illustrated that harmonic priming likely also occurs between different auditory streams, albeit with reduced perceptual facilitation.[120] The generalisation of harmonic priming effects to "spectral processing" is significant on many levels, most notably for its potential to ameliorate potential congruency effects, but requires insightful experimental design in order to deliver on that promise.

The apparent (and surprising) impotence of repetition priming in musical perception, considering its ubiquity in other domains, came under scrutiny once more in a singing task (Hutchins & Palmer 2008), with participants being asked to sing the final note of a short melodic sequence (two to five notes) as soon as possible after hearing it. Control was exercised over contour, melodic range, and rhythm to maintain uniformity, while variance was introduced with respect to the tonal context of the target (tonic or not), as well as whether the target pitch had occurred as the first note of the melody (as prime), and the length of the melody (distance between prime and target, where applicable). A second experiment fixed the length of the melody at five notes, while varying the ordinal location of the prime. A third experiment modified the timbre of the target in order to control for spectral overlap. Their findings showed repetition priming to be active in all of these scenarios, broadly modulated by ensemble singing experience of the participants (though the latter factor did not reach statistical significance). Mean response latency proved to be the better predictor of repetition priming, with all participants clearly separable into fast responders and slow responders. Fast responders tended to show larger benefits of repetition priming than did slow responders. A somewhat unexpected outcome was an effect of the metrical position of the prime, but this was not extensively interrogated, nor were the effects of stimulus rate considered. Once again there was evidence of assisted processing in respect of tonic targets.

---

[119] This particular matter does not seem to have received much attention in the extant research. Considering our relatively less mature understanding of timbre in any quantitative sense (compared to matters of pitch and rhythm, for instance), such observations could well represent the means to important new insights in this regard.

[120] Tillmann *et al.* point to Schoenberg's Klangfarbenmelodie and orchestrations by Marco Stroppa as examples of the application of such techniques.

A more recent investigation of priming in melodic perception employed a timbre discrimination task (Marmel *et al.* 2010). Here twelve two-bar melodies, presented as a piano timbre, were minimally modified to produce variants in closely related tonalities, so determining the tonal context of the final target pitch as either tonic or subdominant. The targets were either a dull or a bright piano timbre. The results showed that, in respect of both musicians and non-musicians, "... tonic target tones were processed more accurately and faster than subdominant target tones." (2010:1020) However, similar results were shown to be produced by Leman's model of auditory short-term memory (Leman 2000), thereby highlighting the potential role of overlapping spectra. A subsequent run of Leman's simulation against pure sine tone stimuli (rather than the spectrally richer piano tones) failed to produce any significant distinction between tonic and subdominant targets, and thus the first experiment was repeated with sine tones, the candidate target timbres being either a pure sine tone or a sine tone plus a single harmonic at twice its frequency. Though this task was found to be rather more difficult by participants (mean percentage of correct responses was but 46.4%), significant evidence was nonetheless found amongst the correct responses to support the priming effects observed in earlier experiments. This is amongst the most recent published evidence supporting a cognitive, rather than a sensory account of tonal expectations. Armed with these results, Marmel *et al* engaged a number of alternative, essentially sensory accounts, including the potential role of virtual pitches, implied harmony, and the possibility of cognitive expectations as merely a "... backup to paramount sensory expectations...", concluding that these accounts were all less likely since "... variability in sensory information suggests that sensory expectations are less reliable as a baseline mechanism for tonal expectations than are cognitive expectations, which do not depend on spectral complexity." (Marmel *et al.* 2010:1025) As regards computational models, they too underscored the relatively better success demonstrated by Bharucha's MUSACT model, particularly as extended according to (Tillmann *et al.* 2000), in accounting for tonal expectations.

Even more recently, typical priming behaviour has been found to manifest as characteristic auditory brainstem responses in both musician and non-musician listeners (Marmel *et al.* 2011). Electrodes places at the top of the head (Cz), earlobes, and forehead recorded participants' involuntary responses to multiple presentations (approximately 4000 of each) of harmonic progressions to related (perfect fifth down/fourth up), unrelated (diminished fifth/augmented fourth/tritone) and repeated chords. Participants watched a subtitled movie (muted) during the experiment, which lasted approximately two hours in each case. After filtering, artefact rejection and averaging, "[s]hort-time Fourier transforms on overlapping 60 ms windows (overlap, 59 ms) were calculated over the 30 – 160 ms portion of the response", and these were band-limited to 30 – 300 Hz before being subjected to 3 × 2 ANOVAs with "...harmonic relationship (related/repeated/unrelated) as a within-participant factor and expertise (musicians/nonmusicians) as a between-participants factor." (2011:506) The results showed no statistically significant difference between the repeated and unrelated conditions, nor any effect of expertise. Harmonic relationship was the only factor that could be causally linked to the measured subcortical responses. These findings augment earlier research indicating divergent brainstem responses to locally consonant and dissonant stimuli, though we will not review more of that line of inquiry here. We simply wish to illustrate how the rich sets of findings arising from response-time measurements have lately been found to correlate well with decidedly more passive neurophysiological measures, once again suggesting the presence of irrepressible, pre-attentive processing in response to musical stimuli.

We will now reflect on the implications of priming research in music for our own study. The robustness of the priming effect in a diverse range of scenarios, in particular its apparently pre-attentive character, recommends its suitability to our investigation of directed motion in music. Within the domain of Western music, harmonic priming has served up evidence of strong expectations for music to "harmonically progress" in particular ways, as evidenced by facilitated processing where such expectations are satisfied. Broadly, these same expectations have been seen to correspond to theories of "directed motion" in Western music theory, encompassing both the global expectation of progression towards the tonic of the primary key, as well as local expectations of progression along the so-called "cycle of fifths", with further evidence suggesting that these are not the only levels implicated, though interactions are apparently non-linear. The evidence is far from a complete account though, with few studies daring to venture meaningfully beyond the conservative constraints of major triads of the tonic, dominant and subdominant, and a marked poverty in terms of rhythmic complexity, the effects of dynamics and timbre, and melodic factors.[121] Studies have tended to treat all factors as polar opposites, with only an occasional suggestion of finer gradations of the priming effect. Also, recent considerations of participants' problem-solving strategies have reemphasised the importance of careful, informed experimental design, suggesting that virtually all results obtained thus far might yet benefit from closer scrutiny.

Significantly, only a single study has been found which forgoes the reliance on Western music theoretical concepts. In the most explicit cases, chord function was invoked directly. A somewhat more subtle invocation of the "cycle of fifths" entered into those experiments seeking to weigh up harmonic priming against repetition priming, but was employed nonetheless. Furthermore, all stimulus materials were constructed from 12tet pitches. Thus there lurks great uncertainty in our conjecture that priming might serve up evidence of assisted processing whilst employing non-Western stimuli. Certainly, such evidence would have far-reaching implications. Unfortunately, we lack the benefit of an appropriate musical theory in terms of which to construct our stimuli. Thus we too (as was Bharucha) will of necessity be led by our intuition in selecting appropriate representational units and signal features. In our case, these intuitions derive from the observations made earlier as regards the Aristotelian roots of Western harmony, particularly the notion of a hierarchy of stability amongst frequency ratios. We intend to adopt such a principle loosely, in the light of

---

[121] Of course, as reported here, all of these matters have received *some* attention, but only in the most superficial ways.

entrainment as the consequence of non-linear resonance, on the supposition that such an approach is most likely to reveal “universal” principles at work.

We therefore take the position that the facilitated processing obtained in respect of musical priming scenarios is evidence of “directed motion”, and that, to the extent that similar measures can be produced within different cultural settings, these will indicate the location, in audio signals, of the physical correlates thereof. In particular, lower response times and error rates in respect of simple perceptual judgements will be judged as evidence of the most expected targets, while relative inhibition may provide evidence of the extent to which deviations from these normals take place.

Appendix C

African Music Theory

Having explored the notion of teleology in tonal music, specifically as a formalism within Western music theory, but also more generally as a percept (as evidenced in the various priming studies reviewed, cf. Appendix B), we will now consider how this might manifest in African music. In the course of reviewing extant methods of understanding such, various confounds will emerge.

To begin with, African musical practise is astonishingly diverse, and has tenaciously rebutted attempts to abstract "universal" principles as regards the construction of musical scales and melodies, principles of harmonic organisation, matters of phrasing, rhythm and metre, instrumentation, and so on. Unanswered questions abound regarding "... the African view of the origin of music, of their distinctions between music and noise, of why music should be made, or of emotions and music." (Merriam 1962:125) In fact, Merriam suggests that this same diversity which so defies generalisation is properly considered a defining characteristic of African music (1962:121,122). Nonetheless, a broad survey of what has been written along these lines does serve up suggestive evidence of some principles which, though not entirely ubiquitous, are apparently strong markers of *African*-ness.[122] What is more, none of these principles are necessarily constrained to Africa, as is seen in various studies considering, for example, the origins of "the Blues" (Oliver 1970; Kubik 1999, 2010a:47), Jazz/Swing and "Boogie Woogie" (Kubik 2010b:50–52), "hot" rhythms (Waterman 1948), or Cuban "clavé" patterns. All of these tend only to further underscore the status of such observed characteristics in music as distinctly "African". In this review, we will be engaging African music at two levels: a broad survey of notable principles in sub-Saharan "traditional" musics; and a more local investigation into salient principles in Xhosa music. Our reason for scoping in this way is largely a matter of convenience, the Xhosa being indigenous to the area in which this study is being conducted, although a broader scope is included in order to support the underlying interest in "universals". Of course, the selected scope in itself does little to ensure such universality, but will serve to filter some of the more specific traits of Xhosa music so as to promote at least somewhat broader applicability.

---

[122] Consider, for example, (Merriam 1959, 1962) and (Ekwueme 1974), who by their titles explicitly endorse such a view. Temperley notes a number of authors who are similarly willing to "... extend their conclusions, to some degree, to traditional sub-Saharan African music in general (Jones, Koetting, Pantaleoni, Chernoff, Kauffman, Arom, Agawu)" (Temperley 2000:66). Von Hornbostel, though acknowledging the distinct identity evoked by any category of world music, warns that "[t]his impression cannot be described in words, and it is necessarily destroyed by analysis." (Hornbostel 1928:38) He nonetheless proposes three outstanding characteristic features of African music: "antiphony [...], part-singing, and highly developed rhythm". (Hornbostel 1928:39)

Ontologically, this study will favour a more "etic" than "emic" approach (Pike 1967:37),[123] though Artur Simon's suggestion of an "ideocultural" standpoint (Kubik 2010b:4) is likely to be the most appropriate label, considering that it is practically impossible to rid ourselves of all our cultural baggage in approaching the music of another culture.[124] While Nettl, on the one hand, believes that only a cultural outsider is capable of mustering sufficient objectivity (Nettl 1964), Kubik warns that: "Characteristically it is part of any culture's enculturation program to cherish the illusion of a universal validity of one's culture-specific thought categories, communication symbols and societal norms. Discovering cultural relativity (cf. Herskovits 1972) is not easy." (Kubik 2010b:86) Therefore, "[e]veryone believes that their own perception of even the most unfamiliar music corresponds with the music's intrinsic structure and meanings." (2010b:107) Much the same critical reflection is already present as early as 1928 in an article which has come in for more than its share of criticism. Von Hornbostel notes "... two contrasting tendencies in our minds both of which falsify the real facts. On the one hand we look at things in our own way and assimilate them to things we know. On the other hand obvious outward differences at first sight strike us so much that we even exaggerate them – overlooking, hereby, essential but less obvious features." (Hornbostel 1928:30)[125] Despite this acknowledgement, it is our stated goal to disconnect from a large part of that baggage by the avoidance of CMN-derived constructs. As Ekwueme puts it: "There is no doubt that the notation employed by classical western music is not completely satisfactory for transcribing African music. It is, in fact known that the staff notation is not even satisfactory for contemporary western music..." (Ekwueme 1974:43).[126] Koetting notes the same in proposing his adaptation of Philip Harland's Time Unit Box System (TUBS) (Koetting 1970:125). Yet, this ideal is hampered somewhat by the predominance of precisely such constructs, however suppressed, in most every piece of writing on African music. What is more, our resolve is likely to attract

---

[123] By a common misinterpretation, etic and emic are equated with "outsider" and "insider" perspectives, respectively. In fact, the distinction has more to do with how units of analysis are arrived at, whether predetermined or discovered in the course of data gathering. Emic approaches tend to produce context-specific results, whereas etic approaches favour universalism, to whatever extent such is possible.

[124] Kubik describes the ideocultural standpoint as that "... in which popular, non-scientific ideas pertaining to the observer's culture become a framework of reference for the analysis of another culture." He lists specific instances of wholesale adoption of terms such as "Hocket-Technique" (Nketia 1962), "hemiola-style" (Brandel 1959), as well as the more generally sustained use of terms such as "organum" (notably by Kirby and Jones), "major", "minor", "melody", "rhythm", and "harmony", each potentially loaded with culturally determined interpretations. Ekwueme adds the bar line to this list (Ekwueme 1974:44), and casts general scepticism over the ability of white researchers in Africa to overcome "... the natural biases of their own cultural upbringing [and/or] unconsciously patronizing attitude toward Africans and things African." (Ekwueme 1974:62).

[125] Von Hornbostel even warns against the dangers of employing "musically gifted" individuals in undertaking the selection and recording of archival materials, lest they be led by cultural bias (Hornbostel 1928:32). Nevertheless, he maintains that African music manifests universal principles, only partially subdued in music which is governed by Western harmonic practise, "... which are the natural outcome of pure melody." (Hornbostel 1928:38) The insightful sensitivity of his earlier statements notwithstanding, ideocultural transference remains apparent in his approach.

[126] An opposing view might be taken by Blacking (1965:15), or seem to be taken by Agawu (1995b:390–393), though closer reading will reveal that, in the latter case, his is not necessarily a call to adopt Western notation, but rather to employ the same notation for all musical analysis. Elsewhere he acknowledges the "... acceptance of certain *a priori* regarding the nature of musical organization", but elects to retain Western music notation "... in order to render the material immediately comprehensible." (Agawu 1986:65)

criticism of promoting the conceptual "othering" of African music, thereby being party to a "... patronizing and pernicious form of conceptual violence." (Agawu 1992:163) To this charge, we can only hope that our results may go some way toward securing our aquittal, or at least pardon.

Given the aims of this research, then, one might reasonably want to engage with a coherent theory of African music in order to find parallel principles at work. Thus, as a convenient (if somewhat arbitrary) point of departure, let us briefly interrogate the notion of "African Music Theory". Firstly, Africa, as a continent, is home to over a billion individuals in fifty-five or more countries,[127] and features some thousands of spoken languages, some spanning multiple cultural groups.[128] The consequent variance in cultural practise is astonishing. The evidence has prompted various scholars to argue for the term African *musics* (rather than African *music* (e.g. Kirby 1932:26,28)), though this postulate does not go unchallenged (Nzewi 1997:31; Agawu 2003:163). Furthermore, the term music (by a Western understanding) translates poorly into the African societal context (Arom *et al.* 2004:10–11; cf. Hansen 1981:709), where singing and instrumental performance are invariantly integrated with dance, social rituals and work (Rycroft 1967:88; Nketia 1974:21–22,27–30).[129] There are furthermore "... strong connections with political, social, and economic life, as well as with religion, dance, drama and folklore, and even with more specialized usages, as an [*sic*] historic device, as a medium of enculturation and education, as a form of social protest … and so forth [...] [M]usic is functional in African culture, from the standpoints of the numbers of people who participate in it, from its intimate connection with other aspects of African life, and possibly from its nonabstractability from its cultural context." (Merriam 1962:123) Finally, the notion of theory, of an inert systematisation of practise bearing no reference to any other domain of knowledge, is seemingly absent in African musical history (Jones 1949:11,18),[130] though this, too, is contested (cf. Ekwueme 1974:36). Thus, the constituted notion of "African Music Theory" apparently fails to stand up. We say *apparently* because, notwithstanding all of the above, various scholars have proposed formalisms to account for particular practises, and a broad reading of

---

[127] Morocco, for instance, is included in this figure although it is not an African Union (AU) member state. Including Somaliland would take the figure to 56. Ongoing territorial disputes could drive the figure either way, depending upon who one consulted.

[128] Agawu argues similarly, though with older figures (Agawu 1995b:384). More specifically, he points us to Mudimbe's (1988) contention that "Africa" is merely a "... construction of European discourse." Joseph shows that even as innocuous a construction as "Zulu music" is, in fact, invalid (Joseph 1983:55). She nonetheless employs the term herself as a convenience, particularly in the light of unrelenting modern trends towards homogenisation (Joseph 1983:56).

[129] As regards the Xhosa of the Lumko district, the interconnectedness of music and dance is embedded in language: "The word for 'to sing' is *uk-ombela*, which means to sing with clapping and body movement." (Dargie 1991:34) In a footnote, Dargie notes that "[i]n the Lumko district people retain traditional uses of words, and the old musicians do not use the school and mission word to sing – uku-cula – except for (stationary) singing in school and church." (Dargie 1991:46 fn. 12)

[130] Or, as Jones states even earlier, and somewhat more explicitly: "... we expect no sort of help from the African with the analysis of cross rhythms. He knows his own drum and how to incorporate it into the whole ensemble, but he has no analytical sense whatever. He cannot say *what* is happening." (Jones 1937:314) Von Hornbostel had already remarked on some evidence of intercultural influence, concluding that "... the Negroes themselves seem to know what is typical of the music of the white race, and alien and inadequate in their own." (Hornbostel 1928:43)

this literature (as follows hereunder) does tend to serve up recurrent themes and thereby to suggest general principles that might be at work at some deeper level.[131] Along these lines, Locke argues for "... the presence in Africa of a shared musical syntax that produces music with the aesthetic quality of everlasting energetic vitality" (Locke 2011:50; cf. Thompson 1984:xiii), and Waterman coins the term "hot rhythm" to describe a similar impression (Waterman 1948). Even Jones notes that "Africa is a large country [*sic*]; tribes living in various parts have widely differing musical practise; yet there is sufficient evidence to warrant the suspicion that underlying these practises there is a common ground of rhythmic structure." (Jones 1937:295)

Of course, the broader the reading (geographically speaking), the more vague such notions become. We will thus declare our interest, for the purposes of this study, as a set of concentric circles.[132] Outermost of these is what is typically termed sub-Saharan,[133] with a more specific interest in Bantu and Khoisan traditions, even more specifically Nguni, and most particularly Xhosa. This gradation will allow us to consider principles which might not be overtly evident within our most specific focus area, but might nevertheless warrant deeper investigation in the light of their strong showing in related areas.

C.1 The Temporal Domain: Rhythm & Metre

Historically, efforts to construct systematic theories about African music have been dominated by studies of rhythm.[134] Von Hornbostel's (1928) remarks[135] are widely cited as the fountainhead of this intellectual tradition,[136] with Jones's "Studies in African Music" (1959a, 1959b) arguably the most substantial early consolidation of the approach. Merriam qualifies the emphasis on rhythm and percussive concepts as subsuming a "... cluster of traits [including] the simultaneous use of two or more metres, the use of hand-clapping as one kind of accompaniment to song, the presence of membranophones and idiophones as

---

[131] Kubik, for instance, finds a highly developed and sophisticated compositional framework evident in Kiganda Xylophone music (Kubik 2010a, pp.250–251)
(Kubik 2010a:250–251)
, a body of knowledge that is evidently lost to modern generations who continue to perform the repertoire without apparently possessing the knowledge to compose new music in the same style (Kubik 2010b, p.108)(Kubik 2010b:108). Locke describes "... the framework of a musical syntax that is well established but largely un-verbalized..." (Locke 2011:49).

[132] Geographically, these are of course neither circles nor concentric, though we hope the reader will indulge us the required metaphoric latitude in the interest of the point being made.

[133] As Merriam points out, the implied inclusion of the "Sudan-Desert" is problematic, and so he proposes the term "south of the Sudan" instead (Merriam 1962:120), though this term does not seem to have significantly displaced the term "sub-Saharan" in the extant literature. We should thus qualify our use of the term "sub-Saharan" as specifically excluding the Sudan.

[134] For a more extensive list of sources to support this contention, consult Agawu (1987:400, 1995b:380–383).

[135] "... we proceed from hearing, they from motion..." (Hornbostel 1928:53) Echoes of the same are to be found in Kirby's comments regarding standardised fingering patterns in bow music in the article "Primitive Music", to be found in pre-1981 editions of Stanley Sadie's "the New Grove Dictionary of Music and Musicians", Jones' "off-hand technique" (1952:36) and in the constant "cycle of fingering operations" found to underlie Mwenda Jean Bosco's *Masanga Improvisations* (Rycroft 1961:82–83).

[136] Actually, Ward's characterisation of African music as dominated by rhythm pre-dates Von Hornbostel's slightly, his "Music in the Gold Coast" article having first appeared in 1927. We might nevertheless distinguish Von Hornbostel's approach as being decidedly more generalist. See reprints in (Ward 1932a, 1932c, 1932b).

outstanding instruments of the orchestra, percussive intonation and attack, and other elements." (Merriam 1959:13) Scholars have questioned the wisdom of this typecasting (Ekwueme 1974:45; Agawu 1995b),[137] yet the fascination with "African rhythm" continues unabated (Jones 1954; Chernoff 1979; Agawu 1995a), quite possibly at the expense of other musical dimensions, and notably to the exclusion of those traditions which have not yielded quite so readily to such a view, Nguni music possibly being a case in point. Merriam notes how the emphasis on drums and drumming, though undeniably "... of great importance in many parts of Africa" (Merriam 1962:121), obfuscates the percussive and rhythmic nature of the music itself, as evidenced in forceful and dynamic vocal attacks, melodic accent, percussive tone qualities, high dynamic levels and a tendency to play all instruments (not just membranophones or idiophones) percussively (Merriam 1959:15).[138] Most recently though, this particular discourse has been preoccupied with the application of set theory and geometric reasoning to rhythmic patterns, thereby seeking to account for coherence and development, particularly in West African dance drumming (Cuthbert 2006; Anku 2007; Toussaint 2010).[139] A number of open issues remain strewn along this historical path: perhaps most glaringly the absence of adequate translations of the English term "rhythm" (Hansen 1981:709; Agawu 1995b:387; Kubik 2010b:5).

