Mariana Montiel · Octavio A. Agustín-Aquino ·
Francisco Gómez · Jeremy Kastine ·
Emilio Lluis-Puebla · Brent Milam (Eds.)

Mathematics and Computation in Music

8th International Conference, MCM 2022
Atlanta, GA, USA, June 21–24, 2022
Proceedings

Welcome
2
Atlanta

🦌 Springer

Lecture Notes in Artificial Intelligence 13267

Subseries of Lecture Notes in Computer Science

More information about this subseries at https://link.springer.com/bookseries/1244

Mariana Montiel · Octavio A. Agustín-Aquino ·
Francisco Gómez · Jeremy Kastine ·
Emilio Lluis-Puebla · Brent Milam (Eds.)

Mathematics and Computation in Music

8th International Conference, MCM 2022
Atlanta, GA, USA, June 21–24, 2022
Proceedings

② Springer

Editors
Mariana Montiel
Georgia State University
Atlanta, GA, USA

Octavio A. Agustín-Aquino
Technological University of the Mixteca
Huajuapan de León, Mexico

Francisco Gómez
Universidad Politécnica de Madrid
Madrid, Spain

Jeremy Kastine
Life University
Marietta, GA, USA

Emilio Lluis-Puebla
National Autonomous University of Mexico
Mexico City, Distrito Federal, Mexico

Brent Milam
Georgia State University
Atlanta, GA, USA

Preface

The 8th Biennial International Conference for Mathematics and Computation in Music (MCM 2022) took place during June 21–24, 2021, at Georgia State University in Atlanta, Georgia, USA. MCM 2022 continued the pattern, initiated in 2007 at the first MCM meeting, of biennial international conferences held on alternating sides of the Atlantic: Berlin in 2007, New Haven in 2009, Paris in 2011, Montreal in 2013, London in 2015, Mexico City in 2017, and Madrid in 2019.

As the flagship conference of the Society for Mathematics and Computation in Music (SMCM), MCM 2022 provided a platform for the communication and exchange of ideas among researchers in mathematics, informatics, music theory, musicology, and related disciplines. It brought together researchers from around the world who combine mathematics or computation with music analysis, music cognition, composition, and performance.

The schedule is available at https://mcm2022.org/. The scientific program featured 28 talks and 10 posters, as well as two panel sessions and two plenary sessions. The presentations were grouped around the following subjects: Mathematical Scale and Rhythm Theory: Combinatorial, Graph Theoretic, Group Theoretic, and Transformational Approaches; Categorical and Algebraic Approaches to Music; Algorithms and Modeling for Music and Music-Related Phenomena; Applications of Mathematics to Musical Analysis; Mathematical Techniques and Microtonality.

On the afternoon of June 23, the conference organizers planned a public outreach event at the Museum of Design Atlanta (MODA). The goals were to (1) engage the general public in the area of mathematics and computation in music, and (2) demonstrate to actual participants how effective outreach activities can be implemented.

Four concerts took place, the majority by SMCM researchers. On the first evening, Emmanuel Amiot, Moreno Andreatta, and Giles Baroin presented a public concert-lecture titled "Music and maths: the geometric match". The second concert was the performance Positive and Negative Spaces by the Terminus Ensemble of Contemporary Music. The third concert was part of the Homage to Jack Douthett, in which there were performances of Jack's own compositions for classical guitar by Octavio Alberto Agustín-Aquino, as well as performances by some of his closest colleagues, and a work by Thomas Noll that is based on Jack's research. The final concert was Emilio Lluis Puebla's concert-lecture, performing Rachmaninoff's Faust Piano Sonata Op. 28.

We received 45 submissions of which 27 long papers and 10 short papers were accepted. All papers were peer reviewed. The submissions came from researchers in 11 countries in North and South America, Europe, and Asia.

We thank the following institutions for providing their infrastructure and human resources for the organization and promotion of MCM 2022:

- Department of Mathematics and Statistics, Georgia State University
- Society for Mathematics and Computation in Music
- School of Music, Georgia State University

- College of Arts and Sciences, Georgia State University
- Museum of Design Atlanta (MODA)
- Universidad Politécnica de Madrid
- Universidad Nacional Autónoma de México
- Universidad Tecnológica de la Mixteca
- Life University

June 2022   Mariana Montiel
                                                              Octavio A. Agustín-Aquino
                                                                        Francisco Gómez
                                                                          Jeremy Kastine
                                                                    Emilio Lluis-Puebla
                                                                            Brent Milam

Organization

General Organizing Committee

Mariana Montiel	Georgia State University, USA
Jeremy Kastine	Life University, USA
Emilio Lluis-Puebla	UNAM, Mexico
Guerino Mazzola	University of Minnesota, USA
Brent Milam	Georgia State University, USA
Octavio Alberto Agustín Aquino	Universidad Tecnológica de la Mixteca, Mexico
Thomas Noll	Escola Superior de Musica de Cataluña, Spain
Robert Peck	Lousiana State University, USA

Scientific Program Committee

The Scientific Program Committee was responsible for the scientific content of MCM 2022. They coordinated and supervised the reviewing process for the submitted papers and prepared the final list of oral and poster presentations.

Mariana Montiel	Georgia State University, USA
Octavio Alberto Agustín Aquino	Universidad Tecnológica de la Mixteca, Mexico
Francisco (Paco) Gómez	Universidad Politécnica de Madrid, Spain
Jeremy Kastine	Life University, USA
Emilio Lluis-Puebla	UNAM, Mexico
Brent Milam	Georgia State University, USA

Local Organizing Committee

The Local Organizing Committee consisted of students and other local participants at the conference who helped with logistical issues.

Reviewers

Carlos Agon	Isabel Barbanchano
Octavio Agustín-Aquino	Giles Baroin
Giovanni Albini	Louis Bigo
Aitor Álvarez	Sonia Cannas
Emmanuel Amiot	Norman Carey
Moreno Andreatta	Rodrigo Castro López Vaal

Elaine Chew
David Clampitt
Richard Cohn
Jose Miguel Díaz Bañez
Andrée Ehresmann
Tom Fiore
Harald Fripertinger
Paco Gómez
Franck Jedrzejewski
Julien Junod
Jeremy Kastine
Carlos de Lemos Almada
Vicente Liern
Emilio Lluis-Puebla
Omar López Rincón
Maria Mannone

Mark McFarland
Brent Milam
Thomas Noll
Robert Peck
Alexandre Popoff
Roberto de Prisco
Lauren Ruth
Robert Schneider
Juan Sebastián Arias
Bill Sethares
Julius Smith
Florian Thalmann
Dmitri Tymoczko
Ciro Visconti
Jason Yust

Collaborating Institutions

Department of Mathematics and Statistics, Georgia State University
Society for Mathematics and Computation in Music
School of Music, Georgia State University
College of Arts and Sciences, Georgia State University
Museum of Design Atlanta (MODA)
Universidad Politécnica de Madrid
Universidad Nacional Autónoma de México
Universidad Tecnológica de la Mixteca

An Afternoon of Math+Music@MODA
(Abstract of Invited Talk)

At the final plenary session of MCM 2019, Paco Gómez discussed the need to promote our field of research through public outreach. In response, in conjunction with MCM 2022, we hosted an event at the Museum of Design Atlanta (MODA) designed to expose the general public to our areas of research, as well as to allow the members of our organization to learn from one another how to implement effective outreach activities in the future.

The event, called Math+Music@MODA, took place at MODA (see museumofd esign.org/) on the afternoon of June 23rd, 2022 from 1:30 PM to 4:30 PM. The following activities were presented:

- Jeremy Kastine (organizer of the event) presented an activity about composing canons with monophonic composite texture. Participants learned how this problem can be formulated in terms of finding maximal cliques of a graph.
- Thomas Noll presented "The Collective Public Fourier Performance." In this activity, three participants control Fourier coefficients by holding flags at varying heights, which are interpreted by mobile devices and processed by a central computer, producing a histogram that indicates how loudly each of seven other participants are to play their assigned note of a diatonic scale.
- Paco Gómez presented "Matherhythm or rhythm is a killer," which puts forward mathematical content - exact division, division with remainder, greatest common divisor, Euclid's algorithm and evenness principle - along with musical content - time span, pulse, rhythm formation, and timelines -, and shows how those mathematical ideas can be used as a tool to understand music and also as a principle for composing music. Participants were able to perform music based on these concepts using Boomwhackers.
- Luis Nuño presented an activity about his "Harmonic Wheel," a physical tool that combines a Tonnetz transformed into a polar grid with a plastic disc containing the lines that define the major, harmonic and melodic minor scales, together with the scale degrees and the symbols of the corresponding seventh chords. The Harmonic Wheel is a powerful and versatile tool for analyzing and composing music, as well as providing an efficient mnemonic notation.
- Maria Mannone presented an activity about the "CubeHarmonic," a novel musical instrument employing the concept of the triad Tonnetz through the physical manipulations of the Rubik's Cube. Participants experienced this instrument firsthand through two mobile apps developed by Maria's colleagues: Takashi Yoshino and Pascal Chiu.
- Gilles Baroin presented a collection of mathemusical virtual reality movies and interactive models. Participants used virtual reality headsets to clearly visualize concepts that would otherwise be difficult to explain and comprehend.

Contents

Algorithms and Modeling for Music and Music-Related Phenomena

Applications of Mathematics to Musical Analysis

Mathematical Techniques and Microtonality

Short Papers

Mathematical Scale and Rhythm Theory: Combinatorial, Graph Theoretic, Group Theoretic, and Transformational Approaches

A Set-Theoretic Model of Meter and Metric Dissonance

Richard Cohn[(✉)]

Department of Music, Yale University, New Haven, USA
`richard.cohn@yale.edu`

Abstract. A set-theoretic of model of musical meter is formalized, building up from time points to pulses to meters to metric relations. The model of metric relations formalizes work on metric dissonance, and refines the displacement/grouping taxonomy under current usage.

Keywords: Musical meter · Set theory · Metric dissonance · Syncopation · Hemiola · Polymeter

1 Introduction

During the 1970's, mathematical set theory emerged in North America as the predominant method for exploring and communicating the structure of the chromatic 12-note universe, emphasizing the properties of its chordal and scalar subsets, and the relations in which they participate [1, 2]. The paradigm was productively transferred onto the diatonic 7-note universe [3], and eventually onto patterns of musical time, with a focus on cyclic rhythms in universes of eight or more beats [4–7]. These cycles are implicitly metric, in the sense that they project at least a unit pulse, and a slower pulse that marks points of cyclic orientation or renewal. But the metric models are minimal, in the sense that they track only those two pulses. In the European tradition, meter is deep, projecting at least three pulses of different speeds, and sometimes as many as seven [8]. The predilection in Asia and Africa, as well as in globally circulating popular and electronic dance music, for cycle-lengths that are powers of 2 or multiples of 6, together with the propensity to bodily entrain (dance) at a rate that is faster than unit and slower than cycle, suggests that deep meter is a broader phenomenon [9–11].

Maury Yeston laid the groundwork for a set-theoretic model of deep meter by defining meter as an inclusion relation between pulses; the transitivity of inclusion invites a recursive application [12]. Lerdahl and Jackendoff further prepared the terrain by defining pulses as sets of time points, and by representing meters as dot arrays with a depth dimension [13]. Using these formulations as a foundation, the first half of this paper proposes a set-theoretic model of deep meter, integrating the perspective of recent studies in the psychology of metric induction.

The second half of the paper builds a model of relations between distinct deep meters. Metric relations, as defined here, are equivalent to what music theorists commonly refer

to as metric *dissonance*: the superposition or juxtaposition of distinct, partly incommensurable meters, which have the capacity to scramble patterns of neural activation [14, 15]. Since the 14th century, metric relations have been associated in Europe with psychological or semantic states such as difficulty, disorientation, conflict, mental instability, and yearning for the unattainable [16–19]. Metric relations are the basis of metric modulations, which are essentially local processes of metric change, and metric form, patterns of metric change across relatively long stretches of musical time. They thus are of central concern to musical analysts who recognize rhythm and meter as dynamic elements of musical structure and experience.

2   A Set Theoretic Model of Meter

The model of metric relations is built up in three stages, each of which converts sets of the previous stage into elements of the subsequent one. Time points are elements of pulses, which are elements of meters, which are elements of metric relations.

2.1   Time point and span

The axiomatic elements are *time points*, which have no properties other than their temporal addresses [20, 21]. Pairs of distinct elements $x < y$, where $<$ represents temporal precedence, give rise to *time spans*, evaluated as $(y - x) > 0$.

   Time points are distinct from the musical events whose onsets mark them [22]. Events and their onsets are *res extensae* that exist in musical sound. Time points are *res cogitantia* that exist in the mind. The correspondence of time points to musical events is not 1:1. There are musical events, such as grace notes, that do not mark time points [13]. Conversely, there are *virtual* time points that are unmarked by musical events [23]. The onsets themselves are smeared across spans, or bins [24], and reduced to points through the mental operation of *quantization* [25].

   Time-span sizes are mentally assigned rather than prosthetically measured, and give rise to comparative rather than absolute values. Augustine of Hippo wrote in the 4th century that "I confidently answer—insofar as a trained ear can be trusted—that this syllable is single and this double…. It is in you, oh mind of mine, that I measure the periods of time" [26].[1] If two adjacent time spans are brief, our mind spontaneously determines whether they are equal. If unequal but integrally proportioned, we subitize the number of concatenated shorter spans that fit the longer span, so long as that number is small.

2.2   Pulse

Definition 1. A pulse P is an ordered set of time points whose adjacent elements are separated by a constant value, $\tau(P)$, the pulse's period.

---

[1] Book 11, Chapter 28, paragraph 34.

The constant-value constraint is often referred to as the *isochrony* property. The property applies to time points, not to the musical events that mark them, which may be literally isochronous (if machine-generated), or notionally isochronous (if human-generated, like a drumbeat), or neither, if the series includes virtual time points (as do most songs and instrumental compositions). It is important to keep in mind this last possibility, since it is easy to default to a prototypical conception of a pulse as a parade of isochronous onsets emitted from a uniform auditory source. Often pulses have gaps, as when a series of alternating half and quarter notes induces a quarter-note pulse, or their isochronous onsets are split into multiple streams, as when a quarter-note pulse is induced by anti-phased half-note pulses in the prototypical rock drummer's alternation of bass and snare.

Pulse periods, equivalent to slow frequencies, can be represented by absolute durations, with Herz or milliseconds, or by relative duration, with standard symbols, or counts of beats or measures.

2.3 Meter

A meter is a set of pulses related by inclusion. It is useful to divide the study of meter into minimal meters, with exactly two pulses, and deep meters, with three or more pulses. Minimal meters are to deep meters as intervals are to chords.

2.3.1 Minimal Meter

Definition 2. A *minimal meter* M is a pair of pulses (P_1, P_2) such that $P_1 \subset P_2$. The definition implies $\tau(P_1) > \tau(P_2)$, and thus that P_1 is *slower* and P_2 *faster*.