Kubik proposes that sub-Saharan African music, generally speaking, may be seen to organise rhythmically along three levels (Kubik 2010b:31–47). Primary amongst these is the "elementary pulse" (loosely, Hood's "density referent" (Hood 1971:114–116), Waterman's "metronome sense" (Waterman 1952:208–209), Koetting's "fastest pulse" (Koetting 1970:127), etc.), a steady stream of temporally equidistant, "unaccented and isomorphous" pulses typically proceeding at a high rate (in the order of 300/s). Ekwueme adds that, in Igbo choral music, "... [t]here may also be a gradual *accelerando* throughout the piece of music in order to increase its rhythmic interest continuously." (Ekwueme 1974:60). Furthermore, some pulses may be physically absent, yet strongly implied by the surrounding auditory context (Koetting 1970:122; Kubik 2010b:32). On the other hand, there are some traditions which are most succinctly accounted for in terms of a double elementary pulse line, rather than merely adopting a reference frame based on the lowest common denominator (Kubik 2010b:34). At a more abstract level lies the "beat", essentially identifiable as the pulses upon which dancers base their movements (Agawu 2006:23). Again, these tend to be evenly spaced, usually coinciding with every third or fourth "elementary pulse", which Locke dubs "ternary time" and "quaternary time", respectively. Indeed, for Locke, the beat is that element in his "metric matrix" which is "... most present to consciousness." (Locke 2011:52) Temperley maintains that

---

[137] Once again, a more extensive list of counter-sources may be found in Agawu (Agawu 1995b:383–387).

[138] "Instead of our emphasis on African expression of rhythm and percussion in terms of drums, drumming, and the various idiophones, then, it would seem much wiser, and more to the point, to recognise that it is African *music*, whether it be vocal or instrumental, accompanied or unaccompanied, which is essentially hard-driving, rhythmic and percussive in its overall effect." (Merriam 1959:15)

[139] Agawu's disapproval is plain enough: "... recent scholarship, motivated by an empiricist illusion, has confined its field of enquiry to what may be termed the mechanical aspects of rhythmic organization." (Agawu 1987:403)

"... other metric relationships [beside duple and triple] – such as quintuple – seem virtually nonexistent." (Temperley 2000:70) At the outermost level lies the "cycle", most commonly a grouping of twelve or perhaps sixteen basic pulses (Kubik 2010b:41), though other cycle lengths are possible (2010b:43). To these three basic levels, Kubik adds a fourth, though only in the case of certain West African cultures, namely the "timeline".[140] Locke proposes the same four levels, attributing specific functions to each. To the "cyclic pattern of the bell" (the timeline), he attributes a function as primary reference defining elapsing time and the length of the cycle, with beats dividing that cycle into an even framework upon which to anchor shorter rhythmic motives and cross-rhythm. Pulses simply "...reinforce precise timing in the polyrhythmic context", while meter results from the division of the cycle length into beats (Locke 1982:243).

The "timeline" element has commanded particular interest due to the tantalizing prospect of exploiting isomorphisms between it and the Western pentatonic and diatonic scales (Pressing 1983), though little more than speculation has emerged to date. Nonetheless, "Pressing's rules […] hold considerable promise for codifying certain aspects of African musical behavior and thereby promoting a cross-cultural understanding." (Agawu 2006:29) Indeed, it is hard to dismiss Pressing's own conviction that "... this commonality must tell us something underlying about perception and the mind." (Pressing 1983:44) Ekwueme puts forward "one dozen principles of African rhythm", which he maintains are, "to a very large extent, … applicable to the music of other African peoples south of the Sahara." (Ekwueme 1974:60–61) Amongst these: "There is invariably a regulative rhythm pattern which acts as a time line, dividing the music into *okele*, and guides all other instruments." (1974:61) Agawu similarly assures us of the "... general consensus that time lines are materially real, widely used, and crucial markers of temporal reference in African ensemble music." (Agawu 2006:3) Though various "timeline patterns" have been identified, particular prominence has nonetheless been given to a set of structurally related West African variants collectively termed the "standard pattern". We would argue that the "timeline" level of rhythmic organisation not be restricted merely to those cultures which employ the "standard pattern", but that it is (as Ekwueme seems to suggest) present in most (if not all) sub-Saharan cultures, albeit often structured according to less provocative and perhaps even less palpable principles.

Reflecting on each of these levels of rhythmic structure, both the elementary pulse and cycle most readily admit to objective analysis (typically by autocorrelation), while the beat is somewhat more elusive, being subject to significant cultural variance (Jones 1934; Kubik 2010b:86).[141] The "timeline", which Jones acknowledges as having first been recognised in the work of Natalie Curtis (Jones 1937:3; Curtis 1920:98), is by all accounts a cultural construct, though the present study is precisely intrigued by the possibility of

---

[140] Agawu (2003) credits the invention of the term "time line" to Nketia (1963:78), mentioning a range of terms having somewhat overlapping meanings, including King's (1960), Locke's (1982), and Pressing's "standard pattern", "bell pattern", "bell rhythm", "guideline", "time keeper", "topos" and Nzewi's (1997) "phrasing referent".

[141] For instance, Kubik presents two different versions of the 7-stroke 12-pulse timeline, each sharing the same pattern of strokes, though differing in the placement of the "beat" (Kubik 2010b:76,82).

revealing universal principles operating consistently in many, if not most of the timeline's manifestations. A further structuring element, as mentioned by Kubik, has received relatively little attention, namely the "tonal-harmonic segmentation of a cycle." (Kubik 2010b:44) We will return to both of these last-mentioned aspects in due course.

A number of polemics have emerged in constructing this model of rhythm in African music. Ongoing debate considers, for instance, the location of the primary temporal reference: albeit the elementary pulse (Koetting 1970:122–123,133; Kubik 2010b:32–33);[142] the beat (Blacking 1967:157–158; Chernoff 1979:48; Locke 1982:221–222; Pressing 1983:5; Agawu 2006:21–23); or the time line (Jones 1959a:53; Pantaleoni 1972:58; Anku 1997:217). Temperley triangulates the apparent majority view, together with Lerdahl and Jackendoff's (1983:73) prescriptions for a "tactus" level within the tempo range of 80-170 beats per minute, as well as Parncutt's (1994) experimental evidence, and concludes that the "beat" is the primary temporal reference, in line with Western norms (Temperley 2000:69–70).[143] Locke briefly reviews the most important counter-views, but ultimately upholds the importance of the "beat" in the light of performance practise (1982:245 fn. 7). Furthermore, there have been some differences in characterising the various levels of rhythmic organisation as "additive" or "divisive".[144] Divisive rhythm is strongly hierarchical, with lower levels being constructed by regular subdivision of the rhythmic units found in levels above, as is canonical to Western music. Additive rhythm, as famously proposed by Jones (1959a:20–21), groups lower level pulses into higher level groupings, where the number of pulses in each component subgroup might be mutually prime. The difference, conceptually, is significant, since it leads to different readings of what is normative and what is novel/deviant. Different interpretations of the time line are also found to "... betray two broad orientations, one qualitative, the other quantitative." (Agawu 2006:4) Qualitative approaches eschew typical Western methods of engaging rhythm, the latter generally rooted in counting of some sort. Instead, they tend to employ, for instance, mnemonic patterns as a form of oral notation, or to leverage language rhythms,[145] or ethical belief systems, or to model themselves on gesture. Quantitative approaches, originating largely in Pressing's (1983) "cognitive isomorphisms", have enjoyed a surge of interest in recent times (e.g. Rahn, Anku), resonating most recently in the work of Toussaint (2003, 2005, 2010). Nonetheless, we are warned

---

[142] Temperley dismisses Koetting's assertion that "[t]he fact that the repetitions of the fastest pulse often group themselves into 'gross' pulses or beats is … incidental" (Koetting 1970, p.122; Temperley 2000, p.69)(Koetting 1970:122; Temperley 2000:69). In Koetting's last paper, though, he concedes that "... [t]he fastest pulse affords the most ambiguous and thus least dogmatic explanation of rhythmic phenomena... But fastest pulse probably is *not* the answer to the question of how African timing is *perceived*." (Koetting & Knight 1986:60)

[143] Of course, Temperley's position is ideoculturally tainted, neither Lerdahl & Jackendoff nor Parncutt having reached their own conclusions in the light of any strong African evidence. However, such conclusions are in line with a more recent mood to reassess the "othering" of African music, in contrast to earlier distinctions drawn by, for example, Waterman (1952:211–212) and Chernoff (1979, pp.40–42,47–54,94–97)
(1979:40–42,47–54,94–97)
.

[144] Kolinski proposes the alternative terms "isometric" and "heterometric", respectively (Kolinski 1973:497).

[145] To the employment of mnemonics and language rhythms, Koetting points out that "... rhythmic timing must be *applied* to the verbal phrase that represents the drum pattern; it is not inherent in it." (Koetting & Knight 1986:62)

that "... the quantitative orientation that facilitates interdimensional – not intersemiotic – comparison is not characteristic of African musical discourse" (Agawu 2006:17), and that its products, for all their purported objectivity, "... have so far not found any corroboration in indigenous African discourse" (Agawu 2006:6).[146] Agawu's fundamentally post-modern stance is to "... deny that any structural feature of African rhythm has an *a priori* validity that excuses it from a cultural test, while also denying that the essential aspects of a cultural view resist structural translation." (Agawu 2006:42) However, each mode of investigation holds its own promise: "Structural analysis explicates possibilities; cultural analysis unveils particularities." (Agawu 2006:26) By the slightest shift of emphasis, we might read herein that any search for universality is by definition bound to be approached from an etic viewpoint, and with the full knowledge that its results merely suggest possible interpretations of the data.

The placement of bar-lines or identification of "starting points" is similarly contentious. For Blacking, "[b]arlines generally mark off the main phrases, and half-bars give some indication of the stresses and the grouping of the notes" (Blacking 1967:35), while Agawu employs his bar-lines for grouping (Agawu 1995a:71,187–188,200). Locke "... inscribe[s] the cadential moment of the ETC[147] on the first onbeat time-point of the first measure", thereby dictating the placement of his bar-lines, and takes issue with the placement of Anku's "regulative time-point" (RTP) (Locke 2011:52).[148] In fact, Locke earlier asserts that "continuous repetition [...] makes the designation of a beginning point somewhat inappropriate" (Locke 1982:225). Kubik advises that "... one must clearly distinguish between motor-accents and the inner reference beat. A bar-line, if anything, should mark the inception of the inner beat, as felt by the performer; it should not follow motor or other accents." (Kubik 2010b:91). To wit, Rycroft published two distinct transcriptions of Mwenda Jean Bosco's "Bombalaka" differing essentially in their placement of bar-lines (Rycroft 1961:95–98, 1962a:96–98; Kubik 2010a:138–139). Kolinski (1973:497), Agawu (1995b:391–392), Temperley (Temperley 2000:71) and others are all critical of Jones's (1959b) "irregular" bar-lines, and Kauffman seems to stand alone in proposing irregular metrical structures (Kauffman 1980:407–412). Still, Agawu maintains that "... it is the regularization of rhythm as meter that ultimately indicates song, [as opposed to] speech." (Agawu 1987:408, cf. 415) In short, there is little, if any, consensus on matters metrical, except perhaps as regards a more modern willingness to accept the presence of a 'tactus' level (Temperley 2000:71), even if, like Arom, one is unwilling to accept any derivatives thereof (Temperley

---

[146] More specifically, Agawu considers the "... cultural relevance of certain structural concepts", and concludes that: "Additive construction, rotation, permutation and isomorphism between pitch and rhythm are problematic when viewed from and African perspective, while variation, embellishment, and broad notions of generation find widespread resonance." (Agawu 2006:41) Perhaps the *coup de grâce* lies somewhere in the observation that, whereas English numerals up to ten employ a single syllable, African languages tend to employ two or more, rendering it highly unlikely that counting could effectively support rhythmic conceptualisation (Agawu 2006:10).

[147] Nzewi's "ensemble thematic cycle" (Nzewi 1997:42).

[148] Locke also takes issue with Anku's assertion that beats "... serve a 'structural' rather than 'metric' purpose." (Locke 2011:53)

2000:93 fn. 6).[149]

A closely related, though somewhat more specific matter addresses the apparent propensity toward "syncopation", whereby African musicians (to Western ears) evidence a marked predilection for placing notes "between" strong beats, rather than on them, as is normative in Western music. Ekwueme, for instance, describes this as follows: "Those pulses which may normally be considered 'weak' beats tend to receive greater instrumental (rhythmic) attacks than so-called 'strong' beats." (Ekwueme 1974:61) Von Hornbostel accounts for this phenomenon in terms of the physiological tension which precedes the sonic event (1928:52), instructing us to "thoroughly change our attitude" and to "place the bar-line before the rest or the up-beat." (1928:53).[150] Blacking summarises Von Hornbostel's position as follows: "Africans think of sounds as bi-products of rhythmical movements, whereas Westerners pay more attention to the sounds than to the movement which causes them." (Blacking 1955:15)[151] Similar sentiments have been expressed by others (Waterman 1952:213; Rycroft 1962b:82–84), though Waterman notes that "[t]he displacement is by no means a random one, however, for the melodic notes not coinciding with the beat are invariably sounded, with great nicety, precisely on one of the points of either a duple or triple division of the beat." (Waterman 1952:213) Ekwueme, on the other hand, attributes such syncopation to "[t]he principle of not allowing the metric units of patterns on different instruments to coincide....", thus dismissing the same as nothing but a style choice (Ekwueme 1974:61). Notwithstanding the range of accounts, Merriam maintains that "... both syncopation and off-beat phrasing are not only characteristic but are among the identifying characteristics of African music south of the Sahara; the device is used nowhere in the world exactly the same way nor with the consistency as it is in Africa." (Merriam 1959:16) Waterman takes this, together with polyrhythm and "jazz phrasing", as central to his notion of "hot" rhythm, a distinguishing characteristic of African-derived musics (Waterman 1948:25,31), though Merriam finds the "hot" concept to be less influential as one moves

---

[149] Temperley reviews various authors' implicit acknowledgements of meter: "Waterman [...] suggested that African music involves a 'metronome sense,' an underlying pulse which is felt but not constantly expressed (1952): this idea has been affirmed by a number of other scholars (Chernoff 1979:49–50,96–98; Locke 1982:245). Jones speaks on a number of occasions of an underlying regular beat, which is often not explicit but is present in the mind of the performer and can easily be supplied if requested (1959a:19,32,38,40). Agawu remarks on the 'secure metronomic framework' underlying complex rhythms of the surface (1995a:110, cf. 189, 193) Nketia expresses a similar view, using the term 'regulative beat' (1963:65, 86–87). Pantaleoni and Koetting also seem to assume some kind of level of beats in African music; however, they differ with the authors cited above as to the nature of this structure..." (Temperley 2000:68–69).

[150] In fact, Von Hornbostel argues, the absence of "syncopation" from any given piece of African music, from a Western point of view, should be regarded as diagnostic of the presence of a "superior division of time" to which that element must properly belong (Hornbostel 1928:53). Waterman takes a similar view of "syncopated" notes as "... tones on the beat of an implied meter at a tempo twice or thrice that of the controlling rhythm." (Waterman 1952:213)

[151] Blacking provides a thought-provoking illustration as regards the *indlamu/ingoma* dance, likely best-known across the world as the high-kicking "Zulu dance": "The *Ndhlamu* stamping dance of the Nguni group in Southern Africa might appear at first to be an obvious contradiction of this: but in fact the tense winding-up of the body is a longer and more significant movement than the stamping release." (Blacking 1955:15–16) We recall what might have been a contrary account being given by South African "crossover" musician Johnny Clegg in the final episode of a television documentary entitled "A Country Imagined", which aired on SABC2 some years ago..

eastwards across the continent (Merriam 1962:120). Rycroft, on the other hand, finds "non-coincidence between words and rhythm" to be one of very few points of similarity between musical practises characteristic of West Africa and those of the "extreme south-east" (Rycroft 1962b:84).

The alternate account argues that Westerners have simply misread the location of the metric accent (Kubik terms this "metrical inversion", and credits its discovery and elucidation to Rycroft (1962a)), and thus the perception of syncopation where there is, in fact, none (Kubik 2010b:96). Ward argues that syncopation is, by definition, only possible against a regularly recurring accent, which he does not find to be present, generally, in African music (Ward 1932b:902). Blacking critiques Von Hornbostel's assertions at length, showing the self-same principles to be at work in the performance of a Western virtuoso pianist (Blacking 1955:15). Dismissing the entire dichotomy, Jones argues against any notion of syncopation (Jones 1954:44), instead favouring an interpretation in terms of the interaction between simple, yet contrasting rhythmic patterns. As he famously puts it: "Rhythm is to the African what harmony is to Europeans..." (Jones 1954:26), and thereby, a "clash of rhythms" is stylistically inevitable. He goes to great lengths to prove that what is elsewhere characterised as being "... syncopated past our comprehension" (Hornbostel 1928:52), is nothing more that "... the combination of […] simple rhythms which make the glorious African rhythmic harmony" (Jones 1934:1–2). In examining Xhosa music, Hansen finds evidence in support of both accounts, though the predominance of one or the other appears to be a reliable marker distinguishing the "Cape Tribes Proper" from the so-called "intrusive clusters" (Hansen 1981:634).[152] She notes that "... off-beat rhythm is characteristic of the music of the northern 'intrusive' clusters (Xesibe and Bhaca) and Mfengu, whereas polyrhythm of multiple metres […] is characteristic of the music of the Xhosa, Thembu, Bomvana and Mpondomise clusters. The Mpondo occupy an intermediary position..." (Hansen 1981:654) Elsewhere she states that "[h]emiola rhythms also occur more frequently in the music of these 'intrusive' clusters" (Hansen 1981:663, cf. 664).[153] Ekwueme also notes the general ubiquity of the crossing of "... duple and triple time, or subdivisions of units of time […] either in vertical or in horizontal combination, or both" in sub-Saharan musics (Ekwueme 1974:60), as does Locke, though the latter also proposes replacing the canonical "three *against* two" by "three *with* two" to better align with his notion of simultaneous, multidimensional temporal perception (Locke 2011:55).[154] Nzewi rejects the "cross-rhythmic"

---

[152] Briefly: the "Cape Tribes Proper", or "older chiefdoms" are regarded as the "original" Xhosa-speaking groupings, with the "intrusive clusters" having been subsequently formed as a result of subsequent waves of southward migration. As such, the "Cape Tribes Proper" are believed to represent older traditions, with the "intrusive clusters" frequently evidencing cultural characteristics apparently imported from further north (e.g. Zulu).

[153] This is interesting to the extent that hemiola resembles polyrhythm on a local scale. Thus, by implication of Hansen's remarks, those traditions which are not characterised by polyrhythm nevertheless seem to employ a limited form of the same. Yet Blacking states that "polyrhythm does not appear to be used extensively by Africans in the Union of South Africa" (Blacking 1955:18–19).

[154] Locke's view finds its more recent precedents in Stone's (1985:139) "mosaic time", and in Avorgbedor's (1987:10) "sound-time maze", though comparable notions are already present as early as 1927: "... whereas any piece of European music has at any one moment one rhythm in command, a piece of African music has always two or three, sometimes as many as four." (Ward 1932c:798)

characterisation somewhat more vehemently, asserting it to be "... antithetical to African social and, therefore, ensemble philosophy. A community/family/team does not work together at cross purposes. This musical structure, which has depth essence, derives from the African philosophy of inter dependence in human relationships." (Nzewi 1997:36) Agawu echoes this sentiment in noting John Collins' insular implication in Peter Bischoff's film "African Cross Rhythm as Seen Through Ghanaian Music" that polytheism, polyglottism and polygamy all draw from the same African source as polymeter, polyrhythm and polyphony, namely, that "Africans generally do things in multiples" (Agawu 2001:192).

Perhaps the entire "syncopation" *vs* "polymetre" debate is a prime example of tinting by the ideocultural lens, and Rycroft considers whether "... we need to revise our terminology" (Rycroft 1962a:100). In fact, Kubik has suggested that the notion of a single metrical structure in a given musical performance might be wholly inappropriate (Kubik 2010b:40). Rather, it may be that individual performers project their own phase-shifted temporal grids onto the sonic whole, viewing their own as primary in each case (Kubik 1965:39, 2010a:277, 2010b:40). For Blacking, "... this polyrhythmic process expresses 'the perfect co-operation of several performers who nevertheless preserve their individuality by maintaining different beats'" (Blacking 1969:18; Hansen 1981:647–648) This would all seem to subtly favour Jones' account rather than Von Hornbostel's, though without necessarily excluding either one. On the whole, such a broader view resonates strongly with more general commentaries on perceived redundancy in African music. What is observed by some to be simple repetition is, by such accounts, actually an intentional and necessary opportunity to perceive the music from multiple perspectives (Kubik 2010a:78; Locke 2011:48).[155] Tracey expresses a similar sentiment with regard to *matepe* mbira music: "... if one is tied down to any one scheme, be it harmonic, metrical or rhythmic, one is missing half the point, which is to appreciate several different conflicting schemes at the same time" (Tracey 1970:42). Locke, too, stresses that "... one who joins the performers in the act of interpreting the polyrhythmic whole will find an ever-changing texture of rich and hypnotic beauty." (Locke 1982:244) Locke enumerates the devices of simultaneous multidimensionality: dualism of tempo, referring to the capacity to experience the musical flow as simultaneously proceeding at different rates, including duple and triple subdivisions, as well as in double-time (augmentation) or cut-time (diminution); meter as matrix, whereby the same phrase may be perceived differently by virtue of its orientation relative to the implicit beat grid; polyphonic perception, referring to the various composite melodies which might emerge in response to a listener's choice of metric orientation; phrase reconfiguration, referring to the wilful shifting of metric orientation by a listener in order to access different permutations of the same phrase; equivocal phrase shape, which invites the listener to "project a temporal shape" onto the incomplete aural image presented; and polysemous phrase shape, wherein a given phrase simultaneously accentuates several beat schemes equally well (Locke 2011:59). Such a meta-perspective would seem to

---

[155] … for which Locke coins the term "simultaneous multidimensionality" (Locke 2011:48,58), with explicit acknowledgement to Kubik's "inherent rhythms", ostensibly a precursor to later formulations of the phenomenon of "inherent patterns".

speak directly to Jones' earlier assertion that "... if we are to solve the problems of African rhythm we must regard it as 'poly-rhythmic,' i.e. a combination of rhythms having their own starting points and their own individuality." (Jones 1934:8) Brandel terms this "the problem of unity in diversity", noting the listener's ability to choose whereupon to focus, whether a single dominant line, or the "Gestalt" of patterns presented (Brandel 1959:113).

Nonetheless, various scholars have sought to identify a coordinating mechanism. For Ward, "[t]he other rhythms may have no possible similarity […] and no connection whatsoever, but on the first beat of the big drum all must coincide." (Ward 1932c:798) Waterman proposes "metronome sense" (Waterman 1952:208–209), an enculturated ability, shared by performers and audience, to impose a steady framework of purely cognitive, temporally equidistant pulses on any given performance (cf. Kubik 2010b:31). In fact, as mentioned earlier, there is compelling evidence to suggest that no single performer has the monopoly in a performance of African music. Certainly, this is the view attributed to Jones (Merriam 1962:127), who implies such by scoring ensembles with individual time signatures and bar lines for each performer, as opposed to the customary shared time signatures and bar lines typically employed in Western music (Jones 1959b passim). Kubik argues that, in noted cases, the individual patterns played by any individual are merely fragments, each only making sense in the way in which it interlocks with other parts (Kubik 2010b:108).[156] The same is subsumed in the term "ensemble thematic cycle" (ETC) coined by Nzewi (1997:42), and is moreover distinguished from polyrhythm in its "... acceptance of melody and harmony as intentional features of African polyphony" (Locke 2011:51). Koetting similarly emphasises the importance of sonority as an aspect of rhythm, also pointing to the "... intimate interrelation of patterns on all component levels...", to the extent that "... members of an ensemble often experience difficulty in performing their pattern alone..." (Koetting 1970:120)[157] This all stands in stark contrast to archetypal Western "common practise" division of musical labour, whereby solo instruments are accompanied by other instruments, these being expected to merely support, but never to challenge the preeminence of the melody. Locke finds widespread evidence linking ETC to "... a sense of rhythmic motion towards a cadence moment of temporary stasis and resolution." (Locke 2011:51) What emerges from such fragmentary performances (in African music) are termed "inherent patterns" by Kubik, these being clearly perceptible melodies which have not been played by any particular player, but have emerged (much like optical illusions) due to the psychoacoustic constraints of human hearing (Kubik 2010a:70–71; cf. Nketia 1974:133–138).[158] In some cases, such as the South African "Reed-Flute Ensembles" described by Kirby (1933), or the Ghanaian variety of the same described by Nketia (1962), it is plain that the listener is expected to constitute single,

---

[156] On the matter of melody, specifically: "*Melody* in Mangwilo is mainly a *resultant* phenomenon. Each part taken by itself does not show much melody, but the two combined usually yield very attractive melodies..." (Kubik 1965:45)

[157] Chernoff reports similarly (Chernoff 1979:53–54).

[158] Tracey notes "... an interesting difference between 'ritual' and 'non-ritual' songs..." of the Sena/Tonga: "mbira parts of the non-ritual songs seem to arise from vocal phrases; [while] vocal phrases in the 'ritual' songs … seem rather to arise *from* the mbira part." (Tracey 1970:38) He, too, employs the term "inherent patterns" in this context.

continuous lines from the disparate parts performed by the instrumentalists. Yet, even where there is only one player (as in performances of *kundi* or *mbira* music), players claim not to play the melodies so much as to allow them to emerge from the interaction of contrasting patterns (Kubik 2010b:123,124,128). Neither is this limited to instrumental music, as Hansen provides "... an excellent example of the polyrhythmic principle applied to voices, with the result that the melodies emerge incidentally from rhythmic conceptualization." (Hansen 1981:647) Here are strong links to a view of music as divination (Kubik 2010b:118–119). One driver behind the formation of such inherent patterns has been studied by Miller & Heise (1950), Bregman & Campbell (1971), Van Noorden (1975) and others as auditory streaming. The other lies in the character of African "tonal languages", whereby melodic patterns are perceived as highly suggestive of particular words or phrases.[159] In both cases, it is not the performer, but rather the listeners (which may include the performer) who impose an interpretive order on the signal received. In the first case, "[p]erception of these patterns is compelling... there may be individual differences in the focus of attention, but nothing like individual constructions of inherent patterns..." (Kubik 2010b:112), and thus this is likely to be a universal percept, while in the second case, it is clearly a cultural construct. The perception, or not, of such patterns is likely to have a significant impact on one's interpretation of the overall musical structure, and clearly implies limits on what may be perceived by a naïve observer or, for that matter, what might be identified by a general computational approach. Given the primary focus of this study, though, we might note that "... [a]lthough the importance of language in Xhosa music should not be underestimated, its influence on the melodies of Xhosa songs is not as strong as has been suggested for other African musical traditions." (Hansen 1981:724)[160]

The corollary to such multifaceted presentation is a fair amount of ambiguity. Overstating any single interpretation is likely to undermine the very effect upon which such music's iridescent quality depends. It is perhaps then less surprising that African musics tend to forgo the obligatory metrical accents of Western music. Instead, the elementary pulses are typically found to be remarkably even with no particular pattern of accent evident. Where accent has been observed, it tends to be quite the opposite to Western norms (e.g.

---

[159] The rhythms of speech may also feature prominently in vocal melodies (Ekwueme 1974:60; Agawu 1987:406), though these are mentioned less often with regard to the formation of inherent patterns than are the pitches employed. Undoubtedly, both characteristics play a role, though their relative contributions remain to be investigated.

[160] Whatever the current state thereof, Kirby nonetheless feels that "... the connection between Bantu speech and Bantu song *must have been* … originally very close, for there are traces still in existence today." (Kirby 1932:26) Von Hornbostel found that speech-tones "... appear to determine the melodic nucleus; but they have no influence on its inborn creative forces [which] direct the further course of the melodic development." (Hornbostel 1928:58) Rycroft explicitly notes adherence to speech-tone requirements in traditional Xhosa songs, as against an apparent disregard thereof in urban Xhosa songs (Rycroft 1959:27–29). Elsewhere is noted the relatively greater compliance to such requirements in bow songs than in choral dance songs amongst the Zulu (Rycroft 1975b:69), though "... on the whole it can be said that speech-tones exert a great deal of influence upon the direction of pitch movement in the melodic line." (Rycroft 1975b:70) A distinction is nonetheless evident between the absolute adherence to speech-tone requirements in Zulu praise-singing as against the more compromised application thereof in true song (Rycroft 1962b:80–81). Broadly, it would appear that "speech-tone emancipation" is a feature of urbanised African styles, and quite possibly an expression of rebellion against tribal conventions (Rycroft 1961:84).

Blacking's reference to *iambic* quantitative division, but without the corresponding agogic accentuation (Blacking 1967:160)). Perhaps an important clue lies in Jones' observation that metric coincidence between the various parts is far more likely to occur at the end of phrases than anywhere else (Jones 1959a:41,84,86,124), as is echoed by Chernoff (1979:56). We take particular note of this phenomenon, which Jones described as African music's "end-directed teleology" (Jones 1959a:49). Indeed, stylistic norms appear to dictate that entry points should never coincide with the supposed "strong beat" (Hansen 1981:624; Agawu 1995a:64,66,110). This phenomenon has attracted much contemplation, including Von Hornbostel's hypothesis that, whereas Western musicians locate the accent in the sounding note, African musicians locate the accent in the physical preparation for the note – the note merely sounding at the point of physical release (Hornbostel 1928:53). Though the evidence which supports such a view (Kubik 2010b:90–91; Hansen 1981:628) is at once novel and thought-provoking,[161] it too has not been without its critics.

C.2 The Spectral Domain: Notes, Scales & Harmony

In all likelihood, the foregoing characterisation of research into systematic principles underlying African music as "dominated by rhythm" is crudely overstated, since it would be an injustice to ignore the tremendous efforts directed at the characterisation of musical scales.[162] Helmholtz's early conclusions (1954:365) did little to discourage the continued examination, over the subsequent century or so, of specific pitches employed in various musical traditions. The data support a view of two distinct drivers, what we might call "bows" and "bars". The latter, referring primarily to the bars of xylophone-type idiophones, but also to the tines of various lamellophones, tend to favour the subdivision of the octave into several approximately equal perceptual increments (Kubik 2010a:170). On the other hand, bow-type chordophones directly exploit the natural harmonic series, resulting in an irregular subdivision of the octave.[163] Furthermore, Kubik suggests that some traditions might exploit the harmonics of only one fundamental (e.g. that of the -Gogo), whereas others would seem to employ two (e.g. -Xhosa) (Kubik 2010a:217).[164] Implementation details vary, of course, with many idiophones (and particularly lamellophones) also evidencing varying degrees of tempering towards harmonic organisation, ostensibly an effort to produce

---

[161] For example: "I witnessed some of the most striking evidence in support of Hornbostel's theory … in Bulawayo … Choirs had to sing set pieces of European composed music … These songs were conducted by the African teachers who coached the choirs: I was astonished to see that several of them gave vigorous up-beats on all the strong beats where I should have given a down-beat … I discussed the matter with some of the African teachers afterwards, and they said that they definitely felt the up-beat to be the strong beat." (Blacking 1955:19–20)

[162] As single examples, consider Ekwueme (1980), Kubik (1985) or Barbour (1972).

[163] Kubik states: "Not all African music is based on tonal systems. And several tonal systems are based on nonsonic principles, for example distance, as seems to be the case in some of the equidistant tonal systems found in certain areas. But many African tonal systems are based on the recognition and selective arrangement of harmonics, sometimes only lower harmonics, sometimes surprisingly high ones, e.g. the 11th harmonic." (Kubik 2010a:217)

[164] In fact, Kubik cites evidence of a -Gogo mouth-resonated vocal technique producing harmonics 4 to 9, thereby suggesting that harmonic-based tonal systems might not be invariantly linked to bows in every tradition (Kubik 2010a:217). In our more specific context, one is here immediately reminded of Dargie's account of *umngqokolo* Xhosa overtone singing technique (Dargie 1991)

harmonic intervals between various notes struck together (or in close temporal proximity). Singing seems to attach more closely to the bow tradition, and there is ample evidence to suggest that vocal melodies parallel bow melodies in their tonal organisation (Rycroft 1971:223; Hansen 1981:668). Certainly, this is the mode more closely associated with Nguni vocal music (Rycroft 1967:88,96–98). Von Hornbostel, it should be noted, would have dismissed all theories rooted in the characteristics of particular musical instruments as un-African since, by his account, all African musical instruments are borrowed (Hornbostel 1928:46,47,48). Rather, he accounts for scales in terms of "natural (psycho-physical) agents" which affect all singers (Hornbostel 1928:35).

Kubik has gone so far as to postulate an uber-bow (which he terms the "!Kung' tonal-harmonic merger model", a hypothetical instrument which accounts for much of the tonal practise found in sub-Saharan Africa, but more specifically including Jones's Nsenga "harmonic cliché" and Tracey's "standard" Shona chord sequence (Kubik 2010a:217–239). In the main, though, African "scales" (if we may refer to them as such) are diverse, and have likely developed in response to somewhat different constraints in different cultural contexts. Nonetheless, the broad categorisations of between four and seven notes to the octave, together with equal or unequal subdivision of that octave remain. Equal subdivision of the octave appears to be favoured in purely instrumental music, while irregular subdivisions (presumably influenced by an awareness of the perceptual implications of the harmonic series) feature more strongly in vocal music, both unaccompanied and instrumentally accompanied, as well as in pure instrumental music having apparent origins in vocal music. Kubik states that: "[w]here string instruments devoid of spectrum-modifying buzzers have played a dominant role, such as … among Khoisan speakers of southern Africa, we are unlikely to observe the development of a tuning temperament." (Kubik 2010a:170) Thus some would find evidence to sustain the claim that "... the music of sub-Saharan Africans is diatonic – that is, uses whole steps and half steps – but may be said to be modal in that the ordering of these whole steps and half steps may not be in keeping with the ordering of the Western European major or minor scale. Scales may be tetratonic, pentatonic, hexatonic, or heptatonic. In many cases, a neutral interval (especially thirds and sevenths) occurs." (Ekwueme 1974:52) The latter-mentioned "neutral intervals" are especially interesting in the light of some of Hansen's observations regarding modern Xhosa songs and church hymns: "I noticed that the 7th and 4th degrees of a heptatonic scale were often altered or even omitted by singers" (Hansen 1981:698). We have long been fascinated by the idiomatic "sound" of modern African choral groups (cf. Rycroft 1991:7), and in particular by their characteristic deviation from Western tuning. We have been simultaneously alarmed at the prospect that such a marker of cultural identity may fade in time, given a widespread popular notion that such deviation is simply due to some inability to "sing in tune".[165] Hansen replies: "In view of

---

[165] We personally recall being party to a conversation wherein a respected choral director made reference to Africans' inability to properly intone semitones. Such notions would appear to underly Kirby's contention that "... [t]he semitone, when it does appear, seems to appear by accident..." As support he cites "... the perennial difficulty experienced by missionaries when they try to teach their converts to sing European hymn tunes." (Kirby

this, it is interesting to note that 'Red' people who have been in little contact with Western music, sing the tritone interval which occurs in their traditional music with ease and accuracy." The tritone interval to which she refers is regarded by Westerners as relatively difficult to sing.[166] Against the broader backdrop of evolving urban styles, Rycroft warns that "... however much the new 'town music' is hailed as an outward sign of progress and success, it is well to remember that it is being paid for heavily out of capital which it may soon not be possible to replace." (Rycroft 1959:29) Indeed, Von Hornbostel had remarked some thirty years earlier "[h]ow dangerous this [Western harmonic] influence is in [African] music as […] can be judged from present conditions among the Zulus, who have hardly preserved any African characteristics even in their melodies." (Hornbostel 1928:42)[167] Taking particular issue with musical developments in south-eastern Africa, he asserts that "... the modern efforts to protect culture are coming too late. As yet we hardly know what African music is." (Hornbostel 1928:60)[168]

The notion of "scale", though not entirely incongruent with musical practise in Africa, does raise its share of issues. Merriam notes one report of sixteen tones to the octave (Ward 1932a:707), and another of African music as being "exclusively pentatonic". He also cites Jones' structural propositions invoking "... a series of conjunct fourths", and that the "... third and seventh degrees are invariably flatted." (Merriam 1959:17). On the other hand, he notes authors including Ward, Tracey and Waterman who subscribe to an essentially diatonic view (Merriam 1959:17). Apart from the absence of a word to signify the concept in any culture studied, and the reticence with which informants gave renditions of what was being asked (Kirby 1932:28), it has been shown that "scales" in African music performance are quite fluid. Aside from the widespread use of rising attacks and falling releases (Merriam 1959:17), particular instances are cited wherein the same sung note, by informants testimonies, might be consistently sounded at different frequencies, as dictated by context (Rycroft 1967:98; Kubik 2010a:193,198–199,394–395). On the other

---

1932:28) Yet Von Hornbostel acknowledges the presence of even narrower steps amongst the *Wedda* (Hornbostel 1928:37), and Rycroft points out that "[a] great number of Zulu songs, including some of their most ancient ones, employ one or more semitone intervals." (Rycroft 1975a:379)

[166] Tovey, too, notes that "... it is interesting that the interval of the tritone […] is a salient feature in both vocal and instrumental music throughout Africa" (Tovey 1935:22). Kubik alludes to possible negative effects of Western musical (mis-) education (of Africans) on African musical practise (by Africans) in a discussion of "metrical inversion" (Kubik 2010b:95–96). Closer to home, a Xhosa music educator (Prof. Sgatya of the University of Fort Hare) asserts that: "By the time the black child reaches the age of five, he is a fully capable musician, the present school method of music soon knocks this potential out of him " (Lucia 1986:197–198).

[167] Rycroft illuminates Von Hornbostel's comments by relating the origins of *makwaya* music to the general repression of indigenous musical practises at mission stations, giving rise to a form of secular venacular song which "... remained fairly close to the familiar mission hymn style." (Rycroft 1991:7) This, Rycroft suggests, is what Von Hornbostel heard.

[168] Ward takes much the same cautionary tone in considering the future of African music in the light of Western influence (Ward 1932b:902). The most recent and more reflective view encountered in the course of this study is once more that of Rycroft: "'Separate Development' along tribal lines […] [f]or educated blacks who had for generations aspired towards citizenship without racial discrimination, [...] was anathema. They strove instead to cherish all the more strongly their Western connections..." More recently: "... the fostering of culture along ethnic lines can easily be a divisive rather than a unifying force. How this impasse can finally be resolved remains to be seen." (Rycroft 1991:6, cf. 1962b:84–85) These word still ring chillingly true today, some two decades on.

hand, the *Ndau mbira* scale has a third note which variably substitutes for both note 3 of the Shona heptatonic scale and note 4 of the Western diatonic scale (Tracey 1972:101). Blacking's study of Venda ocarina music, in particular his measurements of pitches employed on various occasions in performing the "same" notes on the same instruments, leads him to conclude that "... they are concerned with relative patterns of sound rather than exactly-pitched melodies." (Blacking 1959:18–19,21) To wit, Ekwueme asks: "... are accurate measurements of the *results* of a musical performance a true assessment of the *intention* of the musical performer?" (Ekwueme 1974:50) He continues by raising the perceptual character of pitch, as opposed to the physicality of frequency, citing numerous examples of what might be deemed acceptable variance, and directs his reader to pursue the intentions of the performer, rather than the musical manifestation (Ekwueme 1974:51).[169] Von Hornbostel distinguishes "scale" from "mode", arguing the latter's structure, viewed as interlocking tetrachords, to be eminently more suitably to the study of "purely melodic" music all over the world (Hornbostel 1928:36–38). He later notes a typical structure "... consisting of two halves, the first one resting on the upper fifth (dominant), and the second one built analogously on the tonic...", no doubt implying a parallel to modal structure in Western medieval music (Hornbostel 1928:45). Rycroft's analysis of six Tonga and Lala songs is intriguing in much the same way, as "... in all cases the final tone is either the same as, or a 4th lower than the tonal centre." (Rycroft 1954:17, cf. 19) As regards Xhosa music: "The Cape Nguni themselves do not describe the modes they use in their music. They have no special names for the individual notes of the modes, they merely say 'those are the sounds we sing'." (Hansen 1981:683) Also, in some studies of vocal music, slow, systematic (and ostensibly intentional) raising of reference pitch throughout a performance is encountered (Rycroft 1967:101; Dargie 1991:37).[170] This all stands in stark contrast to Western notions of named pitches being tightly bound to particular, standardised frequencies.[171] Furthermore, "scale" flattens pitch space, obfuscating important structural principles. Kirby, for instance, argued for the derivation of scales from a single fundamental (Kirby 1926, 1965), but it has subsequently been shown that every pitch in a given piece might also be more directly related to a specific

---

[169] In summary, Ekwueme's recommendations amount to: incorporating local knowledge of a given piece; comparison of various recordings of the same piece; and gauging performers' reactions to various recordings of a given piece, with additional questioning as required (Ekwueme 1974:51–52). Blacking contrasts the study of Western musical scores against the use of transcriptions in ethnomusicological study, presenting the former as the intentions of the composer, whereas any instance of the latter is but one manifestation of many possible performances. Thus he, too, calls for comparison of multiple recordings of the same piece. "Unless we are specifically studying interpretation, we want to know what a musician sets out to do each time he plays a certain piece of music, not *exactly* what he did on one particular occasion." (Blacking 1959:15) Nonetheless, he challenges Veenstra's distinction between music "to be performed" and music "already in existence" (Veenstra 1958:44), apparently differing on a definition of what it means for music to "exist". Veenstra clearly requires a physical manifestation, whereas Blacking does not.

[170] In a performance of the "Great Eland Song", England notes a constant rise in pitch, "... microtonally, at the rate of approximately one semitone every twelve bars or so..." (England 1967:65), thereafter citing a counter-example (descending pitch) in (Rouget 1961).

[171] Of course, we would be remiss in failing to acknowledge that this is merely a theoretical position. Western music is strewn with examples, in performance, of context-dependent pitch deviation from its own norms. As regards the notion of fixed pitch, such are revealed subtly in studies of *a capella* vocal performance, or more overtly in so-called "blue notes". Similar deviations are studied in the rhythmic domain.

bow fundamental (of which there are most usually two).[172] Analysis of polyphonic textures has further shown that "[t]he unwritten law is that only notes representing harmonics of the same fundamental may be sounded together" (Kubik 2010a:232).[173] In Xhosa music, this determines the selection of *izihlobo* ("friends") tones (Hansen 1981:620,670,686–688), or a set of parallel *iintlobo* variants of the same melody (Dargie 1991:37), though we should immediately acknowledge exceptions encountered in, for instance, Rycroft's study of Zulu *Ugubhu* music (Rycroft 1967:96,97). However, in a canonical Xhosa case, a single instrument is played so as to produce fundamentals a tone apart, say C and D. The fourth, fifth and sixth harmonics evoked (typically by modifying mouth or gourd resonance) would be C, E, G and D, F#, A, respectively. Collating produces the sequence C, D, E, F#, G and A, a so-called hexatonic scale. Now one might observe, for instance, that a G in the vocal line always corresponds with either an E or possibly a C in the bow part, and a D always occurs together with an F# or A. This could lead one to conclude, as Kubik appears to do, that a principle of "single-skipping" is in play, whereby a pitch is paired with the scalar step immediately beyond its neighbour (Kubik 2010a:174). "Several types of multi-part music in east Africa are based on this common technical principle, which is that the harmonizing note for a given note is found in the next note but one of a scale [...] It can almost be considered an 'instrumental technique', not necessarily entailing a preference for certain intervals." (Kubik 2010a:282). Certainly, there are traditions that are best accounted for in this way, presumably those which build more strongly on instrumental music, but a far more succinct explanation simply invokes the requirement of a common fundamental – all the more so when the implicated fundamental is actually present in the bow part. For that matter, Kubik does apply his "!Kung' bow merger" to *silimba* and *kalimba* tunings, arguing persuasively that these show evidence "... linking Bantu musical traditions in the southcentral African belt with the musical heritage of a vanished San population." (Kubik 2010a:233)

Pursuing this line of reasoning may ultimately bring one to consider, as Hansen apparently does, whether the Western preoccupation with melody as identifying characteristic is as appropriate here as is the specific temporal patterning of fundamentals. Von Hornbostel immediately holds a contrary view in asserting this to be a key distinction: that Common Practise Western music "... is built on harmony, all other music on pure melody [which is] not conceived harmonically" (Hornbostel 1928:34). Thus all observed scales, intervals, consonances and polyphony are dismissed as largely coincidental, a by-product of purely melodic processes (Hornbostel 1928:40–41,47). Yet, "... pure melody, too, has 'tonality', its tonality being established by the function and the mutual relations of the notes." (Hornbostel 1928:36) From a decidedly different point of

---

[172] Rycroft coins the term "bi-radical" to describe music built on two fundamental "roots", but furthermore states that "... bi-radical or tri-radical tonality is observable in much, if not all Nguni [choral music]" (Rycroft 1967:97) In the latter regard, he mentions the three fundamentals produced on the *Umakhweyana* (Rycroft 1967:96).

[173] Kubik cites a particularly poignant anecdote related by Marjory Davidson: "One row of keys represents boys, the other represents girls and we choose a boy and girl who go nicely together." He goes on to show how the "smallest girl" is "... undecided and likes two boys equally." (Kubik 2010a, p.234)(Kubik 2010a:234)

view,[174] Kirby considers "Musical Origins in the Light of the Musical Practices of Bushman, Hottentot and Bantu", reiterating Newman's (1919) assertion that "... melody has evolved from harmony" (Kirby 1932:29). Indeed, Hansen goes so far as to suggest that many different "compositions" may be viewed by their performers as being "the same song" by virtue of this very fact (Hansen 1981:698), since they are "contrasting on the surface, but identical in substance" (Hansen 1981:673).[175] The effect, to Western ears, is commonly described as a somewhat irregular alternation between two major triads, most usually spaced a tone or semitone apart.[176] In the former case, as is common amongst the Xhosa, the frequent omission of the "major third" (F#) in the second triad fuels the perception of a pentatonic scale (Rycroft 1966:92–93; Hansen 1981:666), though Rycroft has challenged that perception as regards Nguni music, generally (Rycroft 1975a:379).[177] In fact, the third is similarly enigmatic in Zulu *ughubu* bow songs, where the major third over the higher fundamental is also typically omitted, while that over the lower fundamental is almost arbitrarily sounded as either a major or minor third, apparently at the whim of the performer (Rycroft 1975b:62,67).