The ordering of pulses from slowest to fastest will be preserved as the model develops.

A minimal meter is classified by a function $\beta(M) = \frac{\tau(P_1)}{\tau(P_2)}$, which evaluates the ratio of its constituent pulses. By Definition 1, both pulses are periodic, and by Definition 2, they are related by inclusion. Thus the time points of P_1 are periodic selections of the time points of P_2, and the range of $\beta(M)$ is the positive numbers >1.

Two classes of minimal meters merit special attention:

Definition 3. A minimal meter M is *duple* if $\beta(M) = 2$, and *triple* if $\beta(M) = 3$.

Definition 4. A minimal meter is *normal* if it is either duple or triple.

"Normal" designates a class of meters that have historical significance in Europe and the Americas, and perhaps elsewhere. I intend it as a technical term that implies no judgement of value.

2.3.2 Deep Meter

Definition 5. A *deep meter* M is a set of three or more distinct pulses (P_1, P_2, …, P_k), ordered from slowest to fastest, such that every pair of pulses forms a minimal meter.

A deep meter M is classified by a function, $\beta(M) = \frac{\tau(P_1)}{\tau(P_2)}, \frac{\tau(P_2)}{\tau(P_3)}, \ldots, \frac{\tau(P_{k-1})}{\tau(P_k)}$ whose range consists of multisets of numbers, each of which classifies an adjacent minimal meter, ordered from slowest to fastest. Consider for example M' = (4.5 s, 1.5 s, .75 s, .375 s, .125 s), which also might be notated as (dotted breve, whole, half, quarter, tripleted eighth) or as (36, 12, 6, 3, 1). This meter has five distinct pulses, hence four adjacent minimal meters, whose classification is $\beta(M') = (3\ 2\ 2\ 3)$.

18th-century European compositional theory stipulates that a meter is defined not as a list of pulses, but rather as a selection of two orienting pulses: a tactus, or counting pulse, and a downbeat pulse [27]. In this view, M' (as defined above) is not yet a meter. It is, rather, a genus that gives rise to a multitude of specific meters. For example, depending on which durational values are selected and which pulses are prioritized, M' could be represented using, among others, the following 18th-century meter signatures: $\frac{3}{1}, \frac{6}{2}, \frac{2}{4}, \frac{2}{2}, \frac{4}{4}, \frac{3}{8}, \frac{6}{8}$, and $\frac{12}{8}$, representing eight distinct meters. That historical tradition has left a strong residue in modern textbooks, which classify "meters" according to their signatures, and more subtly in theoretical and perceptual research, which often nominates a single pulse as the tactus [13, 19, 28], implying that a change of tactus, all else invariant, is a change of meter. The view taken here is that reference pulses are external to a model of meter. Meter is a system of relations, which need not be directed or oriented. A change of reference pulse is not a change of meter, and two listeners inducing the same pulses are hearing the same meter, even if their awareness or bodily response is oriented to different speeds.

2.3.3   Properties of Deep Meters

Among the properties of deep meters M = $(P_1, P_2, \ldots, P_k)$ are the following:

Definition 6. The slowest pulse in M, P_1, is its *span pulse*.

Definition 7. The fastest pulse in M, P_k, is its *unit pulse*.

Definition 8. The *cardinality* of M is k.

Definition 9. The *size* of M is the ratio of its unit and span pulses, $\frac{\tau(P_1)}{\tau(P_k)}$, which is equal to the product of the elements in its ordered set $\beta(M)$, $\prod_{n=1}^{k-1} \frac{\tau(P_n)}{\tau(P_{n+1})}$.

Definition 10. A meter M is *saturated* if all elements of $\beta(M)$ are prime numbers.

If a meter M is saturated, then it has no interior gaps in M which could be filled by additional pulses. Saturated meters are particularly significant because of our propensity to fill gaps in the pulse spectrum through spontaneous processes of subjective metricization [28].

Definition 11. A deep meter is *normal* if each of its adjacent minimal meters is normal; equivalently, it is classified a multiset of 2's and 3's exclusively.

2.4 Metric Relations

Definition 12. Two saturated meters are *related* if (1) their unit pulses are identical, $x_k = y_k$, and (2) their span pulses have equal periods, $\tau(x_1) = \tau(y_1)$.

Since both conditions are based on equality, the defined relation inherits the properties of symmetry, reflexivity, and transitivity proper to equality relations.

The first condition insures that X and Y are drawing from a common universe of time points. Recall that not all time points of the unit pulse need be onset-marked. A unit pulse might be a proper superset of two onset-marked pulses in different streams, or presented at different times.

Since $x_1 = y_1$ implies $\tau(x_1) = \tau(y_1)$, the two conditions together entail that X and Y are of equal size. The saturation condition insures that X and Y also have equal cardinality k, permitting disjoint pairing by shared subscript, and thus by identical or similar speed.

Definition 13. Given two saturated meters $P = (P_1, P_2, ..., P_k)$ and $Q = (Q_1, Q_2, ..., Q_k)$ that are related as in Definition 11, pulses p_n and q_n are *associated*, for $1 \leq n \leq k$.

This term will be central to the work carried out in the next section.

3 Kinds of Metric Relations

In recent North-American scholarship, when two distinct meters of equal cardinality are combined simultaneously or successively, they are said to create **dissonances**, which are partitioned into two principal classes, **displacement** and **grouping** [19, 29]. Displacement covers situations often referred to as syncopation, anti-phasing, turning the beat around, and shadow meter. Grouping covers hemiolas and polymeters. I will substitute other terms for "grouping" and "displacement," because both are rooted in properties and conceptions specific to early-modern Europe,[2] and thus introduce impertinent implications for other repertories. Nevertheless, the following classification of metric relations draws the boundary at the same location. Grouping dissonances are further distinguished as simple and complex [31]. Recent work [32–34] identifies and models a third class that combines aspects of displacement and grouping dissonance; the taxonomy developed here defines a hybrid genus that is consistent with that work.

I propose classifying metric relations by a procedure whose components are summarized here, and detailed in the remainder of this paper. First, associated pairs of individual pulses are classified as identical, co-periodic or anti-periodic. Second, metric relations

---

[2] When pulses of the same period but different phases are combined, a displacement model involves determining (a) which of the two pulses is the source, and (b) whether the copy is displaced in the positive or negative direction. The determination is sometimes arbitrary for music that lacks pitch, or whose pitch-combinations don't adhere to historical European principles of dissonance regulation. Grouping is avoided because it is not a property of meters, whose pulses consist of time points that cannot be grouped into "larger time points." The conception of slow pulses "grouping" faster ones is based on a malformation that has likely origins in ancient Greek theories of poetic meter. I elaborate this point in [30].

are classified as ordered sets of associated pulse-pair classes. Finally, the classes of metric relations are mapped onto three genera, which are equivalent to the three classes identified in recent literature on metric dissonance: displacement, grouping and hybrid.

3.1 Pulse-Association Classes

Associated pulse pairs (x_n, y_n) are mapped to one of three classes by a function $\varphi(x_n, y_n)$.

Definition 14. If $x_n = y_n$, then the pulses are *identical*, and $\varphi(x_n, y_n) = $ IDENT.

Definition 15. If x_n and y_n are not identical, but their periods are equal, then they are *properly co-periodic*, and $\varphi(x_n, y_n) = $ CO-P.

"Properly co-periodic" excludes the trivial case where the pulses are identical; as with proper inclusion, this licenses one to cut locutionary corners by dropping the adverb. Pulses that are properly co-periodic share no time points.

Definition 16. If x_n and y_n have unequal periods, then they are *anti-periodic*, and $\varphi(x_n, y_n) = $ ANTI-P.

3.1.1 Constraints on Associate-Pulse Classes

Definitions 13–15 suggest a reformulation of Definition 11: two saturated meters (X, Y) are related if $\varphi(x_1, y_1) \neq$ ANTI-P and $\varphi(x_k, y_k) = $ IDENT. The first constraint is motivated by the observation that when $\varphi(x_n, y_n) = $ ANTI-P, then their intersection set $x_n \bigcap y_n$ is a slower pulse which is an element of X, Y, or both.

An additional constraint governs associated pulses of intermediate speed: if some associated pulse-pair is co-periodic, then all slower associated pairs, up to and including the span pulse, are co-periodic as well. Consider some associate pair $\{x_n, y_n\} \mid 1 < n < k$, and a slower associate pair, $\{x_m, y_m\} \mid m < n$. By definition of meters X and Y, $x_m \subset x_n$ and $y_m \subset y_n$. Assume now that $\varphi(x_n, y_n) = $ CO-P. Then $x_n \bigcap y_n = \emptyset$. Accordingly, $x_m \bigcap y_m = \emptyset$, and thus $\varphi(x_m, y_m) = $ CO-P.

3.2 Metric-Relation Classes

The classification system for associated pulse pairs (x_n, y_n) serves as the basis for classifying the relation between the meters (X, Y) of which they are elements.

Definition 17. Let (X,Y) be meters related as in Definition 11. Then the relation X R Y is classified by a function $\varphi(X, Y) = (\varphi(x_1, y_1), \varphi(x_2, y_2), ..., \varphi(x_k, y_k))$, whose image is a multiset, or *string*, of pulse-association classes.

If the classification of associated pulses were unconstrained, we would quickly suffer a combinatorial explosion of metric-relation classes. Fortunately, the constraints already adopted, plus one additional one proposed below, filter out most combinations. To review:

1) All strings end with IDENT;
2) No string begins with ANTI-P;
3) CO-P is preceded only by CO-P.

A fourth constraint is adopted to eliminate redundancy caused by adjacent identity relations. Since consecutive identity-pairs $(IDENT, IDENT) = (IDENT)^2$ are structurally no different than a single identity-pair IDENT, they do not profit from independent investigation. This motivates the mapping in (4), which reduces the cardinality to one that was already inventoried at a previous level of k.

(4) $(IDENT)^n \rightarrow IDENT$, for n > 1.

The strings that survive these filters, up to a metric depth of five pulses, are listed in Table 1, where they are labelled from (a) to (m). Superscripts count consecutive repetitions of a term. The comments in the final column serve as the basis for assigning the strings to genera of metric relations in the next sub-section.

Table 1. Metric-Relation Classes $\varphi(X, Y)$ for meters up to a depth of five pulses

k =	Label	$\varphi(X, Y) =$	Comments
1	(a)	IDENT	Universal root
2	(b)	(CO-P, IDENT)	Root of co-periodic genus
3	(c)	$((CO\text{-}P)^2, IDENT)$	Left-extension of (b)
	(d)	(IDENT, ANTI-P, IDENT)	Root of anti-periodic genus
	(e)	(CO-P, ANTI-P, IDENT)	Root of hybrid genus
4	(f)	$((CO\text{-}P)^3, IDENT)$	Left-extension of (b)
	(g)	$(IDENT, (ANTI\text{-}P)^2, IDENT)$	Internal expansion of (d)
	(h)	$(CO\text{-}P, (ANTI\text{-}P)^2, IDENT)$	Internal expansion of (e)
	(i)	$((CO\text{-}P)^2, ANTI\text{-}P, IDENT$	Left extension of (e)
5	(j)	$((CO\text{-}P)^4, IDENT)$	Left-extension of (b)
	(k)	$(IDENT, (ANTI\text{-}P)^3, IDENT)$	Internal expansion of (e)
	(l)	$((CO\text{-}P)^2, (ANTI\text{-}P)^2, IDENT)$	Left extension of (h); internal expansion of (i)
	(m)	(IDENT, ANTI-P, IDENT, ANTI-P, IDENT)	Elided concatenation of (d)

3.3   Three Genera of Metric Relations

Through a procedure to be described in this section, the thirteen classes listed in Table 1 reduce to three genera, one of which comes in two species.

Definition 18. The *co-periodic genus* consists of strings of the form ((CO-P)^a, (IDENT)^b), for a, b ≥ 1.

The co-periodic genus corresponds to what is characteristically referred to as extended syncopation, anti-phasing, shadow meter, turning the beat around, or displacement dissonance.

Definition 19. The *anti-periodic genus* consists of strings of the form ((IDENT)^a, (ANTI-P)^b, (IDENT)^c), for a, b, c ≥ 1.

The anti-periodic genus corresponds to what is referred to as hemiola, polymeter, and grouping dissonance. If b = 1, as is the usual case, the relation is simple. This is the hemiola that is familiar from Baroque cadences in triple meter, from 3-over-2 polymeters, and from the alternation of ternary and binary subdivisions of half-measures, which arise characteristically, for example, in Korean *pung'mul* and Spanish *flamenco* [8, 35]. If b > 1, the hemiola is complex [31, 36]. This formation involves other co-prime polymeters, including 4-over-3, 9-over-2, 9-over-8, and so forth.

Definition 20. The *hybrid genus* consists of strings of the form ((CO-P)^a, (ANTI-P)^b, (IDENT)^c), for a, b, c ≥ 1.

These are equivalent to the hybrid forms identified in [32–34]. Most of the examples analyzed in those writings are composed by Johannes Brahms, suggesting that an alternative label might be the *Brahms Genus*.

Leaving aside the universal root (a), the only entry in the table that does not fit this taxonomy is the final one, (m), where a pulse of intermediate speed functions simultaneously as the span pulse of a fast simple hemiola and as the unit pulse of a slow simple hemiola. As this is an elided replication of an existing genus, I am reluctant to establish a new genus to contain it. Otherwise, I conjecture that the three-fold taxonomy introduced here exhaustively covers metric relations of yet greater depth (k > 5), which are (in any case) of diminishing frequency since they approach the limit of the number of pulses that can be simultaneously tracked or entrained [28].

4 Extensions

This paper has adopted several limitations which could be loosened in future work. First, saturated meters need not be normal; higher primes could be substituted for 2's and 3's without affecting other aspects of the model [12]. The simplest case would be anti-metric 5-over-2, as in Holst's Mars movement from "The Planets," and the finale of Ravel's string quartet. Second, relations among three or more meters, which have been noted in the analytical literature, could be explored. Figure 1 sketches some of the simplest possibilities:

(a) Three meters, each pair of which is simply anti-periodic [12].
(b) Three meters, each pair of which is co-periodic. For an example from Schumann, see [18].