As regards melodic contour, Von Hornbostel proposes that "[t]he natural motion of melody is downward […] from tension to rest" (Hornbostel 1928:34), while Rycroft describes the influence of speech-tones on melodic contour as "'sentence intonation' that provides something like a 'carrier wave' of gradually descending pitch […] which is modulated […] by the high, low or falling speech-tones pertaining to individual syllables." (Rycroft 1975b:69) The preference for descending scalar passages is even noted in the guitar improvisations of Mwenda Jean Bosco (Rycroft 1962a:93). Ekwueme essentially concurs with Jones' similar characterisation of African melodic contour in terms of "the teeth of a rip-saw", referring to the steep rises and gradual descents typically encountered (Ekwueme 1974:48). He expands as follows: that "... the steep or sudden rise ... generally marks the beginning of a new musical phrase"; that "... descending intervals are much more frequent than ascending ones, especially intervals of seconds and thirds"; that "... larger intervals, while being less frequent, occur most often in ascending rather than descending motion ... because these leaps generally mark the beginning of a new phrase"; that "... these decrease in frequen[c]y as they increase in magnitude above the fourth, and those larger than a minor sixth are rare"; that "... most African

---

[174] Kirby's (1932) viewpoint is unashamedly evolutionist, in that he argues that this music represents an earlier phase of evolutionary development, thus primarily useful in answering questions about the origins of Western music.

[175] See Figure MH.16 in (Hansen 1981:697) for the melodies of seven different songs, each of which is regarded as essentially equivalent to an older melody, *uNonkala*. Most notable amongst the variants is a theme known as the Chant by Ntsikana, the earliest known transcription of a Xhosa melody (Hansen 1981:696,698). Evidently, the question of what exactly distinguishes a particular piece of music from all others, far from being a uniquely non-Western issue, remains topical to the American media and entertainment industry too (Gherman 2008).

[176] Generally, one might view the spacing of fundamentals approximately a semitone apart as characteristically Swazi and Zulu, and approximately a tone apart as characteristically Xhosa (Kirby 1965:199; Rycroft 1966:91–92, 1967:98; Hansen 1981:667). Minor and major third spacings, or at least approximations thereof, are also encountered amongst the !Kung', together with the tone. A more complete discussion of the latter is to be found in (Kubik 2010a:217–224).

[177] Interestingly, Rycroft compares "jews-harp" performances by a Xhosa and a Venda player, noting the latter's omission of the eleventh partial – which would coincide with the missing "third" mentioned here. The Xhosa player, it seems, was far more willing to employ this note whilst playing the "jews-harp", even if equally reticent to do so on the mouth bow (Rycroft 1966:93).

melodies keep within a range of a tenth, and very many have a span of less than an octave"; and that "... in general most Africans sing one note per syllable" (Ekwueme 1974:48–49).

Ekwueme's systematisation of melodic scales encountered amongst the Igbo of Nigeria is particularly interesting, and is moreover offered as widely characteristic of much African practise (Ekwueme 1980:89). Ekwueme extracts from the particular melody under scrutiny what might be termed a "normal form", such generally having the following properties: minimum intervallic distance between adjacent notes; such intervals actually being employed in the song; as too, intervals between non-consecutive notes; and apical range of scale reflecting the range of song (Ekwueme 1980:93). This inevitably results in a "descending" scale, being either tetratonic, pentatonic, hexatonic or heptatonic. Viewed in this way, one cannot but be struck by the symmetries which emerge. The four-note scale in Figure C.1, for instance, is extracted from a song (1980:93, Ex. 1) which evidences two clear "focal points", those being the E and D marked UF (upper focus) and LF (lower focus). The *ambitus* of the song is found to straddle these two foci symmetrically, with the extrema being termed the upper apex and lower apex, respectively. As Ekwueme goes on to point out, "[t]he pivotal notes are the resting notes in the song. It is on them that the cadences fall. The first two phrases of the song end on the UF, and the last two end on the LF." (1980:94) Thus, Ekwueme suggests, the UF tends to act in the manner of a "dominant", while the LF assumes a "tonic" -like function. Moreover, he notes the two potential approaches to the UF, either a descending minor third, or an ascending major second, and thus attributes cadential significance to the occurrence of specific melodic movements.[178]

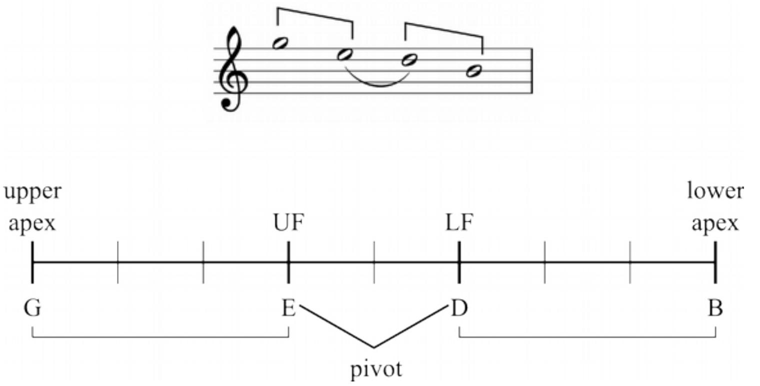


Figure C.1:Pivotal tetratonic with lower focal final (Ekwueme 1980:94)

This account of the "tonal" relationship between the two focal points, here spaced a tone apart, provides an alternative to the view based on bow fundamentals, to which we shall return shortly. An examination of a song employing five notes (1980:99, Ex. 25) produces the configuration seen in Figure C.2. Once more the symmetry is striking, but moreover Ekwueme observes that "... when an even number of notes is employed

---

[178] In fact, there appears to be an error in Ekwueme's discussion, though the intended meaning seems clear enough. Rycroft had earlier drawn a similar conclusion regarding the cadential significance of a falling minor third in an examination of particular bow songs of the Tonga of northern Zimbabwe (Rycroft 1954:18).

in a song, the pivot is formed by the two adjacent notes spanning the smallest interval employed in the song; when an odd number of notes is used, however, the scale is pivoted on one note equidistant in intervallic quantity from the apical notes." (1980:100)

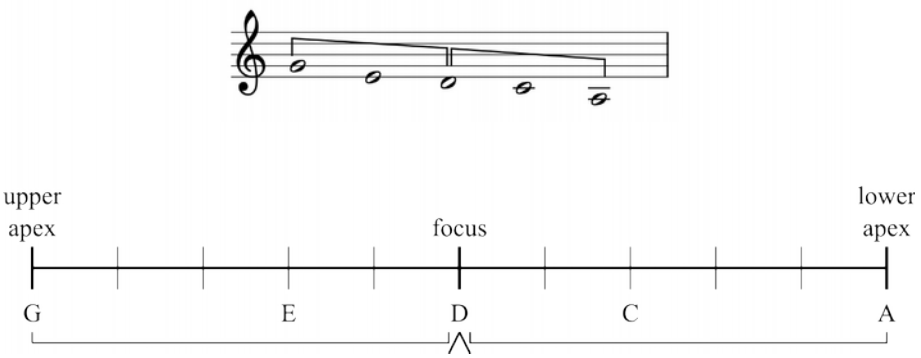


Figure C.2: Pivotal pentatonic with focal final (Ekwueme 1980:100)

This might be a significant insight in as far as it allows one to link scalar structure to typical harmonic movement. In this case, the midpoint of the song cadences on the A (the lower apex), while the D has been reserved for the final. As has already been mentioned above, much Nguni music is deemed to be hexatonic, and Ekwueme provides three examples (cf. Figure C.3 and Figure C.4). The first two of these, (1980:94, Ex. 12 a & b) and (1980:97, Ex. 17), both produce virtually the same schematic (cf. Figure C.3)., though there are interesting details which distinguish these two examples.[179]

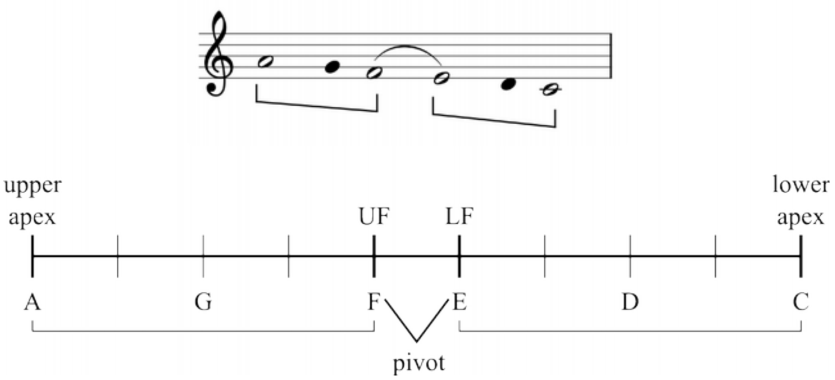


Figure C.3: Pivotal hexatonic (Ekwueme 1980:99)

Ekwueme's third hexatonic example (1980:101, Ex. 28) is especially interesting in that (at a glance, at least) it is at once reminiscent of much Xhosa music which appears to alternate between two tonal centres spaced a whole tone apart. Moreover, this example engages the matter of "neutral 3rds and 7ths". Ekwueme finds that two of the notes in his transcription do not match the pitches intended by Western notation, these being a note midway between A and A♭, as well as one between E and E♭. He thus notates these as "semi-

---

[179] Specfically, the first example demonstrates that the upper focus may be associated with the lower apex, and the lower focus with the upper apex, yielding interlocking fourths. The same example also ends on the lower apex, thus yielding a *pivotal hexatonic with lower apical final*. The second example illustrates a *lower focal final*.

flats", meaning that the notes should be flattened (downward deviation in pitch) by only half of the usual amount. The schematic in Figure C.4 should make this clear enough.[180]

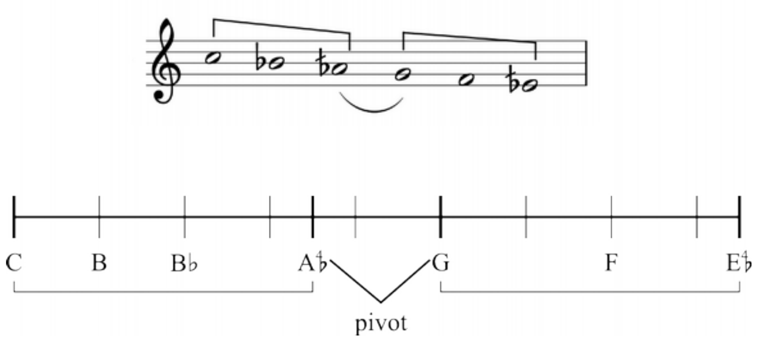


Figure C.4: Pivotal hexatonic with lower mid final (Ekwueme 1980:103)

The net result is that the song alternates, not between two major triads, but between a *perfect triad* (or *neutral triad*) and an *imperfect triad*. The former (F, A𝄳, C), though spanning a perfect fifth, has neither major nor minor quality. The latter (E𝄳, G, B♭) spans an "imperfect" fifth, and is not evenly bisected. Though Ekwueme is not explicit on this point, one might easily imagine that the alternation between the symmetry of the *perfect triad* and the asymmetry of the *imperfect triad* might be instrumental in achieving a sense of teleology in this song. The same example also illustrates the possibility of a *lower mid final*, here the F. One more example will demonstrate the extension of some of these ideas into a heptatonic, possibly diatonic context. The song studied in this case (1980:103, Ex. 35) is understood to be the product of Western influence, and is thus variously performed with or without the "half-flattened" thirds and sevenths. As predicted above, the odd number of notes results in a single focal point, flanked symmetrically by the apices. The final here falls on the lower apex.

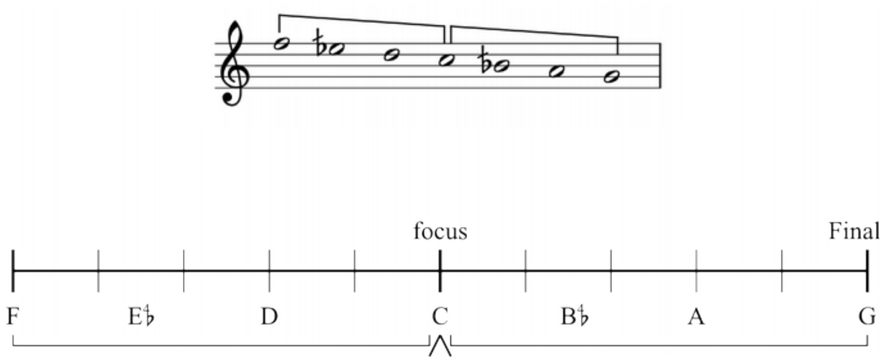


Figure C.5: Pivotal heptatonic scale with lower apical final (Ekwueme 1980:104)

---

[180] Ekwueme's approximation of a neutral third to the ratio of 9 : 11 is handy, though flat by a little more than 3 cents compared to the geometric mean that he presumably intended, considering his assertion that this neutral third "... has the unique character of dividing the perfect 5th with two equal 3rds." (Ekwueme 1980:101)

Ekwueme's emphasis is consistently on recognising and appreciating the symmetry obtained by examining Igbo scales in this manner, such being essentially a musical manifestation of broader themes in all other African art forms. This study, however, is much more interested in how such symmetry might account for perceived tensions and expectations, as might be found, for instance, in the alternation of focal points, or of perfect and imperfect triads. To what extent such an approach is applicable to Nguni (or specifically, Xhosa) music, remains to be seen. Certainly, the claim to common Bantu origins is tenuous, but if such a link does exist, it might well have embedded itself in parallel schemes of musical organisation.

While admitting to the likelihood of some imitation of Western harmonic practise by Africans, Rycroft nevertheless asserts that "... harmonic progression cannot be regarded as wholly unAfrican." (Rycroft 1961:85) Regarding harmonic practise in Ghanaian music, Ekwueme quotes Nketia: "It is not the harmony of block chords built on triads or secondary chords and their inversions, or a system of progression based more on chord relationships rather than on melodic movement. Traditional African harmony has its own logic." (Ekwueme 1974:53)[181] That logic, by Ekwueme's account, is deeply rooted in the nature of tonally inflected languages. If multiple parts are to present the same text, and that text dictates certain melodic inflections, then those parts must employ "parallel, or at least similar, movement" in order to sustain the meaning of that text, a point noted earlier by Kirby (1926, 1932:27), and found to be "... more prevalent among the Swazi and Xhosa than the Zulu." (Rycroft 1967:98)[182] Of course, the lingering (Western) impression, originating to some extent in earlier comments by Von Hornbostel (1928:41) and Kirby (1965:242), is that such parallelism merely echoes the harmonic practises of medieval Europe, a style known as *organum*, and with that impression the concomitant view of a less-developed stage of musical evolution. However, the strict parallelism actually implied by such accounts is challenged in practise, with ample evidence of deviation, ostensibly in service of melodic and scale considerations (Kirby 1932:28–29; Kubik 2010a:174). Where Von Hornbostel notes parallelism in major and minor thirds, being varied in apparently diatonic fashion, he questions whether this "... has grown out of negro music spontaneously or if it is due to European influence." (Hornbostel 1928:42) Merriam, too, notes the prevailing notion that harmony is absent from African music (Waterman 1952:208-209 notwithstanding), suggesting this to be largely due to divergent definitions of harmony. Dismissing both Von Hornbostel's view rooted in polyphony (Hornbostel 1928:41), as well as Ward's more "essentially harmonic" view (Ward 1932a:707, 1932c:797), he proposes instead that we consider "... the simultaneous sounding of more than one pitch by different singers or instrumentalists", a phenomenon which is common enough in African music to be considered characteristic (Merriam 1959:17). It is interesting to note Ekwueme's suggestion that "[i]t appears that triadic harmony is

---

[181] Ekwueme does not give bibliographic details for this quotation. Nonetheless, Nketia's characterisation of Western harmonic practise is here deemed self-evident, and unlikely to attract serious criticism.

[182] Nonetheless, Tracey notes a distinction between the typical parallelism employed north of the Zambezi, and the predominance of "... contrary motion in vocal, mbira and other types of music, such as panpipe ensembles..." to the south (Tracey 1970:39). He later mentions the possibility of oblique motion (Tracey 1970:40).

used only at cadences" (Ekwueme 1974:55), particularly in the light of Rycroft's apparently contrary conclusions about Nguni vocal polyphony that "[u]nison, after chording, may possibly have 'cadential significance' in certain items...." (Rycroft 1967:103). In any event, Ekwueme raises the possibility of a "... system of employing harmony to enhance a rhythmic emphasis or accent" (Ekwueme 1974:56), as well as underscoring the importance of lower and upper pedal tones (which then dispense with the formal requirements of speech-tones), "harmony by imitation", and "harmony by overlapping" (Ekwueme 1974:57–60). Of Nguni vocal polyphony, Rycroft notes that "[d]espite additional chromaticism, and occasional suggestions of major-minor antithesis, indigenous hemitonal root progression still provides the basis for chordal movement, and there seems to be no trace of Western harmonic schemes (Rycroft 1967:100). Countering Von Hornbostel's (1928:40–41) implication that African harmony is a mere product of chance, Ekwueme asserts that "African harmonic principles, nonetheless, suggest a system of organization based on melodic and other considerations which have their roots in the laws of nature." (Ekwueme 1974:60) Such roots are elegantly (albeit speculatively) circumscribed by Kubik's !Kung' tonal-harmonic merger model (see Figure C.6) (Kubik 2010a:217–239), particularly as regards the physical (as opposed to abstract) foundation thereof. What is more, and has particular relevance to this study, is Kubik's assertion that "... the harmonic system in the southcentral African Bantu musical cultures is distinctively different from tonal-harmonic systems in other parts of Africa...", proposing the "merger model" and its degree of fit as evidence of San heritage throughout much of southern Africa, at least as far up as latitude 15° South (Kubik 2010a:214,239,240). While it may seem that this model is far more elaborate than what is required here, particularly in view of Xhosa music's ostensibly "bi-radical" character, it would nonetheless be prudent to keep these principles in mind.

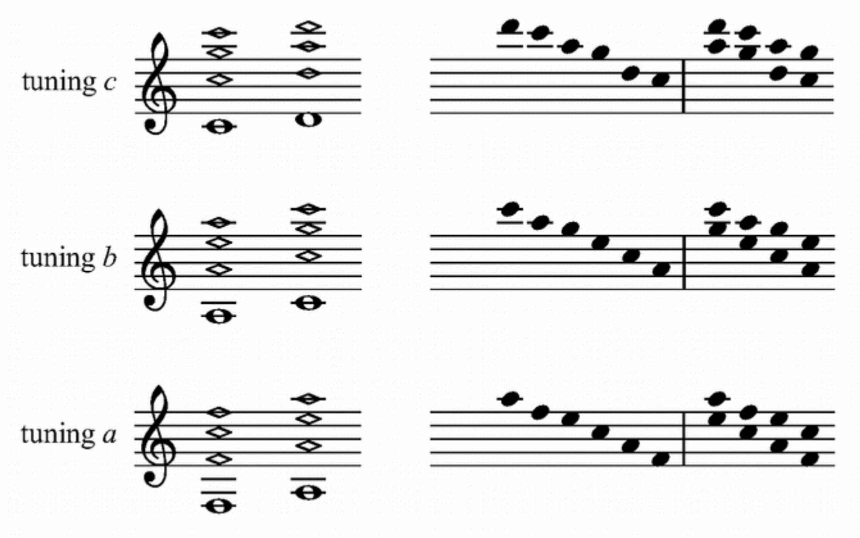


Figure C.6: !Kung' tonal-harmonic merger model (Kubik 2010a:218–219)

The "merger" of all of the proposed tuning phenotypes yields a hexatonic scale which, as Kubik goes on to show, remains implicit within apparently heptatonic music in that such music typically alternates between tonal sections built on distinct hexatonic scales (Kubik 2010a:238). The tetratonic tuning "c" might then have appeared to adequately account for Xhosa musical practise where it not for the missing "thirds" (E and possibly F#). By Hansen's account, these two notes originate in the fifth harmonics over each fundamental (Hansen 1981:667), an option which is explicitly excluded from Kubik's model (Kubik 2010a:218), this despite Kubik's assertion that Hansen's data "seems to suggest" traces of a San heritage (Kubik 2010a:211–212). Instead, Kubik requires (in the context of this model) that a pentatonic be viewed as the merger of any two tunings, and a hexatonic as the merger of all three. However, the scale produced in this way contains a perfect fourth (F), rather than an augmented fourth (F#). It would seem, then, that Kubik's tonal-harmonic merger model requires extension and/or modification before it is to be of use here. A model which does provide a better fit to Hansen's data, without modification, is reported amongst the -Nkhumbi/-Handa group in southwestern Angola, this time including partials up to the fifth (Kubik 1985:44), and Andrew Tracey suggests this to be the appropriate model in the case of many peoples in southern Africa (cf. Fig C.7 below).

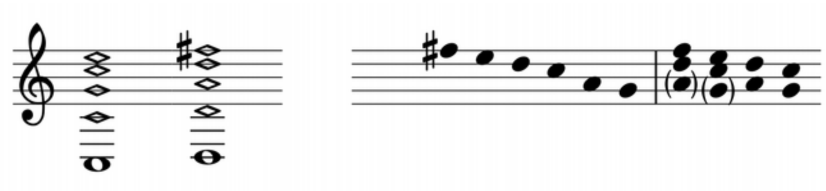



Figure C.7: The tonal-harmonic system of the -Nkhumbi/-Handa (Kubik 1985:45)

On the matter of which harmonics to include, Rycroft raises the point that, given a particular fundamental on a mouth-resonated bow together with the "normal" dimensions of the human mouth, there is a specific range within which partials may be resonated with adequate effect (Rycroft 1966:93,94). Though Rycroft reaches his own conclusions empirically, this approach might yield an interesting analytical approach. Rycroft also points out that we should not be too hasty in assuming Western or even Just tuning for the intervals separating the fundamentals, nor even perfect integer relationships for the partials generated thereby, for in a study of recordings of *ugubhu* songs performed by Princess Magogo, he notes inharmonicity in the flattening of all partials employed, such being in the order of 4 cents or more (Rycroft 1975b:61).