(a) Three simple anti-periodic pairs

(b) Three co-periodic pairs

(c) A product network of anti-periodic and co-periodic pairs, with hybrid pairs on the diagonals

(d) Two simple (3:2) and one complex (4:3) anti-periodic pairs
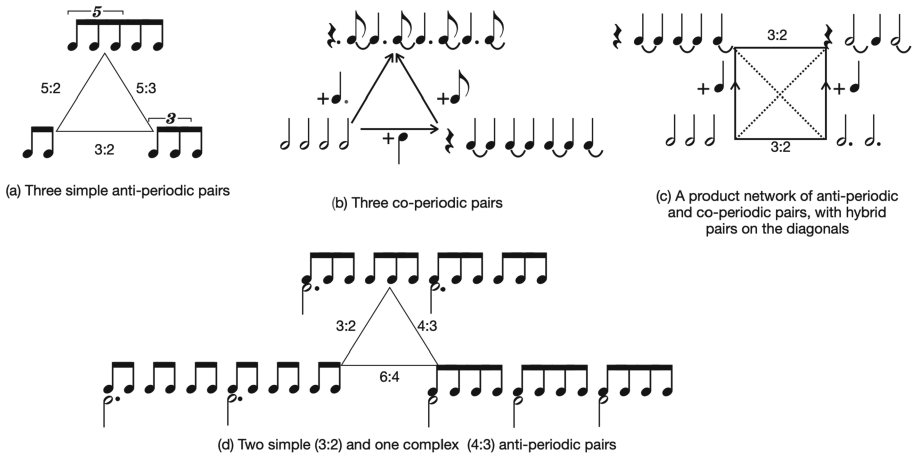
Fig. 1. Four portraits of multiple metric relations.

(c) Four meters in a product network [4], with anti-periodic pairs on one axis, co-periodic pairs on the other, and hybrid pairs on both diagonals.

(d) Three meters, two pairs of which are simply anti-periodic in a 3:2 ratio, the third pair of which is complexly anti-periodic in 4:3 ratio. For examples from Dvorak and Brahms, see [31].

There may be some motivation to regard (c) and (d) as underlying two of the identified genera. For any pairing of meters at opposite vertices of (c), one or both of the meters at the complementary vertices may be implicitly present, even if not explicitly articulated. Thus, any binary hybrid metric relation may be viewed as the product of a co-periodic relation and an anti-periodic one, a quaternary design whose intermediate pulses may be concealed or under-articulated. Similarly, any binary complex anti-periodic relation between two pulses may be viewed as the elision of *n* simple relations, a ternary design with the *n – 1* intermediate terms (*Vermittlungen*) elided out [36, 37]. Underlying these structural proposals is a *Gestalt* hypothesis about cognition: that explicit gaps in a well-defined structure are imagined to be notionally present. This same hypothesis is invoked at earlier levels of the model, where it was posited that gapped sets of onset-marked time points are completed by virtual time points, and that gapped sets of pulses on the speed continuum are filled (saturated) by processes of subjective metricization.

References

1. Babbitt, M.: Set structure as a compositional determinant. J. Music Theory **5**(1), 72–94 (1961)
2. Forte, A.: The Structure of Atonal Music. Yale University Press, New Haven (1973)
3. Clough, J.: Aspects of diatonic sets. J. Music Theory **23**(1), 45–61 (1979)
4. Lewin, D.: Generalized Musical Intervals and Transformations. Yale University Press, New Haven (1987)
5. Morris, R.: Composition with Pitch Classes. Yale University Press, New Haven (1987)

6. Cohn, R.: Transpositional combination of beat-class sets in Steve Reich's phase-shifting music. Perspect. New Music **30**, 146–177 (1992)
7. Rahn, J.: Turning the analysis around: Africa-derived rhythms and Europe-derived music theory. Black Music Res. J. **16**, 71–89 (1996)
8. Cohn, R.L.: The dramatization of hypermetric conflicts in the Scherzo of Beethoven's ninth symphony. Ninet.-Century Music **15**, 188–206 (1992)
9. Hesselink, N.: Rhythm and folk drumming (Pung'mul) as the musical embodiment of communal consciousness in South Korean village society. In: Tenzer, M., Roeder, J. (ed.) Analytical and Cross-Cultural Studies in World Music, pp. 263–287. Oxford University Press, New York (2012)
10. Tenzer, M.: Gamelan Gong Kebyar: The Art of Twentieth-Century Balinese Music. University of Chicago Press, Chicago (2000)
11. Arom, S.: Time structure in the music of Central Africa: periodicity, meter, rhythm and polyrhythmics. Leonardo **22**(1), 91–99 (1989)
12. Yeston, M.: The Stratification of Musical Rhythm. Yale University Press, New Haven (1976)
13. Lerdahl, F., Jackendoff, R.: A Generative Theory of Tonal Music. MIT Press, Cambridge (1983)
14. Mayville, J.M., Jantzen, K.J., Fuchs, A., Steinberg, F.L., Scott Kelso, J.A.: Cortical and subcortical networks underlying syncopated and synchronized coordination revealed using FMRI. Hum. Brain Map. **17**(4), 214–229 (2002)
15. Vuust, P., Roepstorff, A., Wallentin, M., Mouridsen, K., Østergaard, L.: It don't mean a thing…: keeping the rhythm during polyrhythmic tension activates language areas (Ba47). Neuroimage **31**(2), 832–841 (2006)
16. Aluas, L.F.: The 'Quatuor Principalia Musicae': a critical edition and translation with introduction and commentary. Ph.D. dissertation, Indiana University (1996)
17. DeFord, R.: Tactus, Mensuration, and Rhythm in Renaissance Music. Cambridge University Press, Cambridge (2015)
18. Malin, Y.: Metric displacement dissonance and romantic longing in the German lied. Music. Anal. **25**(3), 251–288 (2007)
19. Krebs, H.: Fantasy Pieces: Metric Dissonance in the Music of Robert Schumann. Oxford University Press, New York (1999)
20. Aristotle: Physics, Book 4. Translated by W. D. Ross. Clarendon Press, Oxford (1960)
21. Kramer, J.: The Time of Music. Schirmer Books, New York (1989)
22. Boone, G.M.: Marking Mensural Time. Music Theory Spectrum **22**(1), 1–43 (2000)
23. Hasty, C.: Meter as Rhythm. Oxford University Press, New York (1997)
24. Danielsen, A.: Here, there and everywhere: three accounts of pulse in D'Angelo's 'left and right'. In: Danielsen, A. (ed.) Musical Rhythm in the Age of Digital Reproduction. Ashgate, Burlington, pp. 19–36 (2010)
25. Desain, P., Honing, H.: The formation of rhythmic categories and metric priming. Perception **32**(3), 341–365 (2003)
26. Augustine: Confessions and Enchiridion. Translated by Albert Outler. Westminster Press, Philadelphia (1955)
27. Mirka, D.: Metric Manipulations in Haydn and Mozart: Chamber Music for Strings, 1787–1791. Oxford University Press, New York (2009)
28. London, J.: Hearing in Time: Psychological Aspects of Musical Meter, 2nd edn. Oxford University Press, New York (2012)
29. Kaminsky, P: Aspects of harmony, rhythm, and form in Schumann's Papillo, Carnaval, and Davidsbündlertänze. Ph.D. dissertation, The University of Rochester (1989)
30. Cohn, R.: An Analytical Model of Musical Meter (2022, in preparation)
31. Cohn, R.: Complex hemiolas, ski-hill graphs and metric spaces. Music. Anal. **20**(3), 295–326 (2001)

32. Samarotto, F.P.: 'The body that beats': review of Harald Krebs. Fantasy pieces: metrical dissonance in the music of Robert Schumann. Music Theory Online **6**(4) (2000)
33. Chung, M.: A theory of metric transformations. Ph.D. dissertation. University of Chicago (2008)
34. Popoff, A., Yust, J.: Meter networks: a categorical framework for metrical analysis. J. Math. Music **16**(1), 29–50 (2022)
35. de Cisneros Puig, B.J.: Discovering flamenco metric matrices through a pulse-level analysis. Anal. Approach. World Music **6**(1), 1–24 (2017)
36. Murphy, S.: On Metre in the Rondo of Brahms's Op. 25. Music. Anal. **26**(3), 323–353 (2007)
37. Leong, D.: Humperdinck and Wagner: metric states, symmetries, and systems. J. Music Theory **51**(2), 211–243 (2007)

New Insights on Diatonicity and Majorness

Franck Jedrzejewski[(✉)]

Université Paris Saclay/CEA, Orsay, France
franckjed@gmail.com

Abstract. This article contributes to the study of diatonicity in tunings with N equal divisions of the octave. The new definition we are proposing of a diatonic scale is based on two concepts. The first is well known since it concerns generated scales studied by Norman Carey and David Clampitt. The second is much less known. It is based on the sets of progressive transposition introduced by the French composer Alain Louvier. From these two concepts, we formulate a new definition of microdiatonic scales which is entirely characterized by two fundamental parameters. Then we define the majorness of a scale, by introducing an interval equivalent of the tritone by using limited transposition sets. We conclude this article by observing that under these conditions diatonicity and majorness are two different characters which do not necessarily exist in all tunings.

Keywords: Diatonic scales · Diatonicity · Diatonic theory · EDO · Major scales · Relative minor scales · Microintervals

Is there an analogue to the standard major or minor scales in a given equal division of the octave with N notes rather than 12? The present article addresses this question on the basis of mathematical and music-theoretical arguments set up by the French composer Alain Louvier under the name of *modes of progressive transposition*. It explores cross-connections to related work by other authors and especially with Neo-Riemannian theories. In certain ways, this article takes up the concepts given around the concept of diatonicity in [11,12] in a simpler but different form, and consider what the concept of tonality could be. As already mentioned in the cited article, a historically and theoretically interesting source to this question is the Wyschnegradsky's *24 Preludes* opus 22 composed in 1916. Diatonic modelization has been studied by many authors: E. Agmon [1,2], G. Balzano [3], and M. Broué [4].

In the first section, we study sets of progressive transposition in the light of David Lewin's injection number and interval function, and show their unicity. In the second section, we define and state the properties of microdiatonic scales. In the third and last section, we introduce a generalized concept of major keys, based on the existence of a generalized tritone. We also defined a generalized minor scale for a particular generalized major scale. In most cases, the results

are based on the uniqueness of scales with certain properties. As we wanted to distinguish major chords from minor chords, all classes are considered only up to transposition (and not up to inversion).

1 Progressive Transposition Scales

Introduced by Alain Louvier in [16], *sets of progressive transposition* are particular cases of Douthett's P-Cycles [9,10]. Louvier had a brilliant musical career. In 1968, he won the last annual Prix de Rome for musical composition. He has been director of the Conservatoire de Paris (1986–1991), where he also taught music analysis and orchestration (1991–2009). In his paper, he explains how to build sets of progressive transposition (PT) in the quartertone and thirdtone universe. More generally, PT-sets are defined in the N-equal division of the octave.

Definition 1. *Let $1 < m < N$ and $\gcd(m, N) = 1$. The set X is a set of progressive transposition if the transposition of X at level m has only one note of difference with X,*

$$\operatorname{card}(X \cap T_m(X)) = \operatorname{card}(X) - 1$$

This also means that there exist $x, y \in \mathbb{Z}_N$, such that $T_m(X) = (X \backslash \{x\}) \cup \{y\}$. In particular, if X is a PT-set, then X has N different transpositions and X is not a set of limited transposition.

Most of the properties of PT-sets can be established using the properties of the injection number and the interval function ifunc, introduced by Lewin in [15]. Let $(S, G, \operatorname{int})$ be a generalized interval system. The *injection number* of X into Y for f, denoted $\operatorname{inj}(X, Y)(f)$ is the number of elements x of X such that $f(x) \in Y$

$$\operatorname{inj}(X, Y)(f) = \operatorname{card}\{x \in X, f(x) \in Y\} = \sum_{x \in X} 1_{(f(x) \in Y)}$$

It is easy to show that if f is a permutation of S, then

$$\operatorname{inj}(X, Y)(f) = \operatorname{card}(f(X) \cap Y)$$

and

$$\operatorname{inj}(X, Y)(f) = \operatorname{inj}(X, Y)(f^{-1})$$

If f is the transposition $T_n(x) = x + n \mod N$, since T_n is a bijection, the injection number is linked to the interval function by the relation

$$\operatorname{inj}(X, Y)(T_n) = \operatorname{card}(T_n(X) \cap Y) = \operatorname{ifunc}(X, Y)(n)$$

The interval vector is defined for $n \in \mathbb{Z}_N$ by its coordinates,

$$\operatorname{iv}(n) = \operatorname{ifunc}(X, X)(n) = \operatorname{card}(X \cap T_n(X))$$

where the first coordinate is the cardinality of X since $\operatorname{ifunc}(X, X)(0) = \operatorname{card}(X)$.

Proposition 1. *X is a PT-set if and only if its complement X^c is a PT-set.*

Proof. It follows from the property of the injection number that if X is a PT set at level m,

$$\text{card}(X \cap T_m(X)) = \text{inj}(X, X)(T_m) = \text{card}(X) - 1$$

That implies that a single element x of X is mapped by T_m in the complement of X. And since T_m is a bijection, $\text{card}(X) - 1$ elements of the complement X^c are mapped to itself, and one element to X. Thus the complement of X is a PT set at level m. Conversely, the same reasoning applies to the complement. Therefore, X is a PT set at level m is an equivalence with X^c is a PT set at level m, since the complement of X^c is X.

A set of progressive transposition is invariant by transposition. The simple example is the 12 major (or minor) keys, in the common 12-EDO.

Proposition 2. *X is a PT-set if and only if all of its transposition is also PT.*

Proof. Let X a PT-set for index m. Then for all $n \in \mathbb{Z}_N$,

$$\text{card}(T_n(X) \cap T_m T_n(X)) = \text{inj}(T_n X, T_n X)(T_m) = \text{inj}(X, X)(T_n^{-1} T_m T_n)$$
$$= \text{inj}(X, X)(T_m) = \text{card}(X \cap T_m(X))$$

thus $T_n(X)$ is a PT-set.

Theorem 1. *X is a PT-set for index m if and only if*

$$\text{iv}(X)(m) = \text{card}(X) - 1$$

Proof. We have

$$\text{card}(X \cap T_m(X)) = \text{inj}(X, X)(T_m) = \text{ifunc}(X, Y)(m) = \text{iv}(X)(m)$$

If X is a PT-set then $\text{card}(X) - 1 = \text{card}(X \cap T_m(X))$ and thus $\text{iv}(A)(m) = \text{card}(X) - 1$. Conversely, if $\text{iv}(A)(m) = \text{card}(X) - 1$ then $\text{card}(X \cap T_m(X)) = \text{iv}(A)(m) = \text{card}(X) - 1$ and thus X is a PT-set.

For instance, if $N = 12$, and $m = 7$. The major scale $X = \{0, 2, 4, 5, 7, 9, 11\}$ has $\text{iv}(X)(7) = 6 = \text{card}(X) - 1$.

Let us recall now some common definitions. Well-formed scales have been introduced by Norman Carey and David Clampitt in [5]. A *well-formed scale* is a scale where each generating interval spans a constant number of scale steps. The term *maximally even* (ME) was coined John Clough and Jack Douthett [6] to refer to scales that are subsets of a chromatic scale and in a well-defined sense are spread out as much as possible within that chromatic (see e.g. [13]). Jack Douthett's set is defined as follows. For $N, k, m \in \mathbb{Z}$, with $N > k$ and $k \neq 0$, the J-function on $\mathbb{Z}$ is

$$J_{N,k}^m(x) = \left\lfloor \frac{Nx + m}{k} \right\rfloor$$

where $\lfloor x \rfloor$ is the floor function (the greatest integer less than or equal to x) and the Douthett's J-set is

$$J^m_{N,k} = \{J^m_{N,k}(x), x = 0, 1, ..., k-1\}$$

We have a characterization of the subclass of $\mathrm{ME}(N,k)$ sets, with $N > k$ and $\gcd(N,k) = 1$ (see e.g. [6, Theorem 3.1, p. 148]). That is the ME sets that are not degenerated well-formed.