C.3 Macro-Structure: Form

Subsuming matters both rhythmic and melodic/harmonic is the structural level known as "form". The archetypal "call and response" form, predominating in vocal or vocal-related music, sequences antecedent solo and consequent chorus sections of approximately equal length (Ekwueme 1974:46; Merriam 1959:16). Such is essentially characteristic of all African music, though some examples employ significant overlap between these parts,[183] while others maintain a more strictly consecutive separation (simple antiphony, or what Rycroft terms "call and (then) response" (Rycroft 1967:90)). As regards Nguni vocal music, however, Rycroft has identified "the principle of non-simultaneous entry" as fundamental (Rycroft 1967:90,101). Solo "call" and choral "response" form "phrase pairs", a single strophe consisting of "... usually between two and seven pairs", thus constituting strophes of between "two and eighteen measures – either of duple, triple, hemiolic or additive metre." (Rycroft 1967:90) "Call" phrases, Merriam suggests, are often more improvisatory in character, while "response" phrases tend to remain invariant (Merriam 1959:16).[184] Furthermore, "... no definite sense of finality attaches to the completion of a strophe. Immediate recommencement is obligatory, apparently in order to maintain the balance between the parts." (Rycroft 1967:91) Nzewi proposes the term "circular futurity" to characterise the philosophical origins of African musical form (Nzewi 1997:43), and Rycroft elects to employ a circular notation in his analyses (cf. Figure C.8 below), occasionally suggesting that a spiral may be more appropriate in particular cases (Rycroft 1967:93,98).

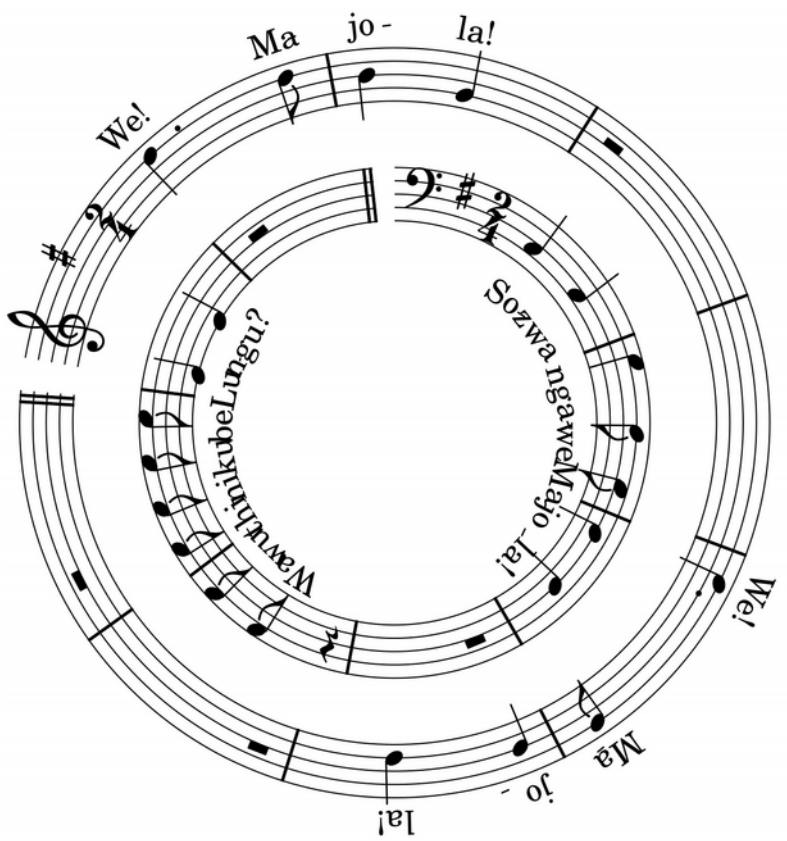



Figure C.8: Rycroft's "circular notation" (1967:91)

---

[183] Von Hornbostel regarded such overlapping as the primitive roots of canon (Hornbostel 1928:45).

[184] Merriam qualifies his use of the term "improvisation" by affirming that all improvisation must function "within controlled limits". It is in this light that he suggests that "... improvisation of text and melody in vocal music are characteristic of African music south of the Sahara." (Merriam 1959:17)

In Xhosa music, form is described in terms of the style of participation. The soloist "leads" (*-hlabela*) while the chorus follows (*-landela*) or "agrees" (*-vuma*) (Dargie 1991:34; cf. Hansen 1981:709, glossary). Hansen has furthermore noted examples of both "single overlap", whereby solo phrases enter shortly before the end of the chorus phrase, as well as "double overlap", whereby solo and chorus phrases both enter approximately halfway through each other (Hansen 1981:618; cf. Dargie 1991:34). Some kinds of music are found to run the gamut from one extreme to the other, as is reportedly common in the *intlombe* ("divination") songs of the "Cape Tribes Proper". As Hansen describes: "After several repetitions of this basic melodic/rhythmic pattern (during which slight variations are introduced in the pulse groups... the music as it were 'goes to pieces', as Francesca Mgudlwa described the Cape Nguni singing style. The different voice parts (including the chorus which subdivides) sing the same and different rhythmic motifs, repeating them ostinato in each repetition of the strophe, with the result that these rhythmic motifs merge into 'lines' of melody, into chordal harmonies, and eventually into one great tonal mass." (Hansen 1981:653) She continues by highlighting the rhythmic nature of the process, and that the accumulation of rhythmic figures to create melodies "... is analogous with hocketting: instead of singers singing isolated tones at specific points in the strophe, they sing rhythmic motifs." (1981:653) Such principles have been attributed more generally to sub-Saharan musics (Ekwueme 1974:61), as too the less common phenomenon of "phrase shift", whereby the soloist progressively advances the entry point until it is nearly in phase with the chorus (Rycroft 1967:93). While simple antiphony presents a rather cleanly delineated form, the kind of development described in Hansen's account of *intlombe* songs is much harder to pick apart. Furthermore, Western analysis of form is typically predicated on at least a clear sense of the start, middle and end of a piece of music, though such ideas are seldom clear in African music. Thus it is hardly surprising when Ekwueme asserts that: "Very few theorists... have attempted to establish a system by which the complete structure of a piece of African music may be determined." (Ekwueme 1974:47)

C.4 In Search of Directed Motion

If we now believe that we have some tentative basis whereupon to consider harmonic "progression", and thus directed motion in African music, it should already be apparent that, in the absence of any consistently operative "functional hierarchy of discord and concord" (Rycroft 1967:101), Riemannian (1896) formulation in terms of chord function has little (if any) relevance.[185] Indeed, Von Hornbostel had already concluded that "... sequences of thirds […] and triads […] do not express or impart a feeling of repose like consonant chords, nor a feeling of tension craving for relaxation like dissonances..." (Hornbostel 1928:49) At best, we might be looking at a back and forth between two chords, with neither necessarily representing repose more

---

[185] However, for a more extensive evaluation of the application of the more salient Schenkerian concepts (e.g. *Urlinie*, *Bassbrechung*, foreground, middleground, and background) to non-Western music, see Stock (1993). An analysis of an Igbo song, with discussion of the merits of the approach, is undertaken in Ekwueme (1975b), though this is fundamentally challenged by Agawu (1986:73–74). Nonetheless, Agawu (1995a) also later draws on Schenker, though "... the transfer can only be accomplished in a restricted sense." (Agawu 2006:30)

than the other. Nonetheless, ideocultural transference doubtless account for the common practise of applying terms such as "tonic" and "counter-tonic" to these two "tonal centres". Neither seems to be universally regarded as more tonally stable than the other, though within a given piece, either might appear to assume that very function (Hansen 1981:405–407,408). Rycroft, while characterising Mwenda Jean Bosco's guitar songs as variously employing "Hepta-Do mode" or "Hepta-Sol mode" or even "Hepta-Fa mode"[186] based on the scale and final "cadence" employed in each case, hastens to add that "[i]t should not be inferred, however, that the scale of values which Bosco may attach to harmonic relationships such as Tonic/Dominant, or Perfect cadence, etc., is necessarily identical with conventional Western feelings towards such phenomena." (Rycroft 1962a:90–92) Tracey justifies his application of the term "tonic" by reporting "... sufficient feeling of repose, return or cadence to justify the use of the term […] as a convenience for analysis only." (Tracey 1970:42) By an alternate account, there might well be an tacitly understood hierarchy of stability, though some pieces would then specifically appear to begin and end in an "unstable" tonality (which is comparatively rare in Western music). What might be more pertinent is the simple fact of change, that movement arises from the pattern of tonal alternation, but towards what goal?[187] The problem is obfuscated somewhat in heptatonic traditions, where it might be possible to project Western diatonic thinking without due regard to the pitfalls which lie in wait. Kubik alludes to such misreading in his comments dismissing the all-too-prevalent view of Kwela as hybrid (Kubik 2010b:45–47,95). It would thus seem prudent to dispense with any narrow interpretation of "chord function" here. Still, Hansen's informants offer the terms "*ekhaya*" (meaning at home) and "-*nene*" (meaning to the right) to characterise the two fundamentals employed (Hansen 1981:683), strongly suggestive that one is regarded as base, the other as antithesis.[188] Intriguing, at the very least, is the implied spatial characterisation which seems to suggest a sense of physical transfer between the one tonality and the other. Certainly, such "shifting tonality" vies for the status of "universal" in Bantu music (Blacking 1959:23; Hansen 1981:679), though Merriam immediately notes our scant understanding of "... intersense modalities in which concepts from one sense are are transferred to the description of another." (Merriam 1962:125)[189]

---

[186] There appears to be an error in the article, but the context clearly dictates what is meant here (Rycroft 1962a:91).

[187] Rycroft suggest that "... the two fundamentals […] serve as 'roots', and their alternation provides a functional equivalent to 'harmonic progression' in Western music." (Rycroft 1975b:62) Thus, "[t]he artistic intention would seem to be that of maintaining an ever-changing balance between all the musical constituents – through temporal, chordal and root contrast, in addition to other features of their relationship." (Rycroft 1967:103)

[188] Rycroft employs the slightly less suggestive terms "principal root" and "subsidiary", justifying such on the basis of their frequency of occurrence. He also points out that either tone may assume either function (Rycroft 1967:96). Nketia similarly asserts that "...[t]he functional relationship of the various notes of a scale can be varied. Just as each note may be used as a final or ending tone, so may each note of the scale be used in specific contexts as a prefinal tone... Because the function of the constituent notes of songs is variable, there is no single pattern of any given scale, but several patterns." (Nketia 1974, p.159)(Nketia 1974:159)

[189] "Thus in our own culture, we speak of music as 'hot' or 'cool,' 'blue,' 'rough,' 'sharp,' 'high,' 'shallow,' 'sweet,' and use various other terms which in actuality refer to the senses of sight and touch. If we turn to Africa, we find, for example, that the Ashanti refer to music as 'hard,' that the Shi refer to what we call a high tone as a small or weak tone, and to what we call a low tone, as a big or strong tone, and that the Basongye make virtually no such transfers." (Merriam 1962:125)

What then constitutes the goal towards which African music might be directed? Western tonal music ultimately reveres the establishment of tonality, with sub-goals dictating how this is best achieved, but apparently this is not at issue here. However, as found in the earlier review of directed motion in Western music, a number of studies focused on tension-relaxation, implication-realisation, expectation-resolution, or some such variant in order to reveal progression toward the goal. As shown previously, such phenomena have been found to correlate particular music theoretical constructs to empirically measurable percepts by way of priming studies. This leads me to consider whether the teleology of non-Western music might similarly be revealed by focusing on evidence of tension or expectation, and subsequent testing, by way of priming experiments, of hypothetical goals suggested thereby. Damocles' sword looms large, though, in that all evidence found in the literature study is immediately subject to ideocultural tainting by the respective researchers. Certainly, no substantially "emic" accounts were uncovered which might confirm whether tension or expectation feature to any extent at all in African musical perception. Thus, the primary empirical data to be gathered in this study might confirm or refute a significant portion of what has already been written in this regard. In the review which follows, we consider various authors' implicit or explicit acknowledgement of the presence and role of tension in African music.

In the earlier, more general reviews of directed motion, tension and expectation, it will have been demonstrated that such notions are deemed by many to be central to the musical experience (cf. Chapter 2 & 3, Appendix B). However, it will also have become clear that the terms employed admit to wide-ranging interpretations. For most authors, these phenomena seem to be self-evident, with only a handful submitting to any form of empirical validation. Brandel reports "[d]istinct opposition of lines, or counter-rhythm" in certain *Watutsi* drum music, "... the total result [of which] is a complex pull in two directions." (Brandel 1959:113) Waterman proposes to unpack the mechanism at work here: "If the objective beat occurs ahead of time, the auditor, unprepared for it, perceives it and assigns to it the additional importance always accorded the unexpected […] If the objective beat is delayed, the period of suspense between subjective and objective beats likewise increases the auditor's awareness of the rhythm. When the objective, audible beat occurs halfway between two subjective pulsations, as is frequently the case, both mechanisms operate to give the off-beat tone heightened significance." (Waterman 1952:213) Rycroft comments on the "'off-beat phrasing' between voices and percussion" occurring in alternate bars of a Tonga girls' pounding sound, where "polyrhythmic opposition […] resolves into rhythmic concord or agreement" (Rycroft 1954:22) Agawu would appear to report "tension" in African music wherever he encounters contradiction, albeit between agogically accented speech-tone and meter (1995a:64), between background regularity and foreground irregularity (1995a:110), or in contrametrical grouping of verbal phrases (1995a:67). Presumably then, it is a perceived contradiction that leads him to find tension in a "... syncopated first bar (with a short-long-short pattern)" being "resolved" in the next bar (1995a:68), or in "... a typical 'linear cross-rhythm' or hemiola, an alternation of sub-groups of 3 and 2 [...], phrased in such a way that each second bar resolves the tension generated in the first. " (1995a:192) On tension of the latter sort, though, he introduces a further mechanism: "In a sense, the triplet group has greater mobility than the duple; it moves towards a goal, creating the sense of tension leading to resolution." (1995a:134) On occasion, Agawu finds his tension in the alternation of pitched percussion, "... whereby a progression from high to low connotes movement from tension to resolution." (1995a:109), or even more explicitly in a particular water drum performance wherein "... the tuning in fourths confers a tonal accent on the higher of the two pitches. […] We have interpreted this as a progression from tension to resolution." (Agawu 1986:71)[190] Perhaps the most extreme kind of contradiction, for Agawu, lies in merely omitting the nominal downbeat entirely from the performance (Agawu 1986:79),[191] evincing what Kubik describes as "... a black hole... i.e. a point where energy is swallowed up so completely that its material manifestation disappears." (Kubik 2010b:101) However, all of the foregoing manifestations of "tension" remain tightly bound up in culture-specific prescriptions of what is deemed normative, and what is thereby deemed contradictory. With such an axiomatic point of departure, one might remain sceptical of the extent to which Lerdahl and Jackendoff's GTTM rules are able to fit the data, as Temperley goes to great lengths to show (Temperley 2000 passim).[192] If tension is to be revealed by signal processing alone, then it must yield to a computational conceptualisation.[193]

Kubik, in exploring the role of auditory perception in African music, proposes that "[n]umerical patterned structures are probably innate in our human perceptual apparatus; that they are projected on sonic stimuli is universal. What is culturally determined is the manner in which these inner reference structures are hooked onto the auditory stimuli – where the listener believes to recognize particular bearings in the mass of an auditory perceptual unit. […] Relativistic pivot points also determine the cross-cultural perception of tonal-harmonic structures." (Kubik 2010b:86–87) Of course, merely "projecting" "numerical patterned structures" is likely to result in far more misses than hits, but Kubik clarifies this later in observing that the "... ability to perceive [inherent patterns] is universal and closely linked to the neurological characteristics of the human perceptual apparatus. Perception probably operates from a neurologically determined system of adapter points which form fields defined by numerical relations... They seem to function like a scanning pattern which is projected over the external stimuli and, wherever possible, brought into congruence with them. Occasionally, our apparatus 'bends' the input to conform. The moment congruence occurs between a

---

[190] However, when Agawu goes on to describe a rhythmic modulation in terms of an instantaneous phase shift in the prevailing meter, it might appear that he has been ideoculturally deceived by a water drum tuned in fourths! (Agawu 1986:71) Certainly, his evidence in this regard seems disproportionately speculative (cf. Agawu 1986:81 fn. 7).

[191] For Agawu, this is "... the essential feature of a piece which thrives on the tension between externalized and internalized downbeats." (Agawu 1986:79)

[192] To the anticipated criticism that the application of Lerdahl & Jackendoff's GTTM to African music would amount to an "etic" imposition, Temperley argues that this is equally true of its application to Western music: "Indeed, the whole premise of music cognition […] is that there are processes and structures involved in cognition to which we do not have direct introspective access, but whose reality can be established by other means." (Temperley 2000:91).

[193] Or, to put it somewhat more broadly: "The scientific investigation of African music […] should provide *mensural* results." (Jones 1937:302) This poses no problem to Jones, who asserts that "... wherever there is an African song, there is also some regular *and mensural* rhythm on which it is based; further it is an *objective* rhythm and is produced by mechanical means and is therefore able to be measured mechanically." (1937:303)

structured inner field and a set of external stimuli, we perceive a gestalt. If congruence cannot be achieved, we experience a chaos of (discarded or suppressed) stimuli. Most external stimuli are unstructured and chaotic." (Kubik 2010b:114) Though a fairly high level of abstraction persists in this description, one can hardly miss the potential to link this to discussions of entrainment encountered earlier, and thereby to a plausible physiological mechanism to account for the same. In particular, the reference to an "apparatus" which "'bends' the input to conform" suggests the presence of a potential field, and thus of a 'tension' which drives perception toward a particular goal. The application of this idea to Kubik's "inherent patterns" would be more complex than what is immediately required here, but the simple acknowledgement of such a tempering mechanism in the perception of African music serves us well.

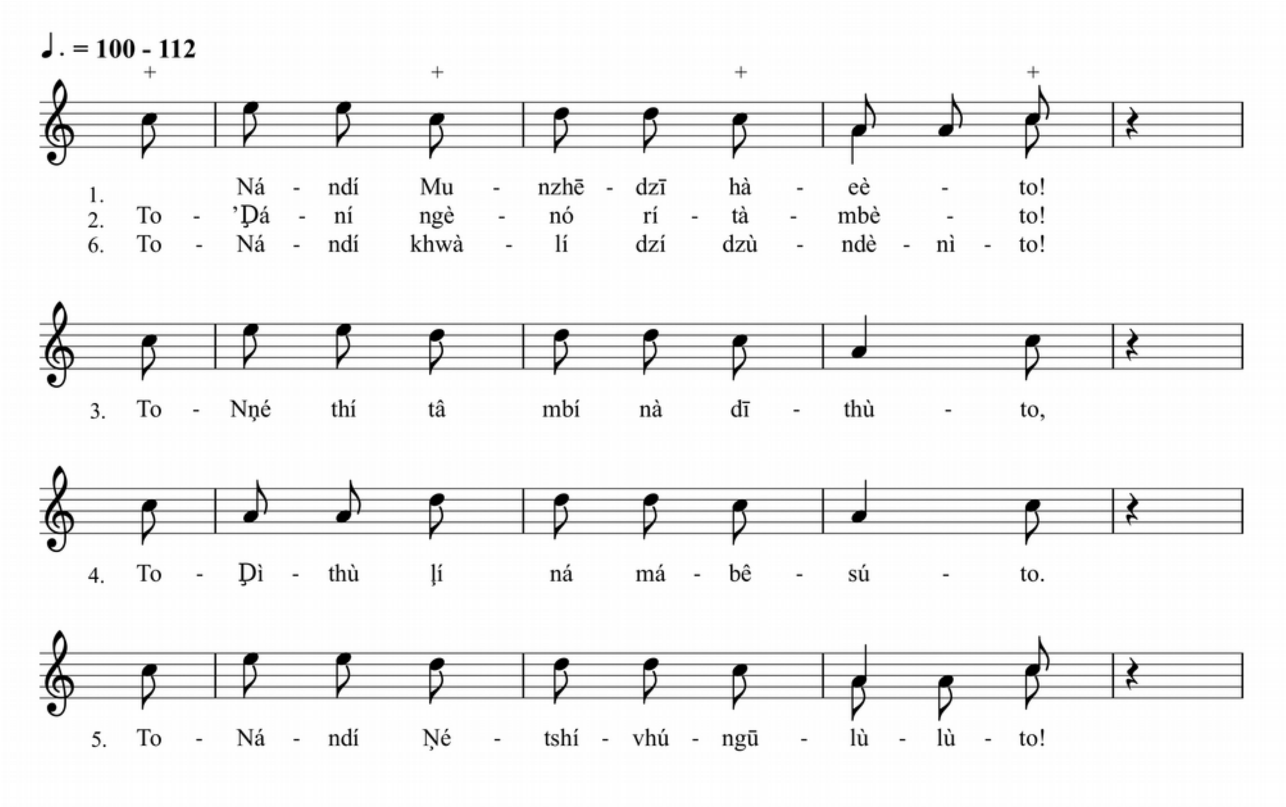


Figure C.9: Excerpt from "Nandi Munzhedzi Haee" (Blacking 1967:92)

Now, while we might have reason to believe that our perceptual apparatus favours a particular, ordered interpretation of the incoming auditory stimulus, we nevertheless find widespread and systematic confounding of that interpretation in African music. For instance, "syncopation", "off-beating" or "metrical inversion" all speak, in subtly different ways, to the notion that African music tends to place musical emphasis between, rather than on metrically strong pulses.[194] Locke examines offbeat accents and notes that "[w]hen one of the three important offbeat positions within each main beat is consistently accented […] a strong pull against the main beats is created." (Locke 1982:228) Temperley views this in GTTM terms,

---

[194] Blacking also points out that the term "syncopation" might be loosely applied to polyrhythm (1955:17).

concluding that "... the African mode of perception gives relatively more weight to the regularity rule, and relatively less to the accentual rules." (Temperley 2000:79)[195] He proposes the term "syncopation shift" to account for the tendency to shift notes to a point ahead of the beat, noting the widespread use of such in "... jazz, rock and other kinds of recent popular music..." (Temperley 2000:82–83). These "shifted" notes, he maintains, are not in conflict with the prevailing meter, but are rather reinforcing it. This stands in contrast to Blacking's (1967:157–160) report of an apparently Venda-specific criteria for strong beats, motivated by participants' elicited clapping (cf. Figure C.9 above). Perhaps, Temperley replies, these participants did not aim to clap on strong beats (Temperley 2000:85–86).[196] As regards Xhosa music, Hansen observes that "[i]n all music accompanied by clapping, the vocal effort frequently precedes the physical movement; the strong beats (off-beats) coincide with the singers' arms being raised and extended to clap; this is the act of tension which is resolved in the clap." (Hansen 1981:629) In fact, "[o]ff-beating, or, more specifically, the off-beat phrasing of melodic accents (accompanying word-syllables) against the main beats of the basic clapped and/or danced metre of a song, is a characteristic stylistic feature of the Xhosa musical tradition." (Hansen 1981:642–643) In considering Blacking's endorsement of Von Hornbostel's theory, that tension (and thus metrical stress) resides in preparation whereas resolution lies in execution, Hansen concludes thus: "If this explanation of African syncopation or 'off-beating' cannot be applied universally, it nevertheless applies to Xhosa music." (Hansen 1981:627) Kubik, too, seems to expect the predominance of "motor and acoustic accents on up-strokes (lifting the hand, beater, etc)..." in "... those African musical cultures having a tendency towards relatively little polymeter, and in the absence of asymmetric time-line patterns...", and explicitly cites South Africa as an example hereof (Kubik 2010b:91). Consider then, as Kubik does, the upstroke [ke] in jive/twist/*sinjonjo*: "By its acoustic prominence and anticipation of the beat, reinforced by anticipation of harmonic change, the up-stroke exerts a strong 'pull' on the listener's reference beat." (Kubik 2010b:94) Even more extreme are the displaced accents of *mbaqanga*, which Kubik reciprocally links back to some of Hansen's observations regarding Xhosa music: "In Xhosa and some other [...] there seems to be a