Theorem 2. *Let X a subset of $\mathbb{Z}_N$ with cardinality k prime to N. X is ME set that is also non-degenerate well-formed, if and only if X is a collection of successive images of some element $x \in \mathbb{Z}_N$ by the transposition T_n where n satifies, $nk = \pm 1 \mod N$.*

$$X = \{x, T_n(x), ..., T_n^{k-1}(x)\}$$

Remark 1. ME sets are invariant, under the action of the T/I group. Under this group, we only consider the case where n is the multiplicative inverse of k, because, the sets X and

$$\overline{X} = \{x, T_{-n}(x), ..., T_{-n}^{k-1}(x)\}$$

belong to the same class. Up to transpositions only (i.e. under the cyclic group), we have to consider both X and $\overline{X}$ which are not in the same class.

For each choice of index m, there exists only one class of set of progressive transposition. More precisely, we have the following result.

Theorem 3. *For each generator $1 < m < N-1$ such that $\gcd(m, N) = 1$, there exist a unique class set X (up to transposition) of progressive transposition given by*

$$X = \{x, T_m(x), ..., T_m^{k-1}(x)\}$$

for some element $x \in \mathbb{Z}_N$ and such that

(1) *X has k elements,*
(2) *has N different transpositions,*
(3) *and $T_m(X)$ have $k-1$ notes in common.*

Moreover, if $k = m$, X is well-formed and maximally even.

The set X is often written $G(N, k, m)$ (see for instance [6]). Remark that if $k \neq m$, the set X is not necessary well-formed (WF). For instance, $N = 12$, $m = 7$ and $k = 4$, $X = \{0, 2, 7, 9\}$ is not WF. X is not always maximally even (ME). For instance, for $N = 12$, $m = 7$ and $k = 3$, $X = \{0, 2, 7\}$ is not ME.

Example 1. Let $N = 12$. The integers of $\mathbb{Z}_N$ coprime with N are $\{1, 5, 7, 11\}$. Thus m takes two values: $m = 5$ or $m = 7 = -5$. If $m = 7$, for $x = 5$, the set $X = \{0, 2, 4, 5, 7, 9, 11\} = J^5_{12,7}$ corresponds to the white keys of the keyboard (the C major scale) and its complement $X^c = \{1, 3, 6, 8, 10\} = T_1(J^1_{12,5})$ corresponds to the black keys. If $m = 5$, the situation is dual. For $x = 9$, the set $Y = \{9, T_5(9), ...T_5^4(9)\} = \{0, 2, 5, 7, 9\} = T_{11}(X^c)$ corresponds to the black keys and its complement corresponds to the white keys.

Proof. Since m is coprime to N, the set $X = \{x, T_m(x), ..., T_m^{k-1}(x)\}$ has N different transpositions: X is not a set of limited transposition. Moreover, since the set

$$T_m(X) = \{T_m(x), ..., T_m^k(x)\}$$

has with X the intersection

$$X \cap T_m(X) = \{T_m(x), ..., T_m^{k-1}(x)\}$$

the cardinality $\#(X \cap T_m(X)) = k - 1$ shows that X is PT. If $k = m$, X is a well-formed scale and ME by Theorem 2.

1.1   Properties of PT Sets

In the following, we introduce the multiplicative inverse $r \in \mathbb{Z}_N^\times$ of m such that $rm = 1 \bmod N$), u the integer of $\mathbb{Z}_N$ such that

$$u = \max(-r, r) \tag{1}$$

and v the integer of $\mathbb{Z}_N$ given by

$$v = \min\{\max(r, -r), \max(r, 2r), \max(-r, -2r)\} \tag{2}$$

Theorem 4. *Let $1 < m < N - 1$ and $\gcd(m, N) = 1$, X a PT scale of cardinal k and generator m. X does not have two consecutive notes if and only if $k \leq \min(-r, r)$.*

Proof. Let jm be an element of the set X for $j = 0, 1, 2, ..., k - 1$. For r as above, the indexes $\{j, j + r\}$ lead to two consecutive notes$\{jm, jm + 1\}$ since $rm = 1 \bmod N$. The indexes $\{j, j - r\}$ also lead to two consecutive notes. Thus $k \leq \min(-r, r)$.

Corollary 1. *Let $1 < m < N - 1$ and $\gcd(m, N) = 1$, X a PT scale of cardinal k and generator m. The complement of X does not have two consecutive notes if and only if $k \geq u$.*

Theorem 5. *Let $1 < m < N - 1$ and $\gcd(m, N) = 1$, X a PT scale of cardinal k and generator m. X does not have three consecutive notes if and only if $k \leq v$.*

Proof. Let jm be an element of the set X for $j = 0, 1, 2, ..., k - 1$. For r as above, the indexes $\{j, j + r, j + 2r\}$ lead to three consecutive notes$\{jm, jm + 1, jm + 2\}$ since $rm = 1 \bmod N$. The indexes $\{j, j - r, j + r\}$ and $\{j, j - r, j - 2r\}$ also lead to three consecutive notes. In order to get all the notes in the same set we need to consider

$$v = \min\{\max(r, -r), \max(r, 2r), \max(-r, -2r)\}$$

That is if $r > N/2$, $v = \min(r, -2r)$ and if $r < N/2$, $v = \min(-r, 2r)$, thus we get

$$v = \begin{cases} 2r & \text{if } 1 < r \leq N/3 \\ N - r & \text{if } N/3 \leq r < N/2 \\ r & \text{if } N/2 < r \leq 2N/3 \\ 2N - 2r & \text{if } 2N/3 \leq r < N \end{cases}$$

1.2   The Case N Prime

If N is prime and $N > 2$. The generator m takes $\varphi(N) = N-1$ values, where φ is the Euler totient function. The integers coprime with N are $m = 1, 2, 3, 4, ..., N-1$. The first integer m greater that $N/2$ is $(N+1)/2$.

Proposition 3. *If N is prime and $N > 2$, the set $X = \{j(N+1)/2, j = 0, 1, ..., (N-1)/2\}$ has three consecutive notes.*

Proof. The three consecutive notes are $(m, 3m, 5m)$, with $m = (N+1)/2$. Since $3(N+1)/2 = (N+1)/2 + N + 1 = (N+1)/2 + 1 \mod N$, and $5(N+1)/2 = (N+1)/2 + N + 2 = (N+1)/2 + 2 \mod N$.

Proposition 4. *If N is prime and $N > 2$, the set $X = \{j(N+3)/2, j = 0, 1, ..., (N+1)/2\}$ does not have three consecutive notes.*

Proof. Let $m = (N+3)/2$. The set X can be decomposed as $X = A \cup B$, with

$$A = \{3j, j = 0, 1, 2, ..., (N+1)/4\}$$

and

$$B = \{3j + m, j = 0, 1, 2, ..., (N-3)/4\}$$

and thus X does not have three consecutive notes.

If N is prime, the set $X = \{j(N+3)/2, j = 0, 1, ..., (N+1)/2\}$ is the diatonic scale in the sense of the next section.

2   Generalized Diatonic Scales

Definition 2. *Let $1 < m < N - 1$ such that $2m \geq N + 1$ and $\gcd(m, N) = 1$, a set X is a diatonic scale of generator m if*

(1) *X has k elements, N different transpositions, X and $T_m(X)$ have $m-1$ notes in common,*
(2) *X does not have three consecutive notes $(x, x + 1, x + 2)$ for some $x \in \mathbb{Z}_N$.*
(3) *X^c does not have two consecutive notes $(x, x + 1)$ for some $x \in \mathbb{Z}_N$.*

This definition means that X is a TP scale by (1). (2) means that there is no three consecutive white keys, and (3) means that there is no two consecutive black keys. The condition $(2m - N) \geq 1$ excludes the dual cases.

Theorem 6. *Let $1 < m < N - 1$ such that $2m \geq N + 1$ and $\gcd(m, N) = 1$, a set X is a diatonic scale of k elements and generator m if and only if X is a TP scale with k elements and generator m*

$$X = \{x, T_m(x), ..., T_m^{k-1}(x)\}$$

for some element $x \in \mathbb{Z}_N$ and

$$u \leq k \leq v \tag{3}$$

The set X is denoted by Dia_N^m.

For a given N and a given m as above, there is 0 or 1 diatonic scale of generator m. For instance, for $N = 13$, the generator m belongs to the set $\{7, 8, 9, 10, 11\}$. For each value m, the triplet (r, u, v) given by Eqs. (1) and (2) leads to a set of values of k by Eq. (3). For $m = 7$, there is no diatonic scale since $u > v$ in the triplet $(r, u, v) = (2, 11, 4)$. For $m = 8$, $(r, u, v) = (5, 8, 8)$, there exists a diatonic scale of $k = 8$ elements,

$$\mathrm{Dia}_{13}^{8} = \{0, 1, 3, 4, 6, 8, 9, 11\}$$

and for $m = 11$, there exists also a diatonic scale of $k = 7$ elements.

$$\mathrm{Dia}_{13}^{11} = \{0, 1, 3, 5, 7, 9, 11\}$$

The other values of m lead to a triplet with $u > v$. From these scales, we can choose a unique diatonic class scale (up to inversion and transposition) by choosing the smallest m.

Definition 3. *The (optimal) diatonic scale* Dia_N *of* $\mathbb{Z}_N$ *is the diatonic scale* Dia_N^m *obtained for the smallest generator* m *such that* $2m \geq N + 1$.

Example 2. For $N = 13$, the smallest m is $m = 8$, $\mathrm{Dia}_{13} = \mathrm{Dia}_{13}^{8}$. The smallest $m = 7$ is excluded since $\{0, 7, 1, 8, 2, 9, 3\}$ is not a 7-diatonic scale (it has three consecutive white keys). For $N = 10$, there is no diatonic scale. The optimal diatonic scale always exists as long as $N \geq 12$.

Example 3. If $N = 18$, $m \in \{5, 7, 11, 13\}$. If $m = 11$, the ten first multiples 0, 11, 2, 13, 4, etc. form a set

$$X = \{0, 1, 4, 5, 8, 9, 11, 12, 15, 16\}$$

whose complement X^c has two consecutive notes. If $m = 13$, the eleven first multiples form a set

$$\mathrm{Dia}_{18} = \{0, 1, 3, 4, 6, 8, 9, 11, 13, 14, 16\}$$

which statifies all the properties of the diatonic scale.

Example 4. If $N = 20$, $m \in \{3, 7, 9, 11, 13, 17\}$. If $m = 11$, the eleven first multiples 0, 11, 2, 13, 4, etc. form a set

$$\mathrm{Dia}_{20} = \{0, 2, 4, 6, 8, 10, 11, 13, 15, 17, 19\}$$

which has no three consecutive notes and its complement

$$\mathrm{Dia}_{20}^{c} = \{1, 3, 5, 7, 9, 12, 14, 16, 18\}$$

has no two consecutive notes. If $m = 17$, the first 13 elements form a diatonic scale

$$X' = \{0, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19\}$$

which verifies the properties of the diatonic scale. But only Dia_{20}^{11} verifies the condition on m (the smallest integer such that $2m > N + 1$).

Proposition 5. *The complement of X does not contain two consecutive notes if and only if $X^c \subset T_1(X)$.*

Proof. If $(x, x+1) \in X^c$, then $T_{-1}(x, x+1) = (x-1, x) \notin X$ since x is in X. Thus $(x, x+1) \notin T_1(X)$. Conversely, since the transpositions are bijections, if $(x, x+1) \in X$ then $T_1(x, x+1) = (x+1, x+2) \notin X^c$.

The following table gives the values of r and v for different values of N and m, and the optimal diatonic scale.

N	m	r	v	Diatonic Scale Dia_N
12	7	7	7	$\{0, 2, 4, 6, 7, 9, 11\}$
13	8	5	8	$\{0, 1, 3, 4, 6, 8, 9, 11\}$
14	11	9	9	$\{0, 2, 4, 5, 7, 8, 10, 11, 13\}$
15	13	7	8	$\{0, 1, 3, 5, 7, 9, 11, 13\}$
16	9	9	9	$\{0, 2, 4, 6, 8, 9, 11, 13, 15\}$
17	12	10	10	$\{0, 2, 4, 6, 7, 9, 11, 12, 14, 16\}$
18	13	7	11	$\{0, 1, 3, 4, 6, 8, 9, 11, 13, 14, 16\}$
19	11	7	12	$\{0, 1, 3, 4, 6, 7, 9, 11, 12, 14, 15, 17\}$
20	11	11	11	$\{0, 2, 4, 6, 8, 10, 11, 13, 15, 17, 19\}$
21	13	13	13	$\{0, 2, 4, 5, 7, 9, 10, 12, 13, 15, 17, 18, 20\}$
22	17	13	13	$\{0, 2, 4, 6, 7, 9, 11, 12, 14, 16, 17, 19, 21\}$
23	16	13	13	$\{0, 2, 4, 6, 8, 9, 11, 13, 15, 16, 18, 20, 22\}$
24	13	13	13	$\{0, 2, 4, 6, 8, 10, 12, 13, 15, 17, 19, 21, 23\}$
25	14	9	16	$\{0, 1, 3, 4, 6, 7, 9, 10, 12, 14, 15, 17, 18, 20, 21, 23\}$
26	19	11	15	$\{0, 1, 3, 5, 6, 8, 10, 12, 13, 15, 17, 19, 20, 22, 24\}$
27	19	10	17	$\{0, 1, 3, 4, 6, 7, 9, 11, 12, 14, 15, 17, 19, 20, 22, 23, 25\}$
28	15	15	15	$\{0, 2, 4, 6, 8, 10, 12, 14, 15, 17, 19, 21, 23, 25, 27\}$
29	17	12	17	$\{0, 1, 3, 5, 6, 8, 10, 11, 13, 15, 17, 18, 20, 22, 23, 25, 27\}$
30	19	19	19	$\{0, 2, 4, 5, 7, 8, 10, 12, 13, 15, 16, 18, 19, 21, 23, 24, 26, 27, 29\}$
31	17	11	20	$\{0, 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 17, 18, 20, 21, 23, 24, 26, 27, 29\}$
32	17	17	17	$\{0, 2, 4, 6, 8, 10, 12, 14, 16, 17, 19, 21, 23, 25, 27, 29, 31\}$
33	26	14	19	$\{0, 1, 3, 5, 6, 8, 10, 12, 13, 15, 17, 19, 20, 22, 24, 26, 27, 29, 31\}$
34	21	13	21	$\{0, 1, 3, 4, 6, 8, 9, 11, 12, 14, 16, 17, 19, 21, 22, 24, 25, 27, 29, 30, 32\}$
35	24	19	19	$\{0, 2, 4, 6, 8, 10, 12, 13, 15, 17, 19, 21, 23, 24, 26, 28, 30, 32, 34\}$
36	19	19	19	$\{0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 19, 21, 23, 25, 27, 29, 31, 33, 35\}$

3   Generalized Major and Minor Scales

In the common equal temperament with 12 notes, the tritone plays an important role in the modulations (changing keys). As a set of limited transposition, and as

it belongs to the dominant seventh chord and to the diminished seventh chord, the tritone articulates the change from one key to another. Its unstable character sets it apart. The unicity of the tritone in major scale characterizes its key.

Proposition 6. *Let $\mathcal{D}$ be the set of divisors of N minus $\{1, N\}$. For $d \in \mathcal{D}$, and k such that $N = dk$, the set*

$$L_d = \{dj, j = 0, 1, ..., k\}$$

and its transposition $T_x(L_d)$ for all $x \in \mathbb{Z}_N$, are sets of limited transposition.

Example 5. For $N = 12$, $\mathcal{D} = \{2, 3, 4, 6\}$. The set $L_6 = \{0, 6\}$ and $T_x(L_6) = \{x, x + 6\}$ are sets of limited transposition. The sets $L_4 = \{0, 4, 8\}$ and $L_3 = \{0, 3, 6, 9\}$ are also sets of limited transposition. If N is prime, the set $\mathcal{D}$ is empty and the N-EDO does not have sets of limited transposition.

Proposition 7. *Let $N = dk$. The set L_d has exactly k transpositions.*

The following theorem is the engine for changing keys.

Theorem 7. *Let $\mathcal{D}$ be the set of divisors of N minus $\{1, N\}$ and X be a set with N transpositions. Suppose $L \subset X$, $L = T_x(L_k)$ for some x, $d \in \mathcal{D}$, and $N = dk$. Then L is included in exactly d sets of class X, namely*

$$\{X, T_k(X), ..., T_k^{d-1}(X)\}$$

Proof. Since $X = L \cup (X \backslash L)$ and L is a set of limited transposition

$$T_{kd}(X) = T_{kd}(L) \cup T_{kd}(X \backslash L) = L \cup T_{kd}(X \backslash L)$$

thus $L \subset T_{kd}(X)$.

Definition 4. *Let $\mathcal{D}$ be the set of divisors of N minus $\{1, N\}$ and d the greatest element of $\mathcal{D}$, and k such that $N = dq$, the* generalized tritone *is the class set (up to transpositions)*

$$L_d = \{dj, j = 0, 1, ..., q - 1\}$$

Proposition 8. *Let $1 < m < N - 1$ and $\gcd(m, N) = 1$, X a PT scale of cardinal k and generator m. X contains a generalized tritone if and only if $k \geq d(q-1)+1$, where d is the greatest element of $\mathcal{D}$ and $N = dq$.*

Proposition 9. *Let X be a diatonic scale with generator m in the N universe. X contains at least one generalized tritone.*

Definition 5. *Let $1 < m < N - 1$ such that $2m \geq N + 1$ and $\gcd(m, N) = 1$, X is generalized major (or micro-major) scale if*

(1) *X is a set of progressive transposition,*
(2) *X does not have three consecutive notes,*

(3) *X contains a unique generalized tritone.*

Theorem 8. *Let m be such that $\gcd(m, N) = 1$, r the solution of $rm = 1$ mod N and v as above, X a PT set*

$$X = \{x, T_m(x), ..., T_m^{k-1}(x)\}$$

of cardinality k. X is a generalized major scale if and only if

$$d(q - 1) + 1 \leq k \leq v \tag{4}$$

A generalized major scale does not necessarily exist. If N is prime, there is no generalized tritone and thus no major scale, although there exists a diatonic scale.

Example 6. If $N = 12$, $m \in \{5, 7\}$. The tritone is the class set $\{0, 6\}$. Since the parameters are $r = v = 7$, X does not have three consecutive notes if and only if $k \leq 7$ and X contains a transposition of the tritone $\{0, 6\}$ if and only if $k > 6$. Therefore the PT set $X = \{0, 2, 4, 5, 7, 9, 11\}$ contains a unique tritone $\{5, 11\}$. In this case, the major scale is the same as the diatonic scale.

Example 7. If $N = 14$, $m \in \{9, 11\}$. The generalized tritone are transpositions of $\{0, 7\}$. If $m = 9$, $r = 11, d = 7$ and $v = \min(r, -2r) = 6$. X does not contains three consecutive notes if and only if $k \leq 6$ and X contains a transposition of the tritone $\{0, 7\}$ if and only if $k > 7$. Therefore there is no major scale associated with $m = 9$. The scale

$$X = \{0, 3, 4, 7, 8, 9, 12, 13\}$$

with $m = 9$, $k = 8$, is well-formed, PT and has a unique tritone, but has three consecutive notes.

If $m = 11$, $r = 9$ and $v = \min(r, -2r) = 9$. k needs to be less or equal 9 and greater than 7. So the candidates are $k = 8$ or 9. The set formed by the nine first multiples of m which is the diatonic scale

$$3 \quad 0 \quad 11 \quad 8 \quad 5 \quad 2 \quad 13 \quad 10 \quad 7$$

contains two generalized tritone and can not be a generalized major scale. Thus

$$\text{Dia}_{14} = \{0, 2, 4, 5, 7, 8, 10, 11, 13\}$$

has 9 elements. For $k = 8$, the first eight multiples define the major scale with a unique triton

$$\text{Maj}_{14} = \{0, 2, 5, 7, 8, 10, 11, 13\}$$

Although the scale Maj_{14} is a PT set, a subset of the diatonic scale, and has a unique triton, Maj_{14} is neither maximally even nor well-formed. But it has a unique triton and no three consectutive notes, which is an important criterion to keep the major character and avoid the feeling of chromaticism. This example shows that the choice arises between well-formed set and TP set. Is it more important for a major scale to be well formed scale or to be a PT set?

Example 8. If $N = 15$, $m \in \{2, 4, 7, 8, 11, 13\}$. The divisors of 15 are $\{1, 3, 5\}$ so $d = 5$ and $q = 3$. The generalized tritone is $\{0, 5, 10\}$. There exists a diatonic scale for $m = 13$

$$\text{Dia}_{15} = \{0, 1, 3, 5, 7, 9, 11, 13\}$$

but no major scale. A major scale needs to take 11 multiples of m, thus has cardinality 11, but in this case X has three consecutive notes, which contredict the definition.

Example 9. If $N = 21$, there exists a diatonic scale for $m = 13$ (or $m = 8$)

$$\text{Dia}_{21} = \{0, 2, 4, 5, 7, 9, 10, 12, 13, 15, 17, 18, 20\}$$

but no major scale for same reasons as for $N = 15$.

3.1   Case N Even

If N is even, we have a graphical representation of these scales. The set $\{N - m, (N/2 - 1)m\}$ is a generalized tritone. The major scale with tonic 0 in the N universe is

$$\text{Maj}_N = \{0, m, 2m, 3m, 4m, ..., (N/2 - 1)m, N - m\}$$

In this case, the diatonic scale is equal to the major scale. Its graphic representation is in the frame $(N - m, N/2)$ which has the multiples of $N - m$ along the x-axis and the multiples of $N/2$ along the y-axis.

$$0 \text{---} \quad N - m$$
$$|$$
$$(N/2 - 1)m \text{---} \cdots \text{---} 4m \text{---} 3m \text{---} 2m \text{---} m$$

A linear representation can be obtain if the two first notes $\{0, N - m\}$ are added after m on the right of the graph. The two ends of the graph are then separated by the unique generalized tritone of the major scale. The relative minor scale has one more generalized tritone $\{2m, 2m + N/2\}$

$$\text{Min}_N = \{0, 2m + N/2, 3m, 4m, 5m, ..., (N/2 - 1)m, N - m\}$$

The note m is replaced by $2m + N/2$. The two scales are in the same neighborhood (differ by only one note).

$$0 \text{---} \quad N - m \qquad\qquad\qquad 2m + N/2$$
$$| \qquad\qquad\qquad\qquad\qquad\qquad |$$
$$(N/2 - 1)m \text{---} \cdots \text{---} 4m \text{---} 3m \text{---} \quad 2m$$

Example 10. If $N = 12$, $m = 7$. Maj_{12} is the C-major scale, and Min_{12} is the A-minor scale (harmonic minor scale). If $N = 14$, we find the previous results. The relative minor of the major scale based on $x = 0$ is the scale obtains from the major scale replacing 11 by 1. $\text{Min}_{14} = \{0, 1, 2, 3, 5, 8, 10, 13\}$.

References

1. Agmon, E.: A mathematical model of the diatonic system. Journal of Music Theory **33**, 1–25 (1989)
2. Agmon, E.: The Languages of Western Tonality. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39587-1
3. Balzano, G.J.: The group-theoretic description of 12-fold and microtonal pitch systems. Comput. Music. J. **4**, 66–84 (1980)
4. Broué, M.: Les tonalités musicales vues par un mathématicien. In: Le temps des savoirs, Le code, Dominique Rousseau, Michel Morvan (dir.), vol. 4, pp. 41–82. Odile Jacob, Revue de l'Institut universitaire de France, Paris (2001)
5. Carey, N., Clampitt, D.: Aspects of well-formed scales. Music Theory Spectr. **11**, 187–206 (1989)
6. Clough, J., Douthett, J.: Maximally even sets. J. Music Theory **35**, 93–173 (1991)
7. Clough, J.: Diatonic interval cycles and hierarchical structure. Perspect. New Music **32**(1), 228–53 (1994)
8. Clough, J., Engebretsen, N., Kochavi, J.: Scales, sets, and interval cycles: a taxonomy. Music Theory Spectr. **21**(1), 74–104 (1999)
9. Douthett, J., Kranz, R.: Construction and interpretation of equal-tempered scales using frequency ratios, maximally even sets, and P-cycles. J. Acoust. Soc. Am. **107**, 2725–34 (2000)
10. Douthett, J., Hyde, M.M., Smith, C.J.: Music Theory and Mathematics, Chords, Collections and Transformations. University of Rochester Press, Rochester (2008)
11. Jedrzejewski, F.: Generalized diatonic scales. J. Math. Music **2**(1), 21–36 (2008)
12. Jedrzejewski, F.: Hétérotopies musicales. Modèles mathématiques de la musique. Hermann, Paris (2019)
13. Johnson, T.A.: Foundations of Diatonic Theory: A Mathematically Based Approach to Music Fundamentals. Scarecrow Press, Lanham (2008)
14. Lewin, D.: A formal theory of generalized tonal functions. J. Music Theory **26**, 23–60 (1982)
15. Lewin, D.: Generalized Musical Intervals and Transformations. Yale University Press, New Haven (1987)
16. Louvier, A.: Recherche et classification des modes dans les tempéraments égaux. Musurgia **4**(3), 119–131 (1997)

Parsimonious Graphs for Selected Heptatonic and Pentatonic Scales

Luis Nuño(✉)

Communications Department, Polytechnic University of Valencia,
Camino de Vera S/N, 46022 Valencia, Spain
lnuno@dcom.upv.es, harmonicwheel@gmail.com

Abstract. Tonal music is based on major, melodic and harmonic minor scales. In some cases, the harmonic major scale is also used. In this paper, four additional heptatonic scale types, derived from them, are considered. The harmonic characteristics of these eight scale types are analyzed by the trichord- and tetrachord-type vectors, which list, respectively, the number of times each trichord and tetrachord type is contained in a set type. Then, a novel parsimonious graph is provided, called 7-Cyclops, which relate those scales by single-semitonal transformations. On the other hand, their complements are eight pentatonic scales, whose harmonic characteristics are also analyzed and the corresponding parsimonious graph, called 5-Cyclops, is given. These graphs highlight the cycles of fifths and fourths, which are the only possible circumferences linking the same scale types in these graphs. Other parsimonious transformations, like moving one note by a whole tone, are easily found in these graphs, too. The acoustical relationship between those heptatonic and pentatonic scale types is analyzed by the pentachord-type vector, which lists the number of times each pentachord type is contained in a set type. With the inclusion of a musical example, all this information is intended both for theorists and composers.

Keywords: Parsimonious transformation · Heptatonic scale · Pentatonic scale · Cyclops · Trichord-type vector · Tetrachord-type vector · Pentachord-type vector · Cycle of fifths · Cycle of fourths

1 Introduction

The major scale is the basis of Western music. Although perfectly well known, it is worth reviewing now some of its main characteristics. It consists of seven notes showing great acoustical affinity among them, to the extent that they constitute a "complete and versatile" set. Thus, most popular songs – and not so popular – are composed in a major key. Our musical notation, based on the staff and the key signatures, is ideal for writing music in major keys. The piano, a crucial musical instrument, is especially suitable for playing in the C major key. The names of the notes are seven – instead of twelve –, precisely those of the C major scale. The term octave (a Latin word for "eighth") indicates its extension (including an ending tonic), whereas the terms whole tone and semitone describe the

types of intervals between two consecutive notes in it. As well, the quality of the intervals (perfect, major, minor, …) are defined with respect to a major scale. In summary, the major scale is a cornerstone of music theory and composition.

Other kinds of scales can be directly derived from the major one. Thus, by choosing any of its notes as the tonic, we obtain seven *modes*, from which the Aeolian one constitutes the minor scale, which is also prevalent in Western music. In this case, its sixth and seventh degrees can be natural or altered (raised by a semitone), thus giving rise to the natural, melodic, and harmonic minor scales.

On the other hand, the complement of a major scale is a major pentatonic one. In general, pentatonic scales – not only the major pentatonic – have been used since ancient times by many different cultures. Although they predominate in Eastern countries (China, Japan, India, Java, etc.), they are also used in several Western styles, such as Classical, Scottish, Andean, Jazz, etc.

This study is based on the standard twelve-tone chromatic system ($\mathbb{Z}_{12}$) and uses the nomenclature of *Forte names* and *set classes* [1], which here will be called *scale classes*. Additionally, the *non-inversionally-symmetrical* ones are split into two *scale types* related by *inversion*, named "a" and "b", following [2]. Under these premises, the major and major pentatonic scales are the most even set types with seven and five notes, respectively, both possessing an exclusive property: apart from the set types with one or eleven notes, they are the only set types that can be transformed into the same set types by a *single-semitonal transformation*.[1] For example, by raising in the C major scale the note F by a semitone, we obtain the G major scale. This property, in the case of major scales, gives rise to the order of sharps and flats, as well as the cycle of fifths [3, 4], which is essential in the theory of *modulation*, that is, the change from one key to another. In this respect, given a key, its "nearest" keys are those having one sharp or flat more or less in their key signatures [5, 6]. This also applies to the minor keys, since they have the same key signatures as their relative major ones.