---

[195] In fact, to counter the view that "African music possesses more syncopation, that is, more conflicts with the underlying metrical framework; and African listening requires a greater ability to maintain a steady beat despite conflicting accents", Temperley proposes that "... the greater tendency of Western listeners to shift their metrical structures in response to phenomenal accents might be seen as a greater sensitivity to metric shift in the music... Perhaps in some cases, African listeners would 'mishear' such music, regarding the fluctuations as syncopations and attempting to maintain a perfectly regular beat throughout." (Temperley 2000:79) Rycroft has expressed the same perceived dichotomy: "... it would seem that, so deeply within him lies the *over-riding numerical order scheme* that he finds it quite unnecessary to emphasise it by equally spaced intensities as is done in our own (pre-contemporary) music. Similarly one might instance our own deep feeling for 'key', which is so well established that we have no need of an incessant tonic or dominant drone as in some less harmony-based folk music..." (Rycroft 1954:26–27)

[196] "We should be cautious about admitting this possibility, since a major source of our evidence for meter rests on the assumption that informants' clapping – especially elicited clapping that is not usually done – is an indication of (some level of) the meter they perceive. Still, it is quite clear that clapping patterns do not always indicate strong beats. In some songs, clapping patterns are irregular (see Jones 1959 II:2); in others, a single clapping pattern may be aligned with the song in different ways (Jones 1959 I:17; Agawu 1995:67-8); in some drum ensemble pieces, several simultaneous and conflicting clapping patterns may be used, despite the general assumption that only one primary meter is present (Jones 1959 I:116). Allowing the possibility that elicited clapping patterns do not always indicate listeners' perceptions of meter adds a major complication to the empirical study of meter; but it is probably something we should consider." (Temperley 2000:85–86)

common stylistic trait with regard to the relationship between the beat and melodic-harmonic patterns, namely the anticipation of the themes' and chords' entry one or two elementary pulse-units ahead of the dance steps (the beat), and clap beats or other objectified or merely thought reference-lines.” (cf. Hansen 1981:732; Kubik 2010b:101) Still, “... in most African traditions the inner beat remains associated with down-strokes, however much this fact may be veiled by a general tendency to underaccentuate beat 1 [of 4]” (Kubik 2010b:91–92). Dargie's examination of songs from the Lumko district reveals sporadic, yet systematic displacement of claps, footfalls or other sounds, though apparently in the opposite direction: “The **movements** of voice, hand and foot (and of bow-stick etc.) begin on the same pulse, but the **sounds** are staggered”, leading him to produce transcriptions showing “the *springing points* of clap and dance beats, not necessarily where the *sounds* fall.” He relates these back to Rycroft's “near-miss” placement technique (Rycroft 1962b:83,84, 1971:240; Dargie 1991:34–35,47). In fact, Rycroft singles out Xhosa dance-songs for “... what seems to be a subtly calculated off-beat relationship between word syllables and the regular dance-step and hand-clap rhythm (Rycroft 1962b:83). He proposes that “... in Xhosa singing, instead of the best being made to coincide with the *release* of a consonant – into a vowel, so that the onset of the vowel is on the beat, as is our own practice – it coincides with the initial *closing* or thrusting movement of the consonant […] so that the commencement of the vowel invariably occurs later, a little after the beat.” (Rycroft 1962b:84) In the context of the Zulu work-song (cf. Figure C.8 above), he argues that the vocal sound must cease “... due to glottal closure – at the actual moment of maximum exertion. It is, in fact, a physiological necessity.” (Rycroft 1967:93, cf. 1962b:82–83) Interestingly, such off-beat vocal phrasing against on-beat clapping or dancing, as is found amongst the Xhosa, stands in contrast to its complement found further north, i.e. a common penchant for “... rhythmic regularity in the vocal line, contrasting with rhythmic subtleties in the accompaniment” (Rycroft 1961:84)[197]

And yet, while some extreme applications of this idea are to be found, for the most part “... it is never retained throughout a song, which would cause the melody notes always to fall regularly between the beats of the basic metre. If this was done, then the listener would readjust, and the total effect lost... off-beat phrasing of accents must threaten but never quite destroy the orientation of the listener's subjective metronome... [it will] tend to heighten musical excitement.” (Waterman 1952:213) As Temperley points out, “... a metrical structure is best regarded as something in the mind of the listener, rather than being present in the music in any direct way.” (Temperley 2000:67) Agawu similarly underscores the importance of “... a normative grouping of notes, allowing for the possibility of real and apparent departures from this norm”, to which he ascribes “... the source of much excitement...” (Agawu 1986:69) Chernoff engages the same matter (Chernoff 1979:58,98). Some go as far as to suggest that, in African music, contrametrical organisation predominates over commetrical organisation, without violating the metrical cycle (Kolinski 1973; Arom *et*

---

[197] Rycroft suggests that, in this particular case (Mwenda Jean Bosco's guitar songs), “... the voice really accompanies the guitar, rather than the reverse [, being] a kind of descant, tethered to the cycle of harmonic progressions dictated by the guitar.” (Rycroft 1962a:86–87)

al. 2004:208; Agawu 2006:23).[198] Rycroft, noting a widespread tendency to forgo metric accentuation in African music, remains doubtful whether the "presence or absence of a single 'main' metrical framework [can be discovered] on the criterion of whether or not this actually comes to overt expression through accentuation." (Rycroft 1962a:101) Hansen acknowledges that "[i]n Xhosa music, I found that physical movements (clapping, drumming, dancing) coincide with the vocal effort almost as much as they do not." (Hansen 1981:630) She finds that, in each song, "... off-beating' between voice(s) and claps occurs approximately half the number of claps constituting the strophe", especially as regards the so-called 'intrusive' clusters (1981:633). Even whilst acknowledging off-beating in Xhosa music (1981:642–643), she simultaneously notes the possibility of accounting for such in terms of polyrhythm. "It seems to me, that the presence of off-beat accents in Xhosa music, produced by polyrhythmic techniques [Jones], and also physical movement which compel the vocal effort to precede the physical effort [Von Hornbostel], indicates the application of Jones' and Hornbostel's theories at the same time in one and the same musical tradition." (1981:642)[199] However, she does earlier state that "[i]t appears to be a rule of Xhosa music that melodies should begin on an off-beat; this coincides with the strong initial accent of the melodic pattern, and is always sung with loud emphasis." (Hansen 1981:624–625) Agawu, too, notes that "... in songs, as in drum ensemble music, phrases rarely begin on downbeats." (Agawu 1995a:64, cf. 66,110) So too Nketia, who reports "... a marked preference for phrases which begin before and after the main beats of a gong phrase – that is off the regulative beat." (Nketia 1963:88). On the other hand, Hansen observes that "[v]ery few Xhosa songs exhibit cross-rhythms in which the melody and basic metre are staggered perpetually...", though the Mfengu "... off-beat like anything." (Hansen 1981:645) Therefore, these same melodies might well end on a metrically strong pulse: "This is what Jones has described as the 'teleological trend' in African music: songs 'lean towards the end of the lines: it is at the end where they are likely to coincide with their time background'" (Jones 1959a:49, cf. also 41,84,86,126; Hansen 1981:624). Chernoff is equally certain that, in African music, "... the main beat comes at the end of a dynamic phrase and not at the beginning." (Chernoff 1979:56) Rycroft, upon reviewing some of Jones' evidence, concludes that "... it seems clear that the musical intention of Tonga and Lala performers is the ordered creation of rhythmic 'discord' and its planned resolution into 'concord'..." (Rycroft 1954:23) Elsewhere, Jones clarifies by explaining how the standard bell pattern, combined with regular clapping, comes to be regarded as end-oriented. (Jones 1959a:54) Locke reports that, in performance, the standard pattern is most frequently started on its supposed second note, once more suggesting an end-accented grouping (Locke 1982:220,225). Finally, considering Locke's assertion

---

[198] In fact, Kolinski rearranges Jones's transcription of the Nyayito Dance of the Ewe and finds that "... about 60% of stresses are commetric." Thus he argues: "The extraordinary complexity often encountered in African music results in most cases from a combination of different types of strictly or predominantly isometric patterns rather than from an opposition between isometric and heterometric patterns." (Kolinski 1973:497)

[199] See (Hansen 1981:632–642) for a considered appraisal of these ostensibly opposing views, with notated examples. Blacking also suggests simultaneous application of both processes in the *ngeniso* movement of a Chopi *Ngalanga* dance (Blacking 1955:17–18). Dargie distinguishes between those songs of the Lumko district which use additive rhythms, and those which "... set up different rhythms going at the same time, two against three, three against four..." (Dargie 1991:34)

that "... each moment within the metric matrix has an inherent rhythmic valence..." (Locke 2011:52), the afore-mentioned characterisations would all seem to suggest a systematic leveraging of the perceptual bias toward some preferred ordering, starting each phrase asynchronously to the prevailing metre and, as it were, "allowing the spring to unwind" in progressing towards the end of the phrase.[200] Of course, the very same description could as easily be applied to the progression from unstable to stable sonorities in Western tonal music, already recognised as a driver of directed motion (cf. Chapter 2 & 3).

It may be that some traditions merely bifurcate between asynchrony and synchrony, but the employment of cross-rhythm does present a particularly systematic solution to the task of attaining a more subtle ebb and flow between these two extremes. Von Hornbostel had earlier noted that "[t]he combination of binary and ternary time is characteristic of African metre in general." (Hornbostel 1928:52) According to Agawu, "[t]riplet effects dominate most transcriptions of West African music." (Agawu 1986:79)[201] Temperley applies GTTM's "event rule" to a typical case, concluding that hemiola comprise "three-level metrical structure, which is ambiguous in terms of its middle level but quite clear in terms of its upper and lower levels." (Temperley 2000:80) But it is Locke who recognises the kind of tension we are after, in noting that "[w]ithin any cross-rhythm there is one moment when the beats occur together and there are other moments when the beats occur before or after one another in distinctive ways." (Locke 1982:235) Fundamentally, "... the artful beauty of phrases with 3 : 2 patterning frequently draws upon motion towards the resolving, cadential moment when the two streams temporarily coincide." (Locke 2011:56) Rycroft notes with interest to what extent "... rhythmic concord coincides with the other characteristic 'punctuation signs' of final or semi-final tone and interval, and long note-duration, to achieve climax..." (Rycroft 1954:24) In one specific context, Kubik observes such tension so: "The rhythmic structure of *kuhunga* is based on a cycle number of 12 elementary pulses, which are grouped as triplets. While the drums perform this triple rhythm, the *tutanda* twist their pelvises in a duple movement that creates tension, arising from the contrast between 4 x 3 = 12 elementary pulses (drum rhythm) and 6 x 2 = 12 (pelvic movement)." (Kubik 2010a:372) He also considers the "complex regularity" of music employed in the *mukanda* initiation ritual, similarly based in cross-rhythm, and invokes Jung's proposition that "... structures of strong inner coherence have a compensatory effect in states of severe tension." (Kubik 2010a:364; cf. Jung *et al.* 1964:213) His emphasis is here less on particular musical structures than on the deeply coercive effect of "... their general qualities of complexity, inner logic and harmonious order", yet it is clear that such a systematic approach is valued for more than

---

[200] In fact, Locke proposes that "... the musical rhythm of separate phrases and the entire ETC is shaped by motion towards a commonly felt downbeat..." (Locke 2011:52)

[201] Agawu prefers to account for these in terms of "rhythmic formations in spoken language." (Agawu 1986:79, 1987:406–408, cf. 1995a:34), later proposing a cycle of "causal or organic relationships" between gesture, language, vocal music, instrumental music and dance (Agawu 1987:404). He explicitly refutes Von Hornbostel's (1928:52) notion that "African music is ultimately founded on drumming", invoking language in its stead (Agawu 1987:415) and concluding that "... language, as an intensification of gesture, is the generator of *all* music." (1987:418)

mere immediate aesthetic gratification.[202] The perpetual and progressive transition from synchrony to asynchrony, and back again, is here being harnessed to entrain the initiates' entire beings towards homeostasis. Dargie finds similar cross-rhythmic interaction between body rhythm, clapping and sung words amongst the Xhosa of the Lumko district (Dargie 1991:34–36). However, "... cross-rhythm[s] need not begin on the moment of simultaneity […] in other words, [they] may be phrased in several ways." (Locke 1982:235) Furthermore, "... the musical period of a cross-rhythm may be [a fraction of] the time span of [a cycle]" and "... simultaneous cross-rhythms of different ratio and/or duration may occur...", but it is perhaps most important of all to realise that the components of a given cross-rhythm merely function each as a grid against which more complex rhythms are timed, and thus that "... cross-rhythms do not necessarily appear in their theoretical form." (Locke 1982:233) In fact, Rycroft identifies what seems to be "... embryonic polyrhythm, in the form of real hemiola, the opponents being, not the rhythm patterns of different participants, but different 'phrasings' of the same part, simultaneously: a polyrhythm of melody against accent, in fact, and a case of crossing one's own rhythm." (Rycroft 1954:20) All of this would seem to introduce significant complexity into any algorithmic approach hoping to identify such phenomena.

While "off-beating" and "polyrhythm" both appear to exploit tension at a macro-structural level, a more immediate effect is obtained by transient variation of, or deviation from, that template. Regarding the application of "non-linear offbeat phrasing", Kubik observes that "[t]he tensions created by making the autonomy of one's patterns more prominent are counteracted by the cohesive force of the three inner reference levels. The more one stresses overtly the autonomy of the patterns, the more the three levels are reinforced. It is a subtle 'forward and backward' at the threshold between disorientation and reorientation." (Kubik 2010b:49) Once again we see the importance attributed to the interaction between synchronous and asynchronous elements. What is more, while macro-structural regularity associates with homeostasis, "... a master drummer [...] may superimpose a string of patterns in a contradictory meter […] associated with heightened psychical states...", while in the presence of polymeter, "... the master drummer may change his internal beat relationship [which is] connected with ecstatic psychological states and often occurs at climax levels in a performance." (Kubik 2010b:49–50)[203] Avorgbedor draws comparable conclusions from Yeʋe cult music: "The opposition of free and strict time structures heightens the dramatic moment and places the audience (dancers too) in a situation where temporal orientation is blurred and disembodied temporarily. The role of this ambiguity in the affective determination of an elevated or spiritual experience cannot, therefore, be overstressed... While the monorhythmic patterns convey the idea of an 'arrest of time' (i.e. repetition), the

---

[202] Merriam, too, notes that "... the Basongye do not seem to conceive of music as an aesthetic phenomenon; rather, it is a useful activity which serves certain sharply defined ends..." (Merriam 1962:124) However, "... there is much reason to believe that music, or at least certain types of music, is conceived aesthetically in some African groups, with a well elaborated system of relationship of certain sounds to beauty or ugliness, for example, or with the composition of music and the playing with musical form for its own sake..." (1962:125)

[203] Though it is not entirely clear to what extent Kubik distinguishes between "polyrhythm" and "polymeter", his use of the term "polymetric" is this context (multiple musicians) is perfectly compatible with Kolinski's (1973:501,502) assertion that no listener or performer is capable of "true polymetric perception".

circular dance motion also reinforces the idea of the 'atemporal'." (Avorgbedor 1987:15) This theme (the attainment of ecstasy) recurs in writings about many different kinds of world music (e.g. "Duende" in Flamenco (Lorca 1998)), is already present in Von Hornbostel's early comments regarding African music (Hornbostel 1928:38,59), and is specifically addressed as regards Xhosa music by Hansen: "[t]he transcendental quality of music as 'bringer of *ihlombe*' to both performers and listeners is crucial" (Hansen 1981:719). Nor are the effects of such *ad hoc* variation restricted to the domain of rhythm. Kubik considers the restraint with which *amadinda* players might vary their melodic patterns: "Often the audible result of variation is that of dissonances which bend to raise the tension". Also: "Sometimes [the *okukoonera* part] only follows the *amatengezzi* pattern approximately; it may stop temporarily or a few notes may be left out to increase tension. But all this is employed in moderation..." (Kubik 2010a:309) Once again we are reminded of Waterman's (1952:213) remarks, in that overzealous employment of variation will destroy the very reference from which that variation derives its effect.

Kubik's "tonal-harmonic segmentation of a cycle" (2010b:44) was earlier mentioned as a further important structuring principle. Indeed, "[t]he principle of shifting tonality, which dominates all Xhosa music (and most, if not all African music, as far as I know), is to be observed in the root progressions of the chords which underlie the melody... provid[ing] the tonal and harmonic foundation of the whole song." (Hansen 1981:670) In a footnote, Hansen indicates precedents for this view: "Blacking drew attention to this important feature of African music in Southern Africa in 1959, in his study of Venda ocarina music..." (Blacking 1959:23); and also "Rycroft pointed out that this essential feature 'is also evident, in embryo, even in the Zulu three-tone songs, in which a upper fourth and a lower fourth are contrasted' (1971:235)" (Hansen 1981:704 fn. 19; cf. Rycroft 1959:26, 1971:226–235). In fact, Blacking describes "... a regular shift between second inversions on the the Tonic and sevenths on the sub-dominant" (Blacking 1959:21), though his transcriptions would seem rather to suggest alternation between supertonic (minor) and dominant instead (Blacking 1959:20–21).[204] The key insight, however, lies in noting the relationship between an apparent root movement by tone and a possible underlying root movement by fourth.[205] Rycroft's "tritonic structure", consisting of an octave partitioned into a lower perfect fourth and an upper perfect fourth, separated by a whole tone, is indeed shown to permeate many more elaborate Zulu tetratonic, pentatonic, hexatonic and

---

[204] Blacking describes a "root-progression" consisting of a "... *canto fermo* which moves a whole-tone above and below a tone which may be regarded as a 'tonal centre'. Combined with the unresolved 'harmonic' progression, this gives the music the quality of perpetual motion, and there is no point which might be called a cadence." (Blacking 1959:21) The unresolved 'harmonic' progression manifests in the alternation between dyads of a fourth ("points of least tension") and those of a tritone, minor third or minor seventh ("points of greater tension") (Blacking 1959:23). He also suggests similarities between this progression and those of "... several other African societies, and even in the Tonic-Subdominant-Dominant strumming that one often hears on guitars and old pianos." (Blacking 1959:23) See further developments of these ideas in Bosco's *Masanga* (Rycroft 1961:85).

[205] To the charge of imposing a Western harmonic framework on the Venda ocarina duets, Blacking contends that his informants' particular choice of ocarinas, from all of the variously tuned options available, serves as clear evidence of a preference for a particular arrangement of relative pitches, and thus of an implicit harmonic framework which is in some way comparable to that of Western music (Blacking 1959:22).

heptatonic scales (Rycroft 1975a:380–382). From this evidence, Hansen concludes that "... penta, hexa and hepta modes may be seen as an extension of this rudimentary three-tone structure by means of an 'evolutionary' process involving 'the accretion of whole tones'." (Hansen 1981:668–669; cf. Rycroft 1971:232, Fig. 27) Nonetheless, the likelihood of ideocultural transference is made explicit when she states that the "... two-tone motif constitutes a sort of 'dominant-tonic' ostinato, and the constant shift from one tone to another represents the shifting tonality which is observed in most (if not all) African music." (1981:679) Such a view has been with us at least since Von Hornbostel's remarks: "a period breaks up into question and answer, tension and relaxation, arsis and thesis, which in antiphony are mostly distributed between solo and chorus, and, as regards tonality, represented, as a rule, by the contrast between dominant and tonic." (Hornbostel 1928:52) In her own data, though, Hansen is even able to relate the tonal centres to specific feet by studying the associated dance movements (1981:672). To the details of such tonal shifts, she further observes that "[a]ll songs which are antiphonally structured, and which feature off-beat phrasing between melody and percussive accompaniment, show a shift in tonality from the 'counter-tonic' (around which the song leader's Call centres) to 'tonic', which is established by the chorus." (1981:699) On the other hand, "[t]rue polyrhythmic songs such as the dance songs of the Cape Tribes Proper, show small shifts of tonality within the strophe, but the tonality remains stable at the beginning and end of the strophe, being centred on the tonic. This makes for stability and cohesion in these songs, in which vocal-rhythmic interplay often reaches considerable complexity, from the linear point of view (even though vertical combinations may not involve more than three or four parts in all)." (1981:699) She notes exceptions to these principles (e.g. "[i]nstrumental music (e.g. bow songs) generally move 'tonic' to 'counter-tonic'" (1981:700)), but reiterates the association of the chorus with the 'tonic'.[206] These general tendencies are found to be congruent with those suggested by Blacking: "...the Western music runs **1** - 2 - 3, **1** - 2 - 3 etc., whereas the African music runs 1 - **2** - 3, 1 - **2** - 3 etc.", and thus, while "[t]he general pattern of Western music is one of Relaxation-Tension-Relaxation", in African music "[t]he pattern is one of Tension-Relaxation, Tension-Relaxation." (Blacking 1955:16; cf. Hansen 1981:626–627).[207] The latter characterisation, in particular, establishes a rather explicit link between tonal function and tension, even if only in the mind of the researcher. Tracey, for one, seems to assign starting points based on the reverse pattern (relaxation → tension), though ultimately deferring the matter by pointing out that "... it is only the demands of paper writing that make it necessary to choose one starting point." (Tracey 1970:42) Can this tension be understood as expectation? Would such expectation reveal its presence in priming experiments? Can we so discover the music's teleology? This is

---

[206] In a discussion of a Tonga bow song, Rycroft seems to suggest that the bow assumes the role of the soloist's call, with the singer providing the response (Rycroft 1954:25), seemingly contrary to what Hansen concludes amongst the Xhosa.

[207] In support of this distinction, Blacking cites aspects of African melody which similarly run counter to Western expectations, specifically that "... the natural motion of melody is downward" (Hornbostel 1928:34) along with Sachs' characterisation of "pathogenic" melody (Sachs 1943:41). He further notes that in the West, "[t]he tendency is to sing up the scale, whereas in Africa the tendency is to do the opposite: some instruments are even tuned from the highest to the lowest note – or, as African musicians say, from the smallest to the largest." Finally, he contrasts the patterns of motion in ballroom and jive dancing as evidence of the same (Blacking 1955:16).

precisely the path of breadcrumbs we are hoping to follow.