A *parsimonious transformation* is a more general concept, where one or more notes move by a semitone or a whole tone (in practice, normally no more than two semitones in total), while sustaining the rest of them [7]. Thus, the *Tonnetz* is a first representation of them for major and minor triads, while [7] provides more complex and interesting graphs. In the nineteenth century, several composers made extensive use of this kind of transformations and a large number of their works are analyzed in [8]. A different approach is given in [9], where pitch-class sets are represented in special spaces called *orbifolds*. In [10], the "most common" trichords and tetrachords are represented in cyclic circular graphs called *Cyclopes*, which allow us to analyze a great number of such musical works in a practical way. In this paper, two novel parsimonious graphs of this kind are developed for heptatonic and pentatonic scales. In each case, eight scale types are chosen following specific criteria. As well, an aforementioned result is shown graphically: the cycle of fifths for the major scales, together with the cycle of fourths for the major pentatonic ones, are the only possible circumferences connecting pitch-class sets of the same type in this kind of graphs [3, 4].

---

[1] A transformation of a pitch-class set (in our case, a scale) consisting in raising or lowering one pitch-class (note) by a semitone, while sustaining the rest of them.

2 Selection of Heptatonic Scales

Tonal music is based on major, melodic and harmonic minor scales [5, 6], whose "extended" Forte names are 7-35, 7-34, and 7-32a, respectively. To complete the numerical series, we can also consider 7-32b (harmonic major) and 7-33 (which we will call "Neapolitan major", following [11]). These are the five most even heptatonic scale types, as they have the least *interval-class vectors*[2] [1], with respect to the lexicographic order. The most even one is, obviously, the major scale (7-35).

In order to obtain other heptatonic scales related to them, we can simply combine two groups of four consecutive notes or *tetrachords*.[3] An example of this procedure is provided by the traditional Indian music, where up to 72 heptatonic scales, called "Melakarta ragas", are obtained by combining different types of tetrachords [12]. However, since the total number of heptatonic scale types is 66, some of those ragas are, in fact, modes of other ragas, the number of different scale types being 36. In any case, both 66 and 36 are too many scale types to develop practical and visually simple graphs relating them.

Another option is to start with a major scale and raise or lower one or more notes by a semitone, as done with the natural minor scale to obtain the melodic and harmonic minor ones. In this case, it seems appropriate to choose the altered notes from the nearest key signatures.

These two procedures are now used to obtain a "reasonable" number of heptatonic scale types, which can be of interest both for theorists and composers.

2.1 Combinations of Two Tetrachords

Let us consider the C major scale. It is composed of the tetrachords C – D – E – F and G – A – B – C, whose "intervallic structures", in semitones, are the same: 221; and the interval between the two tetrachords is 2 semitones. So, we can write the full intervallic structure of C major as 221 2 221; and the tetrachord 221 can be called "major". As well, starting from notes D or A in the C major scale, the first four notes give the tetrachord 212, which we will call "minor". Similarly, starting from E or B, we obtain the tetrachord 122, which we will call "Phrygian". And starting from F, we obtain the tetrachord 222, which we will call "Lydian".

Thus, we can obtain different heptatonic scales by combining any two of those tetrachords. But, to obtain the harmonic minor or major scales, we need another tetrachord: the 131, which we will call "harmonic". Table 1 shows the 25 combinations of these 5 tetrachords, with the resulting intervallic structures and the names and symbols given here to the corresponding scale types. The less common names are taken from [11] and all modes of a scale type are given the same name. Note that, in all cases, the interval between the two tetrachords is such that the starting and ending notes in the scale be the same; or, in other words, the sum of the semitones in every intervallic structure be 12.

---

[2] The vector listing the number of times each of the 6 dyads (intervals from 1 to 6 semitones) is contained in a given set type or set class (in our case, scale type or scale class). It characterizes, to a great extent, the sonority of a set class. In [1], it was called *interval vector*.

[3] The term tetrachord also means "4-note chord". However, throughout this paper, its right meaning will easily be determined by the context.

Table 1. Heptatonic scale types obtained by combining two out of five tetrachords, with their intervallic structures. Scale symbols: M: major, mm: melodic minor, hm: harmonic minor, hM: harmonic major, NpM: Neapolitan major, Npm: Neapolitan minor, hL: harmonic Lydian, hh: double harmonic or Hungarian, WT: whole tone.

1st ↓	2nd →				
	Major	Minor	Phrygian	Lydian	Harmonic
Major	221 2 221 M	221 2 212 M	221 2 122 mm	221 1 222 NpM	221 2 131 hM
Minor	212 2 221 mm	212 2 212 M	212 2 122 M	212 1 222 mm	212 2 131 hm
Phrygian	122 2 221 NpM	122 2 212 mm	122 2 122 M	122 1 222 M	122 2 131 Npm
Lydian	222 1 221 M	222 1 212 mm	222 1 122 NpM	222 222 WT	222 1 131 hL
Harmonic	131 2 221 hL	131 2 212 hM	131 2 122 hm	131 1 222 Npm	131 2 131 hh

The combination of two Lydian tetrachords gives rise to the whole-tone scale (WT), which only has six notes, thus being excluded from this study. The rest of the combinations give rise to eight different scale types, which include the five most even and is a suitable number for developing our graphs. Table 2 shows those scale types with their extended Forte names, the symbols here used to represent them, their *intervallic forms*[4] [2] starting from the tonic, and their interval-class vectors.

Table 2. Heptatonic scale types considered here. The intervallic forms start from the tonic.

Heptatonic scale	Symbol	Intervallic form	Interval-class vector
7-22	hh	1312131	424542
7-30a	Npm	1222131	343542
7-30b	hL	2221131	343542
7-32a	hm	2122131	335442
7-32b	hM	2212131	335442
7-33	NpM	1222221	262623
7-34	mm	2122221	254442
7-35	M	2212221	254361

2.2 Combinations of the Altered Notes from the Nearest Key Signatures

Let us consider again the C major scale. Its two nearest key signatures, both in the order of the sharps and the flats, introduce the altered notes F♯, C♯, B♭, and E♭. Thus,

---

[4] The intervallic form is the sequence of intervals, in semitones, between every two adjacent pitch classes in a set type (in our case, a scale type), including the interval between the last and the first ones, or any of its circular shifts. If it starts from a scale tonic, then it matches the "intervallic structure" previously used in this section.

using the C major scale with any of those notes either natural or altered, gives rise to 16 combinations, which are shown in Table 3, together with the resulting scales. As can be seen, we obtain the same eight scale types as with the previous procedure. In fact, according to [13], they are the ones with a "span" ≤10. Therefore, they will be the heptatonic scale types considered in this study, which are those listed in Table 2.

Table 3. Heptatonic scales obtained from CM by combining the two nearest altered notes.

Altered Notes	-	F → F#	C → C#	F → F# , C → C#
-	CM	GM	Dmm	DM
B → Bb	FM	Gmm	Dhm	DhM
E → Eb	Cmm	GhM	DNpM	GhL
B → Bb , E → Eb	BbM	Ghm	DNpm	Dhh

2.3 Harmonic Characteristics of the Selected Heptatonic Scales

To evaluate the harmonic characteristics of the heptatonic scale types considered here, we can use two generalizations of the interval-class vector: the *trichord-* and *tetrachord-type vectors*, which list, respectively, the number of times each trichord and tetrachord type is contained in a set type. Table 4 shows these vectors for those scale types, where each digit corresponds to a chord type in the order established in [2]. Thus, for example, the first digits from right to left in the trichord-type vector correspond to the augmented, major, minor, and diminished triads. And the first digits from right to left in the tetrachord-type vector correspond to the diminished, dominant, half-diminished, and minor seventh chords. Digits in bold correspond to the trichord and tetrachord types considered in [10].

Table 4. Trichord- and Tetrachord-Type Vectors of the Heptatonic scale types considered here.

Scale	Trichord-Type Vector	Tetrachord-Type Vector
7-22	111333322-011**2212331**	000111110220-000110011111111222**222-0000011110**
7-30a	111212321-312**3321321**	010001110110-011010011112012101**202-2120211110**
7-30b	111123212-321**3321231**	001010110110-011100011110221110**022-2210211110**
7-32a	022222221-122**1224321**	000100000110-111212112011011112**211-0011101211**
7-32b	022222212-122**2124231**	000100000110-111121221100111121**121-0101101121**
7-33	111111111-611**6611112**	011000110000-011110000111111000**110-6110630110**
7-34	022111111-333**3332221**	000100000000-211111111111111000**110-2222211220**
7-35	022002211-344**1151330**	000000000010-222001122110011000**002-1334003110**

3   Parsimonious Graphs for the Selected Heptatonic Scales

Figure 1 is a diagram showing the eight heptatonic scale types considered here, linked by single-semitonal transformations, and where link crossings are avoided. The arrows show how to transform the scale types by raising one note by a semitone (or, in the opposite direction, by lowering one note by a semitone). Note that the major scale is the only scale type that can be self-transformed, which will give rise to the cycle of fifths. When a scale class consists of two scale types related by inversion, they are placed next to each other (7-30a next to 7-30b and 7-32a next to 7-32b). This allows us to clearly see the relations between them, when they exist, as is the case for 7-32a and 7-32b. As well, the links connecting two such scale types with others always consist of pairs of arrows in opposite directions, one for each scale type.

   Arabic numerals indicate the initial and final notes referring to the scale tonics, where 1 to 6 stand for perfect or major intervals, which may be altered with ♯ or ♭, and major and minor sevenths are denoted by Δ and 7, respectively. And the Roman numerals at the middle of the arrows indicate the interval between the scale tonics, *in semitones* (letter "O" means zero). For example, Cmm consists of notes (C, D, E♭, F, G, A, B) and, by raising the perfect fourth (4) by a semitone, the new note is the major seventh (Δ) of the new scale, a "hM" with tonic C – V semitones, that is, GhM = (G, A, B, C, D, E♭, F♯). Or, by lowering in Cmm the major seventh (Δ) by a semitone, it turns into the tonic (1) of the "M" scale with tonic C – II, that is, B♭M = (B♭, C, D, E♭, F, G, A). Other parsimonious transformations can be found in this diagram, particularly those obtained by moving one note by a whole tone. But this will be explained in Sect. 4.2, together with the transformations of pentatonic scale types.
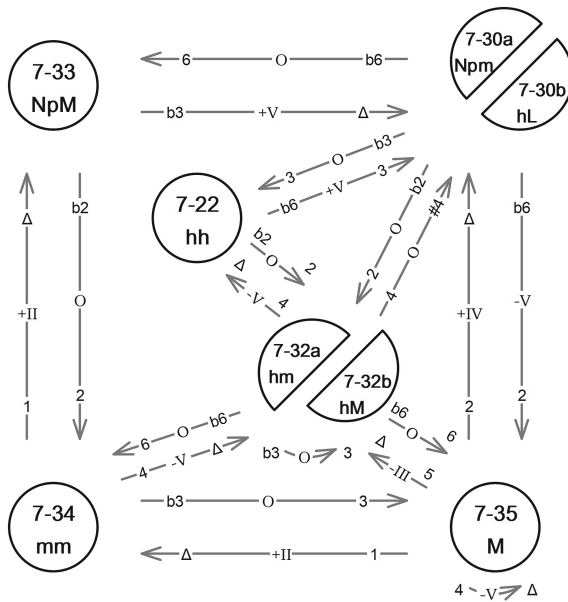
Fig. 1. The heptatonic scale types included in Table 2 with their single-semitonal transformations.

This diagram does not include the scale tonics, so it represents the "local relation-ships" in a more general scale space. Thus, let us now represent the "global relationships" among all the heptatonic scales considered here (with their tonics). To do this, we group them into *voice-leading zones* [8] or, simply, *zones* $\varphi \in [0, \ldots, 11]$, which are the equiv-alence classes defined by the sum of the notes in a scale, modulo 12. For example, CM = (C, D, E, F, G, A, B) is in the zone $\varphi = 0 + 2 + 4 + 5 + 7 + 9 + 11 = 2$ (mod 12). This way, given a scale in the zone φ, the one obtained from it by raising one note by a semitone will be in $\varphi + 1$. And scales related by *pure contrary motion*, as CM and Dhm = (D, E, F, G, A, B♭, C♯) will be in the same zone (in this case, $\varphi = 2$). The final result is given in Fig. 2 in a cyclic circular graph, here called *7-Cyclops*, where φ is actually an angular position starting from "12 o'clock" ($\varphi = 0$ for B♭M) and increasing clockwise. The arrows in Fig. 1 are now substituted by lines whose directions are assumed to be clockwise and no Roman numerals are used, since the tonics are directly given. Because 7 and 12 are coprime integers, in each zone of the 7-Cyclops there is exactly one scale of each type. The links between major scales make up the *cycle of fifths*, which corresponds to the only possible circumference in this graph (the bold line).

A different circular diagram is given in [9, p. 136], which includes the major (there called *diatonic*), melodic minor (there called *acoustic* and whose tonic is the perfect fourth of the corresponding melodic minor scale), harmonic minor, and harmonic major scales, plus three non-heptatonic scales with *transpositional symmetry*: the whole-tone (6-35), *hexatonic* (6-20, also called *augmented*), and *octatonic* (8-28, also called *half-step/whole-step diminished*), whose intervallic forms starting from the tonic are, respec-tively, 222222, 131313, and 12121212. To properly allocate all these scales, 36 angular positions are used, which obviously cannot coincide with the zones considered here, and the notes that change from one scale to the other are not shown.

Another relevant work is [14]. Based on just intonation and "commatic transition series", three groups of heptatonic scales are obtained:

– Hiatal: 7-35, 7-32a, 7-32b, 7-30a, 7-30b, 7-22
– Octatonic: 7-35, 7-34, 7-32a, 7-32b, 7-31a, 7-31b
– Whole-tone: 7-35, 7-34, 7-33, pseudo-whole-tone

where "pseudo-whole-tone" is a whole-tone scale plus one enharmonic note. So, in [14], this scale and the pair 7-31a/7-31b are added to those in Table 2. As well, the nomenclature used there for some scale names differs from the one used in this paper.

For each group, the corresponding diagrams for both the local and global relation-ships are given, although the notes that change from one scale to the other are not indicated. Finally, the diagrams of the three groups are superimposed, both those with the local and the global relationships, the latter resulting in a really complex diagram, so that only the links are shown, but not the scale names. As well, the links between 7-33 and the pair 7-30a/7-30b are not included, since they belong to different groups.

4 Pentatonic Scales

A similar process is now followed for the pentatonic scales.

Fig. 2. The 7-Cyclops, with the heptatonic scales considered in Table 2.

4.1 Selection of Pentatonic Scales

For consistency with previously selected heptatonic scales, we will select the pentatonic scales that are their complements. Table 5 shows these scale types with their extended Forte names, the symbols here used to represent them, their intervallic forms starting from the tonic, and their interval-class vectors. The scales whose symbols are of the form "m…P" and "7…P" are derived, respectively, from the minor (mP) and "dominant" (7P) pentatonic scales, whose intervallic forms, starting from the tonic, are 32232 and 22332. Note that "mP" is a mode of the major pentatonic scale (MP).

We can also obtain these scale types in a similar way as we did for the heptatonic scales by using the nearest altered notes. However, because now we cannot use the concept of key signature as before, we will talk of the nearest *modified* notes. For example, starting from the CMP scale, we search for the notes that, by moving by a semitone, generate other major pentatonic scales. There are just two possibilities: raising E to F to obtain FMP or lowering C to B to obtain GMP. Thus, considering the

Table 5. Pentatonic scale types considered here. The intervallic forms start from the tonic.

Pentatonic Scale	Symbol	Intervallic Form	Interval-Class Vector
5-22	mΔ#4P	33141	202321
5-30a	mΔP	32241	121321
5-30b	Δ#5P	42231	121321
5-32a	m#4P	33132	113221
5-32b	7#9P	31332	113221
5-33	7#5P	22422	040402
5-34	7P	22332	032221
5-35	MP	22323	032140

two nearest modified notes, both raising and lowering, we obtain the Table 6, which is analogous to Table 3 but for the pentatonic scales. As can be seen, again eight different scale types are obtained, which are precisely those listed in Table 5.

Table 6. Pentatonic scales obtained from CMP by combining the two nearest modified notes.

Modified Notes	-	E → F	A → Bb	E → F, A → Bb
-	CMP	FMP	C7P	BbMP
C → B	GMP	G7P	Em#4P	G7#9P
G → F#	D7P	D7#9P	D7#5P	F#Δ#5P
C → B, G → F#	DMP	Bm#4P	BmΔP	BmΔ#4P

The harmonic characteristics of these scale types are shown in Table 7, which includes their trichord- and tetrachord-type vectors. This is analogous to Table 4 but for the pentatonic scales, the same conventions being used here.

Table 7. Trichord- and tetrachord-type vectors of the pentatonic scale types considered here.

Scale	Trichord-Type Vector	Tetrachord-Type Vector
5-22	000111111-000**001111**	000000000010-0000000000000000011**110-0000000000**
5-30a	000100110-101**110101**	000000000000-00000000001001000**100-0010100000**
5-30b	000011001-110**1110011**	000000000000-0000000000110000**010-0100100000**
5-32a	000110010-011**0101210**	000000000000-000000000001000101**000-0000001100**
5-32b	000110001-011**1001120**	000000000000-0000000000001001100**000-0000001010**
5-33	000000000-300**3300001**	000000000000-00000000000000000**000-2000210000**
5-34	000000000-211**1111110**	000000000000-00000000000000000**000-1110000110**
5-35	000000000-122**0030110**	000000000000-00000000000000000**000-0112001000**

4.2 Parsimonious Graphs for the Selected Pentatonic Scales

Figure 3 is a diagram showing the eight scale types considered here, linked by single-semitonal transformations. Since if two scale types are related by a single-semitonal transformation, then so are their complements, this figure is completely analogous to Fig. 1, the same conventions being used here. Note that the major pentatonic scale is the only scale type that can be self-transformed, which, in this case, will give rise to the cycle of fourths.

   This graph also allows us to easily find other parsimonious transformations, particularly those obtained by moving one note by a whole tone. They correspond to two consecutive arrows, where the ending note on the first matches the starting note on the second one. For example, if in C7♯9P = (C, D♯, E, G, B♭) we raise the major third (3) by a semitone, it turns into the major second (2) of an "MP" scale; and by raising again this note by a semitone, it turns into the augmented fourth (♯4) of an "m♯4P" scale whose tonic is C + III – III, that is, Cm♯4P = (C, E♭, F♯, G, B♭). There are 5 parsimonious transformations of this kind, the rest of them being: from "7♯5P" to the same scale type through "7P", from "7P" to "m∆P" and "∆♯5P" through "MP", and from "∆♯5P" to "m∆P" through "MP". The latter is harder to see because there is, or there may be, a voice crossing. For example, transforming C∆♯5P = (C, E, F♯, G♯, B) into C♯m∆P = (C♯, E, F♯, G♯, B♯) or (C♯, E, F♯, G♯, C) can be done by raising B by a whole tone, which crosses C. But, in Fig. 3, this must be done by first raising the tonic (C) by a semitone, giving EMP, and then raising its perfect fifth (B) by a semitone, thus avoiding the voice crossing. There is, however, a simpler way to find this parsimonious transformation. To this end, we have to use Fig. 1 and take into account that: 1) the complements of 5-30a, 5-30b, and 5-35 are, respectively, 7-30b, 7-30a, and 7-35; 2) raising/lowering a note by a semitone in a scale type corresponds to lowering/raising a note by a semitone in its complement; and 3) if a note of a scale type, by moving by a whole tone, crosses a voice, it does not produce any voice crossing in its complement. Therefore, the pentatonic transformation from "∆♯5P" to "m∆P" through "MP", by raising one note by a whole tone, corresponds to the heptatonic transformation from "Npm" to "hL" through "M", by lowering one note by a whole tone, which is clearly seen in Fig. 1 (note ∆ lowers to 2, which lowers to ♭6). Consequently, using the two diagrams, we easily find all transformations of this kind.

   From the local relationships among the pentatonic scales (without the tonics), we can obtain the global ones (with all the tonics). They are shown in Fig. 4 in a cyclic circular graph, here called *5-Cyclops*. This is analogous to Fig. 2 but for pentatonic scales, the same conventions being used here. As well, in each zone of the 5-Cyclops there is exactly one scale of each type. But now, the links between major pentatonic scales make up the *cycle of fourths*, which also corresponds to the only possible circumference in this graph (the bold line).

4.3 Relation Between the Selected Heptatonic and Pentatonic Scales

Apart from the complementary relationship between heptatonic and pentatonic scales, a more acoustical relationship can be found by using a further generalization of the interval-class vector: the pentachord-type vector, which lists the number of times each
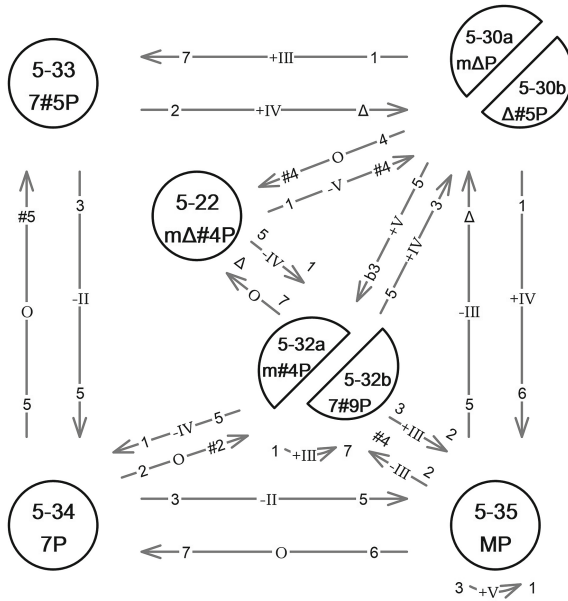
Fig. 3. The pentatonic scale types included in Table 5 with their single-semitonal transformations.

pentachord type (or pentatonic scale type) is contained in a given scale type. Thus, Table 8 shows the pentachord-type vectors of the heptatonic scale types included in Table 2 plus the pair 7-31a/7-31b, considered in [14]. Now, each digit corresponds to a pentatonic scale type in the order established in [2] and those in bold correspond to the ones considered here. Thus, from left to right, the first digit in bold corresponds to 5–22 and the last nine digits (those after the space) correspond to 5-30a, 5-30b, 5-31a, 5-31b, 5-32a, 5-32b, 5-33, 5-34, and 5-35.

As can be seen, 7-35 contains three 5-35 scale types. This is the maximum number of 5-35 contained in any heptatonic scale type (the rest of the heptatonic scale types – not only those in the table – contain no more than two). As well, 7-34, 7-33, and 7-22 contain, respectively, the maximum number of 5-34, 5-33, and 5-22, which are 2, 6 and 2 (in all other cases, they contain no more than one of each of them). Regarding the pairs of scale types forming a scale class, we must take into account that the complement of an a-type is a b-type and vice versa [2]. Then, 7-30a and 7-30b contain, respectively, the maximum number of 5-30a and 5-30b (that is, the inversions of their complements), which is two (in all other cases, they contain no more than one of each of them); and they do not contain their corresponding complements. As well, 7-31a and 7-31b contain, respectively, the maximum number of 5-31a and 5-31b, which is three (in all other cases, they contain no more than two) and do not contain their corresponding complements. And, with respect to 5-32a and 5-32b, it turns out that the heptatonic scale types containing the maximum number of them are, respectively, 7-31a and 7-31b, which is two (in all other cases, they contain no more than one). At least, 7-32a and 7-32b contain, respectively, one 5-32a and one 5-32b (again the inversions of their complements) and do not contain their corresponding complements. Therefore, in all these cases there is, to a greater or

Fig. 4. The 5-Cyclops, with the pentatonic scales considered in Table 5.

Table 8. Pentachord-Type Vector of several Heptatonic scales.

Scale	Pentachord-Type Vector
7-22	0000000001100-00000000000000011111111100111**2**-00000000001100 **000**0**11000**
7-30a	0000000000100-01000100001000100000101001110**0**-00110001110100 **200**0**10110**
7-30b	0000000001000-00100010001001000010100011010**0**-00110010111000 **020**0**01110**
7-31a	0000000000000-00021000000000012000000210000**0**-00002100002100 **003**0**21000**
7-31b	0000000000000-00012000000000021000000120000**0**-00001200001200 **000**3**12000**
7-32a	0000000000000-00010000010000010100110001101-10001011010011 **102**1**10000**
7-32b	0000000000000-00010000100000100100110010011-01000111100011 **011**2**01000**
7-33	0000000000000-11100000001100100000000000000**0**-00110011001100 **110**0**00610**
7-34	0000000000000-00011000000000000100000000000**0**-11111111001111 **110**0**00121**
7-35	0000000000000-00000000010000000000000011000**0**-22111100220011 **000**0**00013**

lesser extent, a clear acoustical relationship between each heptatonic scale type and the inversion of its complement.

5   Example of Musical Analysis

An interesting chromatic excerpt is analyzed in [14]: the Fantasy in C minor, K. 475, by Mozart, mm. 1–25. The involved scales are determined there, although some of them are incomplete. With the nomenclature used in this paper, they are

> Ghh % FhM % DbM Ebmm FNpm Bbhm BbNpm BM % G#Npm
> DbhL F#Npm CbhL BhM Bhm GM % % F#Npm–F#hh F#hM % % %

where each scale or a pair linked by a dash lasts one measure, and symbol "%" means to repeat the previous measure. These scales are represented on the diagram with the global relationships of the hiatal group [14, Example 27], as it contains most of the scale types of this excerpt. However, the Ebmm scale (measure 6) could not be represented there, as it does not belong to that group, but to the octatonic and whole-tone groups. Thus, the scales of mm. 3–7 are then represented on the complex diagram with the global relationships of the three groups superimposed [14, Example 30].

For comparison, Fig. 5 shows the same excerpt on the 7-Cyclops, which includes all the required scales, the initial one being marked with a double line. As can be seen, although the composition is in C minor, none of the scales Chm, Cmm, or C natural minor (EbM) are used. Nevertheless, the initial scale, Ghh, includes the C minor chord and is played starting with C (its fourth mode). Then, the scales move counterclockwise in the diagram, that is, in the direction of flats (with back and forth movements), until reaching GM, whose tonic (at the bass) is the dominant of C. However, it does not resolve to any C minor scale, but to several scales with tonic F♯, a tritone away from C, the last four measures being based on the F♯ major chord.

Fig. 5. Mozart, fantasy in C minor, K. 475, mm. 1–25. Analysis with the 7-Cyclops.

6 Conclusions

Eight heptatonic scale types, together with their pentatonic complements, are selected following specific criteria. Their harmonic characteristics are analyzed and an acoustical relationship is found between each heptatonic scale type and the inversion of its pentatonic complement. Two novel parsimonious graphs, called 7-Cyclops and 5-Cyclops, are provided, which relate those heptatonic and pentatonic scales by single-semitonal transformations. Other parsimonious transformations, like moving one note by a whole tone, are easily found in them, too. As well, these graphs highlight the cycles of fifths and fourths, which are the only possible circumferences linking the same set types in this kind of graphs (apart from the trivial cases of set types with one or eleven notes). Finally, an example of musical analysis is included, so all this information is expected to be of interest both for theorists and composers.

References

1. Forte, A.: The Structure of Atonal Music. Yale University Press, New Haven (1973)
2. Nuño, L.: A detailed list and a periodic table of set classes. J. Math. Music **15**(3), 267–287 (2021)
3. Balzano, G.: The group-theoretic description of 12-fold and microtonal pitch systems. Comput. Music. J. **4**(4), 66–84 (1980)
4. Cohn, R.: Maximally smooth cycles, hexatonic systems, and the analysis of late-romantic triadic progressions. Music Anal. **15**(1), 9–40 (1996)
5. Schönberg, A.: Theory of Harmony, 3rd edn. University of California Press, Berkeley (1983)
6. Piston, W.: Harmony, 5th edn. W. W. Norton and Co., New York (1988)
7. Douthett, J., Steinbach, P.: Parsimonious graphs: a study in parsimony, contextual transformations, and modes of limited transposition. J. Music Theory **42**(2), 241–263 (1998)
8. Cohn, R.: Audacious Euphony: Chromatic Harmony and the Triad's Second Nature. Oxford University Press, New York (2012)
9. Tymoczko, D.: A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice. Oxford University Press, New York (2011)
10. Nuño, L.: Parsimonious graphs for the most common trichords and tetrachords. J. Math. Music **15**(2), 125–139 (2021)
11. Ring, I.: The Exciting Universe of Music Theory (2022). https://ianring.com/musictheory/scales/
12. Popley, H.A.: The Music of India. Oxford University Press, London (1921)
13. Hook, J.: Spelled heptachords. In: Agon, C., Andreatta, M., Assayag, G., Amiot, E., Bresson, J., Mandereau, J. (eds.) MCM 2011. LNCS (LNAI), vol. 6726, pp. 84–97. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21590-2_7
14. Žabka, M.: Dancing with the scales: subchromatic generated tone systems. J. Music Theory **58**(2), 179–233 (2014)

An Interactive Tool for Composing (with) Automorphisms in the Colored Cube Dance

Alexandre Popoff[1]([✉]) , Corentin Guichaoua[2] , and Moreno Andreatta[3,4]

[1] Paris, France
al.popoff@free.fr
[2] STMS (UMR9912), CNRS, Ircam, Sorbonne Université,
Ministère de la Culture, Paris, France
Corentin.Guichaoua@ircam.fr
[3] CNRS/Institute for Advanced Mathematical Research, ITI CREAA,
University of Strasbourg, Strasbourg, France
andreatta@math.unistra.fr
[4] IRCAM, Paris, France
andreatta@ircam.fr

Abstract. The "colored Cube Dance" is an extension of Douthett's and Steinbach's Cube Dance graph, related to a monoid of binary relations defined on the set of major, minor, and augmented triads. This contribution explores the automorphism group of this monoid action, as a way to transform chord progressions. We show that this automorphism group is of order 7776 and is isomorphic to $(\mathbb{Z}_3{}^4 \rtimes D_8) \rtimes (D_6 \times \mathbb{Z}_2)$. The size and complexity of this group makes it unwiedly: we therefore provide an interactive tool *via* a web interface based on common HTML/Javascript frameworks for students, musicians, and composers to explore these automorphisms, showing the potential of these technologies for math/music outreach activities.