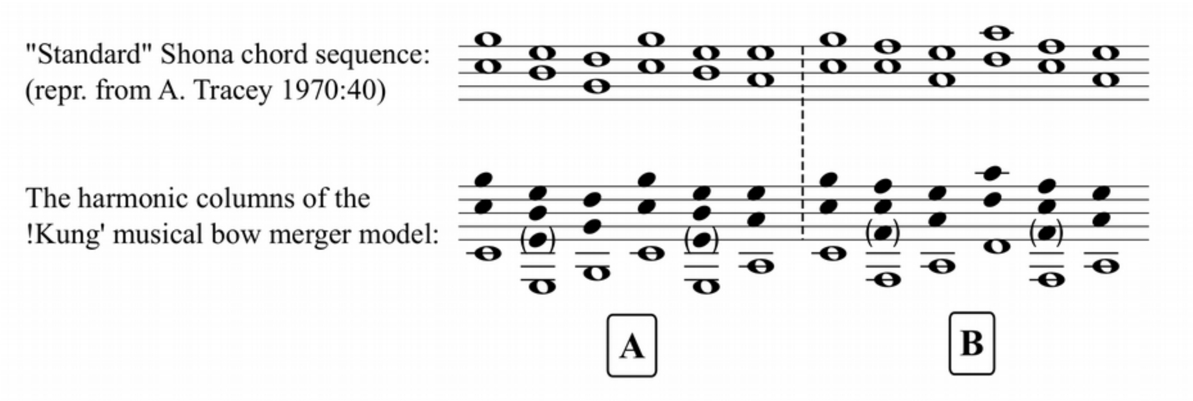


Figure C.10: !Kung' tonal-harmonic merger model applied to the "standard" Shona chord sequence

A reconsideration of Kubik's !Kung' tonal-harmonic merger model might be of value here, if only for the potential that such might hold in revealing somewhat more general principles at work. The most elaborate application thereof is perhaps in regards to the so-called "standard Shona chord sequence" (cf. Figure C.10 above) as described by Tracey (Tracey 1970:38–43). Suggesting that "... each harmonic movement from chord to chord must embody something that is right to the Rhodesian ear" (1970:40), Tracey enumerates a number of structural characteristics, including the long descending scalar passage from the E in the second chord, together with the strict alternation of fourths and fifths in the alternate part.[208] The sequence is also distinguished by its alternation, for the most part, of ascending root movement by alternating thirds and fourths, reportedly "... found also in many, perhaps all other mbira songs." (Tracey 1970:40) Tracey also suggests that every third chord might group perceptually to "... correspond with the general southern African tendency of harmonic movement to alternate up and down by one step or tone." (1970:40) What speaks most to Kubik's model, though, is Tracey's suggestion that the *motif* of the first half is being answered a fourth higher by a contrasting statement of the same (1970:40). This, Kubik suggests, is evidence of bipartite structure, essentially a meta-merger of two complete bow mergers spaced a fourth apart, giving "... absolute congruence in sonic content and spatial layout (inversions) between the bichords in the 'standard' Shona chord sequence and the harmonics of the four harmonic columns of the !Kung' merger model" (Kubik 2010a:238–239). Elsewhere, in surveying the arrangement of keys on various lamellophones of the lower Zambezi valley, Tracey again notes the separation of primary tonal centres by the interval of a fourth (Tracey 1972:92,94). Neither Tracey nor Kubik, however, explicitly attributes any particular significance to this conspicuous fourth, though it would be enticing to speculate whether this particular interval is selected precisely for its perceived effecting of directed motion (cf. Chapter 3). Tracey does go on to point out that the "standard" sequence might also be encountered in a modally transposed form (Tracey 1970:43),

---

[208] This, Tracey points out, stands in contrast to the parallelism encountered in Blacking's study of the Venda *tshikona* reed-pipe dance (Blacking 1967:177; Tracey 1970:40).

seemingly a parallel to the *miko* system of Kigandan *amadinda* xylophone music (Kubik 2010a:261–263),[209] as well as noting that the "third note of the scale" appears to have an important function as a "mediator between the four quarters of many songs" (Tracey 1970:43). Stylistic idiosyncrasies such as these would likely confound any overly simplistic application of the !Kung' tonal-harmonic merger model, and Kubik does not venture any further down this path.[210]

Merriam considers the misplaced emphasis on "... the use of drums and idiophones" in African music, citing large "... large groups […] where drums are practically unknown", or at least heavily restricted in their use (Merriam 1962:121). Rycroft asserts that the music of "Bantu-speaking peoples of the Nguni group living in the extreme south-eastern corner of Africa" is predominantly vocal and notably devoid of the "intricate percussion patterns typical of Central and West Africa" (Rycroft 1959:26, 1975a:351),[211] and Kubik finds timelines to be less significant a structuring element in South African music, placing greater emphasis on "motor and acoustic accents on up-strokes" (Kubik 2010b:91). "In most other parts of the continent, particularly in its eastern and southernmost regions, [timelines] were unknown before the mid-20th-century impact of popular music from the Congo and other parts of Africa." (Kubik 2010b:57) Yet Hansen maintains that "[t]he clap patterns (or their equivalents in drum beats and dance steps) are the most important structural device in this sort of music. They provide what Jones has called the 'time background', and what the Cape Nguni refer to as a 'rim' or 'edge', against which the voice parts can merge and conflict temporarily." (Hansen 1981:649) There is more than a little contention here. Not only does Hansen invoke the self-same term that Jones used to describe what Kubik has elsewhere dubbed the "timeline", but she evidently finds that this concept is significant enough in Nguni music-making to have earned a Xhosa label.[212] What is more, there might well be a conflagration of concepts here, since Kubik has advised us that the timeline is distinct from the reference beat, the latter being more closely associated with the dance steps. If this is the case, we might want to restrict our focus in this regard to clap patterns, which intuitively seem to function much like the more esteemed bell patterns of West Africa. Indeed, Merriam singles out clapping as the most universal percussion device used across Africa (Merriam 1962:121). This is not to say that we expect to find identical patterns in such far-flung traditions, only that these deserve to be carefully reconsidered in the light of their possible function as an "important structural device" (as has already been posited above). Certainly, Hansen

---

[209] Such "modal transposition" is for Von Hornbostel merely evidence that the player "... realizes melody above all as an act of motility, regarding its audible quality rather as a side-issue, although a desirable one." (Hornbostel 1928:49) By this he privileges "unity of 'form' in succession, i.e. melody" over "the disconnected chord and its consonance" (*ibid.*)

[210] As an aside, Kubik does note the spacing of the lowest notes in Tracey's "kalimba core" as mirroring the spacing of the lowest fundamentals in the !Kung' tonal-harmonic merger model (Tracey 1972, p.88,99; Kubik 2010a, p.234) (Tracey 1972:88,99; Kubik 2010a:234).

[211] Elsewhere, Rycroft cautions against "... attribut[ing] the predeliction for choral singing among these people solely to missionary influence. There is in fact a strong indigenous precedent for choral singing. Traditional music in South Africa, unlike the instrumental and percussion-based music of more northerly Africans, has always been predominantly vocal. In formal group dances […], the performers sing their own dance music while dancing, without any accompaniment; and their dance songs are polyphonic." (Rycroft 1991:9)

[212] May we remind the reader that terms such as "rhythm", "harmony" and "melody" do not have such labels.

finds no difficulty in ascribing Jones' description to Xhosa music: “As Jones has pointed out, 'a definite mathematical relationship' exists between the number of claps covered during the singing, and the number of beats (pulse groups) in the vocal part. Individual singers may introduce patterns at any point during the singing (i.e. within the execution of the strophe); the metres of these patterns may differ from that of the basic pattern, but all must synchronize with the basic metre.” (Hansen 1981:651) In fact, Jones draws two fundamental distinctions between the rhythmic organisations of hand-clapping and drumming. Firstly, while “the main beats never coincide” in drumming, clapping patterns always have a simultaneous starting point and meet periodically. Secondly, while the master drum patterns undergo variation in response to interaction from the dancers and other musicians, clapping patterns are repeatedly tirelessly, without variation (Jones 1954:39).[213] England's study of *Ʒũ'|'wasi* “counterpoint” illustrates the structural importance of various clap patterns employed therein (England 1967 passim). Still, Kubik cautions against prioritising the timeline over other levels, especially as regards the dance movements. Against the backdrop of authors studying drumming in Ghana and claiming timelines to function as primary reference level, Kubik cites contradictory evidence obtained in an interview with one Maurice Djenda, an ethnology student and musician from the Central African Republic. Asked: “And how do the dance steps behave with regard to the *mayo* [time-line pattern]?”, Djenda replied: “This has no influence on the music. The *mayo* is an accompaniment; one is not guided by that pattern.” (Kubik 2010b:62) Nzewi similarly maintains that the West African bell pattern is simply a “phrasing referent” and “not the structural fundamental” (Nzewi 1997:35), while Locke places equal emphasis on the time line and the beat (Locke 1982:222). Perhaps this is a key insight then, that the various structural levels really do function independently, as well as interacting with each other. This is, at the very least, a shift in emphasis from the typical hierarchy of structural levels employed in Western music. In his continuing discussions with Maurice Djenda, Kubik also explored the notions of additive and divisive rhythm. Jones, it must be noted, conceived of timeline patterns as additive (Jones 1959a:210). No doubt this stems from their characterisation as having “... an asymmetric inner structure, such as 5 + 7 or 7 + 9, against which the 'melodic and rhythmic phrasing of other performers is juxtaposed.'” (Kubik 2010b:57) Still, Kubik's descriptions do not exclude the possibility of symmetric timelines, which might then be perceived as essentially divisive, and clapping is mentioned as a valid method of performing timelines (Kubik 2010b:57). We note too that, whereas many additive pattens cannot be seen as divisive, all divisive patterns may be seen as additive, notwithstanding the fact that many formal definitions explicitly exclude this possibility. As regards Xhosa music, “[a]dditive and divisive patterns may occur simultaneously... In both songs the basic metrical pattern is divisive, and the vocal patterns are nearly all additive. This is the usual combination in Xhosa music.” (Hansen 1981:660) Hansen reports that “[a]dditive basic metrical patterns are rare”, though she does give some unique examples of the same (1981:658,661). But why should we bother with such a contested notion in Xhosa music at all? It is our belief (informed by intuition, and little else at this stage)

---

[213] Though noting Merriam's objection on this point (Merriam 1962:126–127,130 fn. 29), it is nonetheless of interest that such a description would equally apply to so-called “bell patterns”.

that timelines have an important teleological function wherever they are employed. Agawu, in his various hermeneutic readings of the standard pattern, seems to agree (Agawu 2006:9). That this is so for the five- and seven-stroke "standard pattern", we believe to be self-evident, though we are here inclined to part ways with Agawu on the details of his "generative" approach (Agawu 2006:31–32).[214] Temperley's more tentative account of the success of the standard pattern, rooted in its asymmetry, is perhaps more useful here (Temperley 2000:82).[215] Further to this, though, we believe that less-valourised patterns may have been overlooked, yet will prove to be valid and effective implementations of the same principles.

Timelines, as mentioned earlier, are held to represent a distinct level of temporal organisation in African music. The diagrams which follow are adapted from Kubik (2010b).

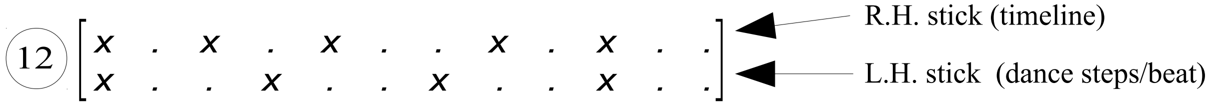


Figure C.11: Nyanja version of the 5-stroke, 12-pulse pattern (Kubik 2010b:65)

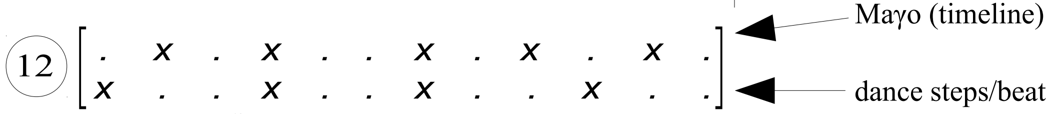


Figure C.12: Mpyεmɔ̃ version of the 5-stroke, 12-pulse pattern (Kubik 2010b:61)

In Figure C.11, we see one possible arrangement of the 5-stroke, twelve-pulse pattern, a rhythmic configuration which Jones (1959a) reported amongst the Yoruba, Ewe, Ila and Tonga. This might be described as a sequence of three events spaced at every second pulse, followed by a gap of three pulses, then followed by another two pulses again spaced with a pulse between. The remaining pulses in a twelve-pulse cycle thus constitute a second gap which occurs before the entire pattern is repeated. What has intrigued so many is made plain by comparing this to the layout of a piano keyboard, as illustrated in Figure C.13.

---

[214] By this, we do not mean to exclude Agawu's approach *in toto*, but merely to carve out a space for the inclusion of a more parsimonious principle. Agawu's notion of "play" is certainly relevant in music, but is essentially inaccessible to the kind of study envisaged here.

[215] Temperley cites Browne's (1981) observation that "... the tonic is the only diatonic scale degree that lies a half-step above one scale degree and a perfect fourth below another." Thus, he contends, the diatonic pattern (and thus the standard pattern) contain built-in cues to orient listeners (tonally in the former case, metrically in the latter). (Temperley 2000:82)

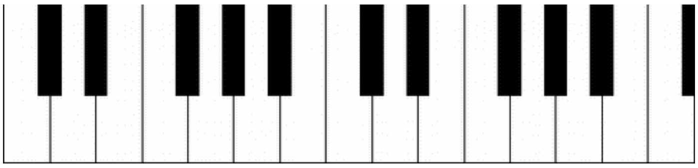


Figure C.13: Canonical keyboard layout

Consider each white and black note as essentially equidistant from both of its closest neighbours (which, in perceptual terms, they are generally held to be), then notice the groupings of black notes. A cluster of three black notes, separated by white notes, is flanked by clusters of two black notes, again separated by a white note each. Each cluster is then separated from its neighbours by two white notes (with no black note in between). Compare this to either of Figures C.11 or C.12. In the first case, a clear parallel is evident between the spacing of strokes and the spacing of black notes extending from the left-most note of each group of three. In the second case, which Jones (1959a) finds amongst the Bemba, Ewe and Lala, the same may be seen starting on the left-most note of each group of two adjacent black notes. Jones concluded that both were one and the same pattern, evidently alternate manifestations of something "... very deep down in the African musical mind" (Jones 1959a:212). Taking these beyond being merely novel observations requires some consideration of the significance of these "black note patterns", commonly termed "pentatonic scales".[216] A pentatonic scale may start on any note, being characterised by the particular grouping and spacing, relative to that reference note, of five notes to the octave as indicated above. So, a major pentatonic built on C, for instance, would include the notes D, E, G and A. This configuration is regarded as maximally tonally stable since it contains none of the most dissonant intervals between any of its notes, namely the minor second, major seventh and tritone. It may be derived by selecting any five adjacent notes along the "cycle of fifths", this in turn representing the closest perceived tonal relations between non-identical pitch classes. This relationship is understood to derive from the particular ratio of the respective frequencies between any two such related notes, being 3 : 2.

Few have been able to proceed meaningfully beyond this point. Koetting, for instance, warns us that: "Fascinating and seductive as these explanations may be, they do not provide an answer to how Africans perceive rhythm. They are not operating principles – they do not actually *account* for what happens in the music." (Koetting & Knight 1986:61) Yet, he does not direct us to interrogate ever more "... non-musical thought patterns, social structures, world view, or anything outside music." (*ibid.*)[217] Rather, he cites Ekwueme's call to "... discover and explain *what* the African musician does musically instead, merely, of *why* he does it" (Ekwueme 1975a:5; Koetting & Knight 1986:61) In fact, there are two matters which

---

[216] Specifically, these are anhemitonic pentatonic scales, since they include no semitone intervals. For the purpose of this discussion we will simply use the term pentatonic.

[217] With somewhat more hyperbole, Koetting elaborates: "Language, calendrical cycles, the complexities of kinship ties, or the attitudes of inter-urban taxi drivers towards the timing of their trips, fascinating though they may be, ultimately cannot account for timing in music." (Koetting & Knight 1986:62)

confound the simplistic account given up until this point. Firstly, the apparently "equal" spacing of notes on the piano, perceptually speaking, is an approximation to the pure ratios proposed above. In the case of the closest relation, the perfect fifth, the deviation is in the order of -2 cents, accumulating to almost -4 cents in the case of those notes separated by two steps along the cycle of fifths. This deviation is deemed to be acceptably small in the case of such closely related pitches. Secondly, pitch is perceived logarithmically, with every doubling of frequency resulting in a pitch rise of one octave. By this principle (loosely, the Weber-Fechner Law), the arithmetic midpoint of an octave (the frequency nominally halfway between its extrema) is perceived as somewhat closer to the upper bound of that octave. Specifically, the pitch having a frequency 1.5 times that of the lower bound (linearly halfway towards the upper bound at 2), is perceived as higher by $\log_2 1.5 = 0.585$ or 1.585 times the perceived pitch of that lower bound. This is approximated in Western musical practise by the ratio $7 : 12 = 0.583$. Thus, on the piano keyboard, we find these two physical "halves" of the octave (in terms of frequency, that is) populated by seven and five pitches, respectively. The confound emerges when we try to apply this same principle to temporal partitioning. Western music theory shows no evidence of such a view of rhythm. Indeed, Western culture, for the most part, favours a distinctly linear view of temporal perception. Still, questions regarding the applicability (or not) of Weber-Fechner to the human experience of time are not entirely unprecedented (Church & Deluty 1977; Gibbon & Church 1981; Gibbon 1981; Brannon *et al.* 2008; Jazayeri & Shadlen 2010; Provasi *et al.* 2011). What is more, questions persist regarding the interaction between various musical domains and the influence thereof on time processing (Firmino *et al.* 2009). It is therefore worth considering whether the isomorphisms observed between the pentatonic scale and the five-stroke, twelve-pulse standard pattern might derive from a common principle. By such a view, the five strokes are selected from the twelve possibilities in every cycle by virtue of their close relationship along a "temporal cycle of fifths". To accede to such a notion as this, we would first have to acknowledge something akin to octave-equivalence in the human experience of time's passing. In other words, that a 20-year-old's perception of the preceding 10 years of his/her life is comparable to a 40-year-old's perception of the preceding 20, or an 80-year-old's perception of the preceding 40. While popular discussion of such a large-scale phenomenon are fairly easy to come by, the scientific evidence focuses on the bisection of far smaller temporal intervals and has, in fact, served up compelling evidence supporting just such a logarithmic view (cf. citations above). This, in turn, allows one to consider whether the asymmetric patterns of strokes in the five-stroke, twelve-pulse standard pattern are, in fact, perceived (or at least, conceived of) as even bisections of "temporal octaves". The interval between the first strokes of each subsequent cycle of 12 pulses (1/12) might thus be logarithmically "bisected" by a stroke at pulse 8/12 (seven pulses later). At double this scale, i.e. over 24 pulses, bisection would occur at pulse 15/24. Admitting the possibility of overlapping processes, we might normalise this (mod 12, since these cycles are 12 pulses long) to account for the recurring stroke at pulse 3/12 in every cycle; similarly so over three cycles (bisection at 22/36, normalised to pulse 10/12) and four cycles (bisection at 29/48, normalised to pulse 5/12). Alternatively, we could simply consider the bisection at pulse 8/12 to have been "echoed" at every 7th pulse thereafter, up to the third repeat.[218] Thus we could account for every stroke in Figure C.11. In order to approach Figure C.12 along the same lines, we merely need to note that the cycle repeats *ad infinitum*, and that this pattern is thus nothing but a phase-shifted version of Figure C.11. More specifically, these two patterns are 180° out of phase, and their complementary, interlocking character with respect to each other is plainly revealed by juxtaposing one on top of the other. Thus similar derivations may be applied in each case by merely selecting one of two reference points. An alternative approach is to consider logarithmic trisection, whereby $\log_2 1.333 = 0.415$ suggests a stroke at 5/12 (0.417), with extension hereof proceeding in precisely the same fashion as suggested above.

The notions above might seem entirely fanciful until once considers pertinent evidence gleaned from video footage. Kubik, recognising the importance of viewing African music in terms of motor patterns (Kubik 1965:35), incorporated frame-by-frame transcriptions of film into his analytical method, subsequently reporting an interesting variance between the upswing and down-strokes of Ugandan and European xylophone players, respectively. Whereas Europeans tended to bisect the temporal interval between strokes into upswings and down-strokes of approximately equal duration, Ugandan players were found to apply around two-thirds of that time to the upswing, with only a third going to the down-stroke (Kubik 1972:33). More generally, Kubik reports that: "Triple subdivision of the space of time between two strokes is frequently found in Africa, the lifting of the striking medium taking two pulses, and the downstroke one. This is also widespread in work situations, for example pounding maize and millet, or using a hammer or mallet. The rhythmic organization of blacksmiths working around the bellows has already struck many observers." (Kubik 1972:34) Of course, Kubik's footage is limited by its frame-rate of 24 f.p.s. (Kubik 1965:37, 1972:33), and though his transcriptions depict a steady upswing over six frames followed by a down-stroke over three frames, it is not entirely untenable that the quantisation imposed by his medium may have obfuscated a logarithmic bisection, as described above.[219]

Notwithstanding the more contentious views being posited here, it is clear that a linear view continues to play a role. The dance steps or beat, for one, retain their linearly uniform spacing, and provide an apparent counter-reference to the timeline. Stone finds the genesis of such temporal duality in Jones (1959a:21–22), stating that "... Jones's crossing of rhythms suggests not only different rhythms, but different principles of constructing rhythms." (Stone 1985:142) She challenges the prevailing linear view by asserting that "... it does not follow that this background grid of an equally spaced pulse must then determine that the beat

---

[218] Such "echoing" would be consonant with widely mooted notions of "circularity" in African music, as evidenced in, for example, Rycroft (1967) or Anku (2000). Agawu, characteristically, questions the implied "othering" of African music as "especially circular", calling instead for the admission of "degrees of circularity" (Agawu 2006:40–41).

[219] To be entirely accurate, a logarithmic bisection of 9 frames would lie on frame 5.265, and more likely be rounded to frame 5. It nevertheless seems unlikely enough that the strokes should have lined up so precisely with the frames to warrant further enquiry on this, particularly in the light of advances in modern high-speed video photography – though finding surviving practitioners is likely to be problematic.

constituted of several pulses at the next level in the hierarchy must also be equally spaced. It is precisely with the recognition of unequal beats composed of equal pulses that ethnomusicology begins to break out of a linear perspective." (1985:140) Her notion of "mosaic time" introduces, amongst other notions, a distinction between "outer time", whereby action is coordinated, and "inner time", which exists only in participants' streams of conciousness (1985:146). By various accounts it appears that music is precisely revered for its ability to transport participants from outer, linear time toward something else. Nzewi, for one, holds that "... autochthonous African musical productions are abstract configurations which demonstrate the fairly common fundamental creative principle of mediating the physical and metaphysical worlds." (Nzewi 1997:13) Merriam suggests that "[i]t seems reasonable to assume […] that in African cultures, music is placed in a specific role in regard to the world as an entity, and particularly to the supernatural world." (Merriam 1962:124)[220] Locke reports that in the Eʋe tradition, "... music with a binary division of the main beats [is considered] to be gay and lighthearted", while that employing triple division is considered "serious and poignant" (Locke 1982:224).[221] Eliade suggests that "... religious man lives in two kinds of time, of which the more important, sacred time, appears under the paradoxical aspect of a circular time, reversible and recoverable, a sort of eternal mystical present..." (Eliade 1959:70; Avorgbedor 1987:13) Ubiquitous as such associations appear to be, the explanations typically forwarded are all grounded in either of the greater perceived complexity of the ternary structures or association with culture-specific values attributed to the number three. Perhaps a further explanation is now apparent. For the moment, we can only assume that linear and logarithmic views of time coexist in African music-making, each representing a different level of perceptual awareness. Though their interaction is, from an ideocultural viewpoint, an important stylistic characteristic of African music, it is unclear whether this is truly "by design".[222] It is certainly possible that this is yet another example wherein the listener is simply invited to switch between different modes of perception within the same performance.

Immediately apparent is that the same ideas may be extended to account for various 7-stroke, 12-pulse patterns, for example the following two discussed by Kubik:

---

[220] As an example, Merriam mentions "... the Lovedu of the Transvaal, where the drum cult serves as a fourth category of forces by which nature is regulated in the interests of man, the other three existing through properties intrinsic in various objects, animate and inanimate, the ancestors, and the control of the queen. […] [T]he drum itself […] serves […] as an abstract concept identified with power, especially in connection with rites of transition." (Merriam 1962:124)

[221] There is more than a little resonance here with medieval mensural notions of *tempus perfectum* (⊙, ○ or ⦶) and *tempus imperfectum* (Ͼ, C, Ↄ, ¢ or Ↄ⃒), the former, being ternary, being historically associated with the Christian "Holy Trinity".

[222] Agawu says much the same of his own approach: "It should be emphasized that the generative process being advocated here is not a report on existing compositional practices as such – although it certainly could be that – but an imaginative reconstruction. At best, a generative proceeding make explicit the logic of structure inherent in a given cultural product." (Agawu 2006:30)

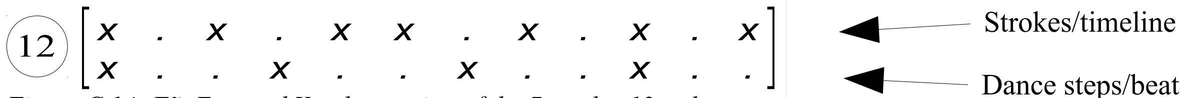


Figure C.14: Fõ, Ewe and Yoruba version of the 7-stroke, 12-pulse pattern (Kubik 2010b:77)

Figure C.15: Igbo version of the 7-stroke, 12-pulse pattern (Kubik 2010b:82)

Figure C.14 maps neatly onto the Western diatonic major scale, or the pattern of notes formed by white notes from C to C on a piano keyboard (cf. Figure C.13). Comparing Figure C.14 to Figure C.12, we might also notice that each pattern is the true complement of the other, with strokes in one perfectly coinciding with open pulses in the other. They also happen to be "retrograde" versions of each other, though retrogradation is generally dismissed as irrelevant to African expression (Agawu 2006:36). This all only goes to sustain the likelihood that both of these derive from a common principle. Figure C.14 would be recognised by Jazz musicians as the "lydian mode", or the pattern of white notes from F to F. By invoking the above-mentioned logarithmic approach to temporal bisection and selecting an appropriate reference point, we can trivially derive either of these.[223] The astute reader will likely have realised by now that this is simply the application of the very same principle from which the Pythagorean scale is derived, here applied to the rhythmic domain. Collins speculates on the origins of Pythagorean ideas, pointing to Pythagoras's membership of an ancient snake cult, originally imported from North Africa, and concludes that "... Africa has provided the same musical octave arrangement of seven notes/pulses on twelve intervals twice over[:] once in melodic and once in rhythmic form." (Collins 1992:11)[224] Having opened the door, so to speak, it might prove fruitful to briefly consider other notable tuning methods, most particularly the system of Just Intonation (JI). Rather than applying a single ratio repeatedly, JI is characterised by simple, super-particular ratios, ostensibly reflecting the perfect consonance obtained between adjacent components in the natural harmonic series. Thus are employed the octave occurring between harmonics two and one (2:1), the perfect fifths between harmonics three and two (3:2), the perfect fourth between harmonics four and three (4:3), the major third between harmonics five and four (5:4), and so on. This would provide an alternative account of the

---

[223] The appropriate reference point, in the case of Figure C.14, is pulse 6/12. In the case of Figure C.15, it is pulse 1/12. Employing logarithmic trisection is also an option in both cases, given an appropriate reference point.

[224] Still, Collins' postulated parallel is based on surface features, with no attempt to relate such to a common physical principle, i.e. a "cognitive isomorphism". A similar link between Pythagorean and ancient Mesopotamian ideas is asserted by Kehinde (Kehinde 2011:99–100), this time with reference to cuneiform tablet-inscribed accounts of lyre tuning, apparently by a "cycle of fifths". The result hereof would have been recognisable as "Pythagorean" tuning.

"diatonic" appearance of the standard pattern, given a logarithmic view of rhythmic time. We must hasten to add, though, that the reported similarity between diatonic scales and rhythmic timelines should not be taken to imply that African timelines amount to the persistent playing of scales in African music.[225] In both cases, though even more so in the latter, it is the use of super-particular ratios which suggests a systematic exploitation of the tensions created by the ebb and flow between synchronous and asynchronous elements. Herein, we believe, lies the true significance of timelines to this study: that they combine different periodic processes in a way which directs perception toward some future point of alignment, hereby providing much of the perceived impetus in the music. This is directed motion. Its goal is alignment. Its effect lies in the expectations raised by our conscious or unconscious recognition of being drawn into entrainment.

In fact, parallels have been found between African timelines and all but one pentatonic, and one heptatonic permutation (Pressing 1983).[226] Yet, these have been dubbed "cognitive isomorphisms" precisely in recognition thereof that they are "products of human musical thinking" (Pressing 1983:38), there being no physical principle apparent to account for such parallels (logarithmic perception of time was nowhere suggested). More recently, such timelines have been distinguished from other possible permutations by phylogenetic principles of "rhythmic oddity" and "off-beatness" (Toussaint 2003, 2005), thereby arguing that these particular structures have been selected to maximise such formal attributes.[227] And yet, despite the obvious allure of such homonymicities, there are many rhythmic timelines which do not appear to map quite so easily.

```
     [ x . x . x . x . | x x . x . x . x ]
(16) [ . x . x . x . x | . . x . x . x . ]
     [ x . . . x . . . | x . . . x . . . ]
```

Figure C.16: -Luvale/-Lwena 9-stroke, 16-pulse pattern (Kubik 2010b:59)

Figure C.16 depicts a 16-pulse cycle,[228] which might immediately appear to be incompatible with the very notion of Pressing's "cognitive isomorphisms", since those would seem to be predisposed towards the division of the octave into 12 semitones. However, the logarithmic midpoint here is 0.585 x 16 = 9.36,

---

[225] As Agawu puts it: "... actual musical manifestation would discourage us from pressing the analogy too far: whereas the standard pattern is a fixed ostinato heard from the beginning to the end of a dance composition, the diatonic scale is rarely employed as an ordered intervallic pattern from beginning to end. Pressing himself meant the analogy – not homology – to be suggestive." (Agawu 2006:17)

[226] Pressing found no isomorphic mapping in either the case of the phrygian mode, nor the pentatonic scale from G# to G# (black notes only).

[227] Pressing showed that "... the standard pattern is almost maximally ambiguous, as it samples several different meters (12/8, 6/4, and 3/2) – and different phases of those meters – almost equally" (Temperley 2000:81; cf. Pressing 1983:46–47)

[228] Two phase-shifted variants of this pattern appear in a discussion of Mwenda Jean Bosco's guitar improvisations (Rycroft 1962a:100).

which lends some credence to the notion that the stroke at pulse 10/16 (nine pulses after pulse 1/16) might be perceived as such. If we do accept this, then iterative application of the same principle will produce the pattern shown. This then raises the question of tolerance. Sources maintain that the elementary pulse is perfectly even (both in temporal spacing and in accent), though none qualify such statements with any quantitative measure of tolerance (Koetting 1970:122; Kubik 2010b:33). In the absolute sense, of course, "perfectly even" is attainable by neither man nor machine, as Koetting (1970:123) does concede.[229] Rather, it is broadly acknowledged that systematic deviation from a metric ideal imparts "groove" to a rhythmic performance, and that such deviation is desirable to the extent that it "humanises" a mechanical performance. Perhaps closer inspection will find such deviations to be better understood, at least partially, in terms of overlapping periodic processes having different periods. In any event, for the purpose of this study, we note that some degree of approximation is typically required, but that such might, for instance, be found to be generally consistent with the kinds of approximations routinely employed in various tuning temperaments. In other words, approximation might be be symptomatic of a structural imperative to align different periods, as is encountered in Equal Temperament's "closing" of the cycle of fifths. Furthermore, approximation might reveal layering of stylistic concerns. Agawu suggests "binarization", for instance, as that kind of approximation which allows one to relate the older, precolonial 12/8 standard pattern to the more modern, postcolonial or neotraditional clavé pattern in 4/4 (Agawu 2006:38, cf. also citations therein; Gómez *et al.* 2007). In both of the aforementioned cases, approximation is expected to reveal underlying ideal structures, each encapsulating some facet(s) of "structural tension" within the music examined. According to Agawu, such "tension" might exist in the discrepancies between "... the effect of the normative metric grouping and that of the actual rhythmic patterns being exposed ... We might, in fact, postulate the presence of *metric dissonance*..." (Agawu 1986:69–70) However, Agawu's comments fall short of recognising the more systematic "polyrhythmic" tension being advocated above, which is perhaps better described by Rycroft's "'cadential' use of 'discord and concord'" (Rycroft 1954:24, 1961:82). On the other hand, systematic tempering of the above sort will likely account for only part of the variance observed between the ideal and its performed manifestation. The balance, we would suggest, should be considered "groove plus noise", being those deviations which embody both the performer's intended inflections and physical limitations. In the case of accomplished musicians, one would of course expect the former to significantly overshadow the latter, though Kubik nonetheless directs us to "... find out in every case what are *intentional* and what are *accidental* (tolerated) deviations from rhythmic regularity." (Kubik 1972:33)

Recognising the distinction between the intentions of the musician and the physical manifestations of

---

[229] In fact, Koetting presents the Time Unit Box System as follows: "TUBS represents the music on approximately the same level of precision that a trained observer hears; for those special analyses in which a more mathematically precise representation of rhythm and sonority is desired, elaborate mechanical tools such as the Seeger Model C Melograph and the Stroboconn would be appropriate." (Koetting 1970:126) Kubik, too, reports tolerance of considerable rhythmic inaccuracy in music intended as perfectly even, particularly at higher tempi (Kubik 1965:37).

those intentions (in the domains of both pitch and rhythm) brings us full circle, in the sense that we are once more confronted with the etic/emic dichotomy. Undoubtedly, limiting ourselves to only such information as can be gleaned from a digital recording (as is envisaged here) is bound to exclude important insights, as is made clear by Dargie's reference to “... an important and constantly encountered attitude of Xhosa musicians – that the listener is presumed to be able to experience everything that the performer experiences – and the evidence is that this is true of fellow-musicians and fellow-villagers.” (1991:36) On the other hand, analysis based on digital signal processing methods presents possibilities that are simply not feasible in typical CMN-derived transcription-based methods. Specifically, the ability to engage arbitrarily large datasets speaks to the core of the very notion of “universality”. However, given a generation of ethnomusicologists trained to dismiss all generalisations with a single counter-example (Nettl 2001:463), we should perhaps adopt the term “fuzzy universality” here, with all of its implications of probability, rather than outright certainty. We will now reflect on those principles which might speak to our goal.

C.5 Reflection

It would seem that the individual events that constitute a particular piece of African music, to whatever extent these are physically manifested, are typically sung or played with “percussive” attack. One might thus expect the onset of such events to be clearly evident in the audio signal. Pulse and cycle structures would be revealed by autocorrelation, and beats and timelines might then be inferred, to whatever extent this is possible, from the remaining events. In particular, the division of the cycle into multiple, overlapping grids, where the number of divisions in each grid is mutually prime, would appear to provide an appropriate search space against which individual events could be matched. To the degree that one can thus identify multiple overlapping cycles, each leading to the same future point of coincidence, we will suggest that directed motion has been uncovered. In the trivial case, hemiola-like structures (3 : 2) are to be expected, but the generality of the approach envisaged here holds within it the potential to expose more complex relationships. For one, logarithmic division of the cycle has already been shown to generate important timeline patterns, and as might be explored more fully within our implementation. Additional complexity will arise from the need to consider arbitrary phase shift of the various constituent cycles relative to each other, and possibly to consider more drastic variation from the rhythmic template. We are nevertheless of the opinion that all matters relating to “off-beating”/syncopation and additive/divisive rhythm will be adequately addressed in the course of the same method, without specific attention thereto. Nor will it be necessary to address the “phylogenetic properties of rhythmic oddity and off-beatness”, since these will be seen to be symptomatic of the imperative to combine cycles in this way.

We suspect that the very same principles are, to a great extent, at work in the spectral domain. Certainly, the combination of distinct frequencies at super-particular ratios is well-known as an account of consonance, and such is implied by the relating of notes to a common fundamental. It is also quite clear that absolute pitch has little value here. Though it may be useful to extract a "scale" in a given case, the relationship between tones therein is likely to be of far greater value than the actual frequencies employed, and this might require that we consider alternatives to the canonical FFT approach. Specifically, the presence of conjunct or overlapping tetrachords, apices and finals might shed light on "harmonic" tensions embedded in the music. Finals, in particular, are expected to coincide with rhythmic concord. However, some variance is once again to be expected from any theoretical ideal, whether originating in physical phenomena such as inharmonicity, in physiological limits to the consistency of performance, or in fluid cultural notions of pitch. The characteristic "pathogenic" contour of African melody is expected to assist in identifying phrases, and the typically greater variation between successive solos as compared to choruses might be helpful in distinguishing between these. It is expected that such will be found to correlate, in most cases, to shifting "tonal-harmonic centres". Phrase overlap and progressive phrase shift will introduce further complexities which are yet to be considered.

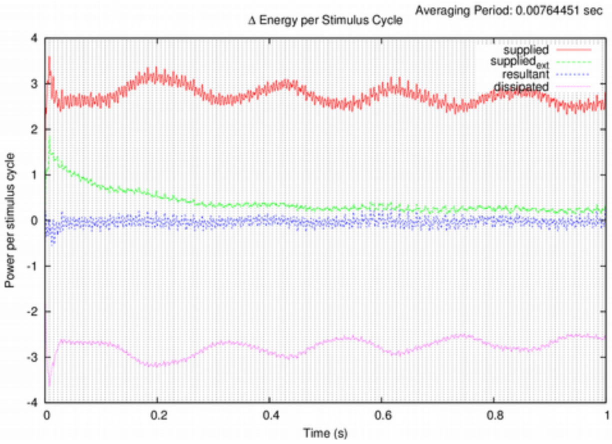
Further to the systematic temporal and spectral tensions playing out in any given piece of African music, one can expect variation on a more local scale, as determined by the performer. Such might play out as small deviations from the rhythmic or tonal patterns established by the overall piece, but we are nonetheless assured that these are, in all cases, stylistically restrained.

What remains, then, is to find a suitably tolerant, yet robust periodicity detector. In order to support the notion of directed motion as a "universal", this component must operate across most (if not all) of the audible bandwidth as a "pitch" detector, but equally so in the sub-audio band (perhaps down to 0.01 Hz) as a "beat" detector. Non-linear oscillators appear to possess many characteristics which endear them to this very task.

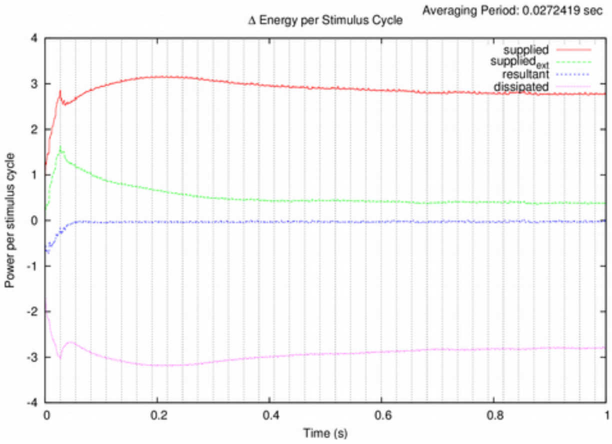
Appendix D

Experimental Data

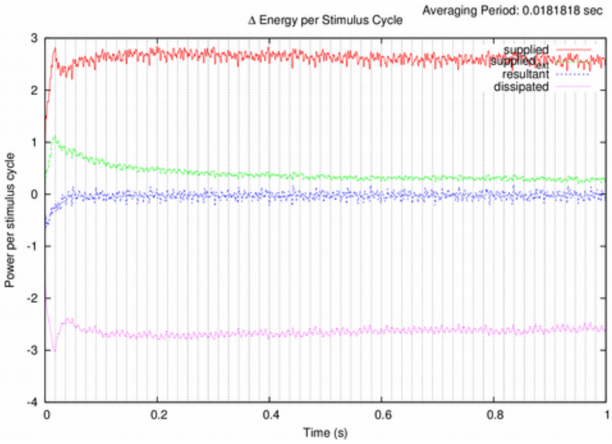
We here present plots of our experimental data obtained in the course of Experiment 3.



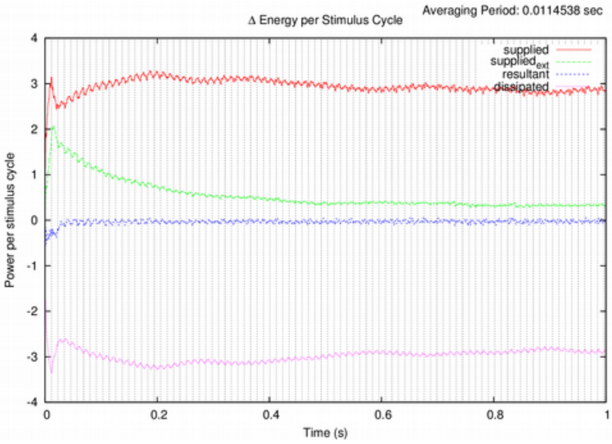
I



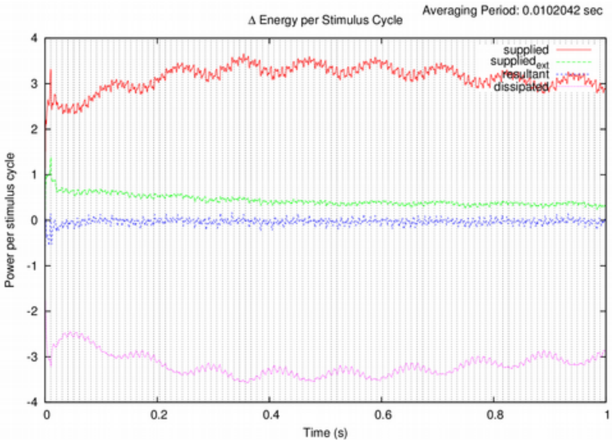
ii



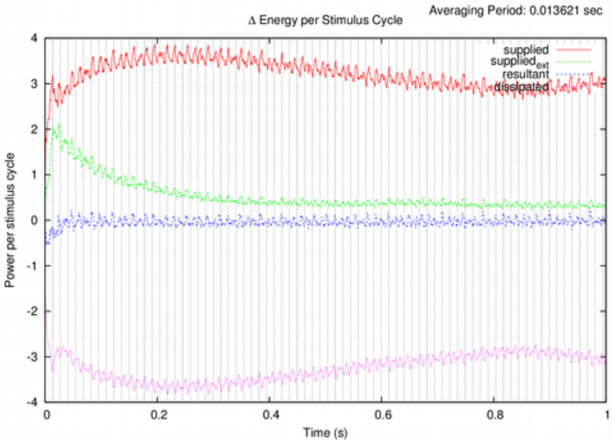
iii



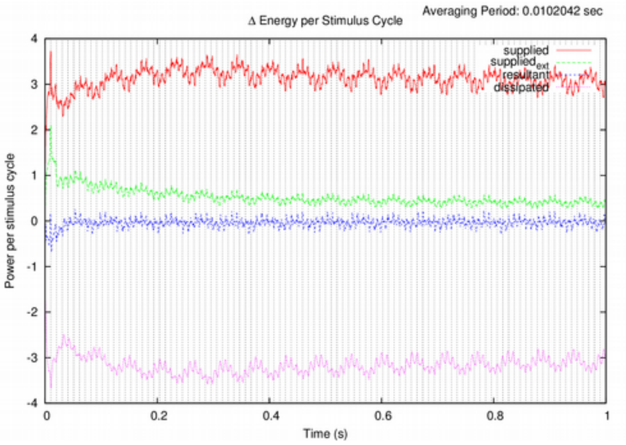
IV



V



vi



viiº

The data above directly supports the positive finding of this research. All other data is to be found in the project repository (see Appendix E).

Appendix E

Project Repository

Components supporting this research are hosted in a Subversion repository at:

`http://musictechnology.co.za/su_msceng_2016`

Specifically, the following folders:

`/data`
`  |-- /cx_nil`
`  |-- /cx_stat`
`  |-- /cx_dyn`
`  |-- /cx_pno`
	These folders contain the experimental results discussed in this thesis:
	• "cx_nil" refers to the absence of learning (Exp. 1)
	• "cx_stat" mimics typical "train-and-predict" behaviour (Exp. 2)
	• "cx_dyn" retains neural plasticity throughout (Exp. 3)
	• "cx_pno" introduces real-world timbres as stimuli (Exp. 4)

`/gfnnviewer`
`  |--/branches`
`  |--/tags`
`  |--/trunk`
`       |--README.txt`
`       |--/src`
	Following typical Subversion repository structure, the "src" folder contains the source tree for the application used to generate the experimental results. The "trunk" folder also contains the Eclipse project files, and should easily be importable into that IDE. Checkout instructions are included in the README.txt file.

`/pythons`	This folder contains various Python modules used in the initial explorations of this thesis.