Keywords: Cube Dance · Binary relations · Monoid action · Interactive software

1 An Algebraic Introduction to the Colored Cube Dance

The Cube Dance is a well-known structure introduced by Douthett and Steinbach in their work on parsimonious graphs between triads [4]. Its definition involves the $\mathcal{P}_{1,0}$ binary relation which relates two pitch-class sets if they differ by a single pitch class a semitone apart. The Cube Dance is then defined as the graph having the major, minor, and augmented triads as its vertices and the set of pairs of triads related by $\mathcal{P}_{1,0}$ as its set of edges. Since the classical neo-Riemannian P and L operations imply the $\mathcal{P}_{1,0}$ binary relation, some recent work [7] has investigated an extension of the Cube Dance wherein further refinements of the $\mathcal{P}_{1,0}$ relation are considered. More precisely, three binary relations
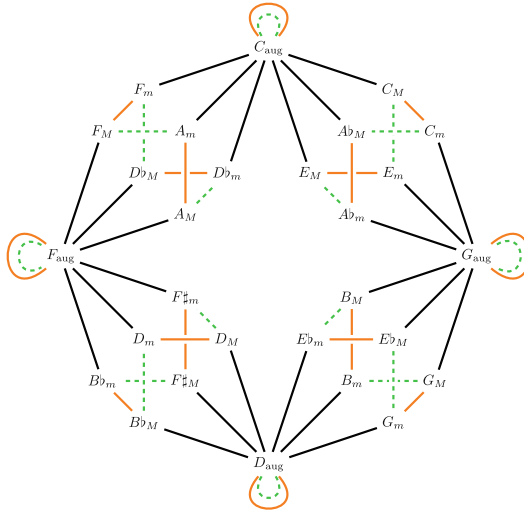

A. Popoff—Independent Researcher.

Fig. 1. The colored Cube Dance graph for the binary relations $\mathcal{U}$ (color: black/BW: black), $\mathcal{P}$ (color: orange/BW: gray), and $\mathcal{L}$ (color: dashed green/BW: dashed gray). (Color figure online)

$\mathcal{U}$, $\mathcal{P}$, and $\mathcal{L}$ are defined on the set of major, minor, and augmented triads as follows. The notation we adopt for these triads is of the form x_s, where x is a pitch class (the root for major, and minor triads, or any note for augmented triads), and s is a subscript ('M', 'm', or 'aug') indicating the type of triad.

Definition 1. *Let X be the set of the 24 major and minor triads and the four augmented triads.*

- *The relation $\mathcal{P}$ is the symmetric relation which coincides with the neo-Riemannian P operation on major and minor triads and is the identity relation on augmented triads.*
- *The relation $\mathcal{L}$ is the symmetric relation which coincides with the neo-Riemannian L operation on major and minor triads and is the identity relation on augmented triads.*
- *The relation $\mathcal{U}$ is the symmetric relation which relates an augmented triad with a major or minor triad if they are related by the $\mathcal{P}_{1,0}$ relation.*

The 'colored Cube Dance graph' (Fig. 1) is then defined as the graph having X as its set of vertices, and the set of pairs of triads related by either $\mathcal{U}$, $\mathcal{P}$, or $\mathcal{L}$ as its set of edges, each edge having a canonically attributed color in the set $\{\mathcal{U}, \mathcal{P}, \mathcal{L}\}$. From an algebraic point of view, these binary relations generate a monoid $M_{\mathcal{U},\mathcal{P},\mathcal{L}}$ with an action on X, which corresponds in categorical terms to the definition of a functor $S\colon M_{\mathcal{U},\mathcal{P},\mathcal{L}} \to \mathbf{Rel}$, and whose structure has been investigated in [7].

Proposition 1. *The monoid $M_{\mathcal{U},\mathcal{P},\mathcal{L}}$ generated by the relations $\mathcal{U}$, $\mathcal{P}$, and $\mathcal{L}$ contains 40 elements and has the following presentation.*

$$M_{\mathcal{U},\mathcal{P},\mathcal{L}} = \langle \mathcal{U},\mathcal{P},\mathcal{L} \mid \mathcal{P}^2 = \mathcal{L}^2 = e, \ \ \mathcal{LPL} = \mathcal{PLP}, \ \ \mathcal{U}^3 = \mathcal{U},$$
$$\mathcal{UP} = \mathcal{UL}, \ \ \mathcal{PU} = \mathcal{LU}, \ \ \mathcal{U}^2\mathcal{PU}^2 = \mathcal{PU}^2\mathcal{PU}^2\mathcal{P},$$
$$(\mathcal{UP})^2\mathcal{U}^2 = \mathcal{P}(\mathcal{UP})^2\mathcal{U}^2\mathcal{P}, \ \ \mathcal{U}^2(\mathcal{PU})^2 = \mathcal{PU}^2(\mathcal{PU})^2\mathcal{P}\rangle$$

The categorical point of view allows us to consider automorphisms of the functor S (i.e. automorphisms of the monoid action), whose general definition has been given in [6,7] and which simplifies in this case as follows.

Definition 2. *The automorphism group $Aut(S)$ of the functor $S\colon M_{\mathcal{U},\mathcal{P},\mathcal{L}} \to$ **Rel** is the group of pairs (N,ν) where $N\colon M_{\mathcal{U},\mathcal{P},\mathcal{L}} \to M_{\mathcal{U},\mathcal{P},\mathcal{L}}$ is an automorphism, and ν is a bijection on X such that we have $p\mathcal{R}q \implies \nu(p)N(\mathcal{R})\nu(q)$ for all $\mathcal{R} \in M_{\mathcal{U},\mathcal{P},\mathcal{L}}$ and $(p,q) \in X^2$. Composition is done term-wise.*

It should be noted that the normal subgroup of $Aut(S)$ of automorphisms of the form (id,ν) is isomorphic to the normal subgroup of graph automorphisms leaving the color of edges invariant. Some of these automorphisms are notably involved in chord progressions in pop music [1,6]. The computation of the full automorphism group of the monoid action is therefore of interest, giving tools for musicians and composers to transform chord progressions in the colored Cube Dance. We establish its structure in the next Section.

2 The Automorphism Group of the Monoid Action of $M_{\mathcal{U},\mathcal{P},\mathcal{L}}$

The automorphism group of the monoid $M_{\mathcal{UPL}}$ itself has been determined in [7].

Theorem 1. *The automorphism group of the $M_{\mathcal{UPL}}$ monoid is isomorphic to the group $D_6 \times \mathbb{Z}_2$.*

Each automorphism N of $M_{\mathcal{UPL}}$ is entirely determined by an automorphism of the subgroup isomorphic to D_6 generated by $\mathcal{P}$ and $\mathcal{L}$, and by the choice of the image of $\mathcal{U}$ by N in the set $\{\mathcal{U},\mathcal{LUL}\}$. The main result of this paper is the structure of the automorphism group of the monoid action $S\colon M_{\mathcal{UPL}} \to$ **Rel**.

Theorem 2. *The automorphism group of the functor $S\colon M_{\mathcal{UPL}} \to$ **Rel** is a group of order 7776 isomorphic to $(\mathbb{Z}_3{}^4 \rtimes D_8) \rtimes (D_6 \times \mathbb{Z}_2)$.*

Proof. We sketch here the methodology for the proof, leaving the full enumeration of the cases to the reader. We denote by $\overline{C_M}$ the set $\{C_M, E_M, Ab_M\}$, by $\overline{F_m}$ the set $\{F_m, A_m, Db_m\}$, and so on.

Let N be an automorphism of $M_{\mathcal{UPL}}$. Assume for example that $N(\mathcal{U}) = \mathcal{U}$. We then look for the possible bijections ν of X: these will obviously map the subset $\{C_{\mathrm{aug}}, G_{\mathrm{aug}}, D_{\mathrm{aug}}, F_{\mathrm{aug}}\}$ onto itself. At this point, we can freely choose the image of C_{aug} by ν: assume for example that $\nu(C_{\mathrm{aug}}) = G_{\mathrm{aug}}$. Since C_{aug}

Table 1. Graphical representation of the possible automorphisms (N, ν) of the functor $S \colon M_{\mathcal{U},\mathcal{P},\mathcal{L}} \to \mathbf{Rel}$. The mapping of subsets X_M is determined by the permutation of augmented chords, by the image of the generator $\mathcal{U}$ by N, and by the action of the group elements g_i in $\mathbb{Z}_3$.

$N(\mathcal{U}) = \mathcal{U}$	$N(\mathcal{U}) = \mathcal{L}\mathcal{U}\mathcal{L}$

The table consists of nine rows of graphical diagrams. Each cell contains braiding/crossing diagrams connecting the following labels.

Top labels (augmented chords): C_{aug}, G_{aug}, D_{aug}, F_{aug} and $\overline{C_M}$, $\overline{G_M}$, $\overline{D_M}$, $\overline{F_M}$

Bottom labels: C_{aug}, G_{aug}, D_{aug}, F_{aug} and $\overline{C_M}$, $\overline{G_M}$, $\overline{D_M}$, $\overline{F_M}$ (or C_m, G_m, D_m, F_m)

Group element labels g_i appear between diagram pairs, drawn from g_0, g_1, g_2, g_3.

is related to the elements in $\overline{C_M} \cup \overline{F_m}$ by the relation $\mathcal{U}$, it is implied by the definition of ν and the fact that $N(\mathcal{U}) = \mathcal{U}$ that $\overline{C_M} \cup \overline{F_m}$ should be bijectively mapped by to the subset $\overline{G_M} \cup \overline{C_m}$. Since the subset $\overline{C_M}$ (resp. $\overline{F_m}$) is an orbit of C_M by the subgroup of D_6 isomorphic to $\mathbb{Z}_3$ generated by $\mathcal{LP}$, we conclude by the definition of ν that either $\overline{C_M}$ is mapped to $\overline{G_M}$ and $\overline{F_m}$ to $\overline{C_m}$, or the other way around. Assume the first case, and note that each of this mapping is entirely determined by the choice of a representative element in each subset and an element of the subgroup of D_6 isomorphic to $\mathbb{Z}_3$ generated by $\mathcal{LP}$. We then have that $\overline{C_m}$ is mapped to $\overline{G_m}$ and $\overline{F_M}$ to $\overline{C_M}$. Since the elements of $\overline{C_m}$ are related to G_{aug} by $\mathcal{U}$, it is implied by the definition of ν and the fact that $N(\mathcal{U}) = \mathcal{U}$ that G_{aug} should be mapped to D_{aug}. Similarly, we get that F_{aug} should be mapped to C_{aug}. By continuing this enumeration, we arrive at the graphical representations of automorphisms given in Table 1. The group of permutations of the set $\{C_{\mathrm{aug}}, G_{\mathrm{aug}}, D_{\mathrm{aug}}, F_{\mathrm{aug}}\}$ is isomorphic to D_8, and this automatically determines the permutation of the set of subsets $\{\overline{C_M}, \overline{G_M}, \overline{D_M}, \overline{F_M}\}$. For a given permutation of this set, each element g_i is a group element in $\mathbb{Z}_3$ determining how the subsets are mapped, assuming a set of representative elements has been fixed beforehand. It can then readily be seen that $\mathrm{Aut}(S)$ is isomorphic to $(\mathbb{Z}_3{}^4 \rtimes D_8) \rtimes (D_6 \times \mathbb{Z}_2)$, a group of order $(3^4 * 8) * (6 * 2) = 7776$. □

3 An Interactive Interface for Composing (with) Automorphisms

Contrary to the neo-Riemannian PLR group, the size and complexity of $\mathrm{Aut}(S)$ makes it hard to use with pen and paper, especially for non-mathematicians wish-
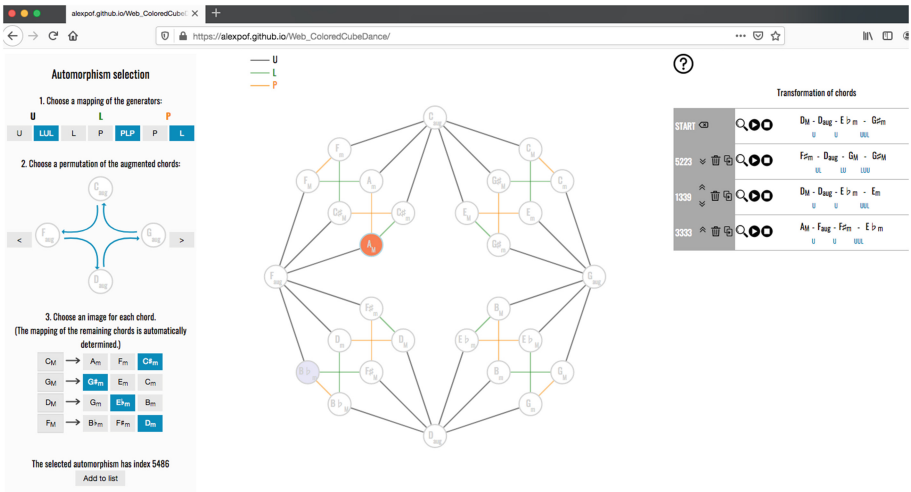

Fig. 2. Screenshot of the web interface for manipulating the automorphisms of the colored Cube Dance.

ing to explore its potential for transforming chord progressions. In this light, we have developed a web interface for the concrete manipulation of the elements of Aut(S), and for their use as chord transformations. A screenshot of this interface is shown in Fig. 2. It uses common HTML and Javascript frameworks [3], thus making it runnable on virtually any web browser without the need for complicated software installations. Such frameworks have already been used for other mathematics/music applications, notably to explore the Tonnetz [5]. The web interface can directly be used from the corresponding GitHub repository [2] and the associated source code is freely available.

As shown on Fig. 2, the left part of the interface corresponds to the interactive choice of an automorphism of the colored Cube Dance. The middle part is an interactive colored Cube Dance: alt-clicking on chords adds them to the current chord progression, which is shown on the right part of the interface, along with its successive transformation by the selected automorphisms. Each chord progression can be played back with the corresponding buttons. In the future, the interface will also feature MIDI capabilities, so that chord progressions could be recorded, transformed, and replayed at will (see [5] for a current implementation of such capabilities). Following the pattern of Table 1, the user first selects a mapping of the generators of $M_{\mathcal{UPL}}$, then a permutation of the augmented chords, and finally a mapping of the major/minor chords through the mapping of a given representative in each quadrant. Once an automorphism has been uniquely determined, the user can hover over chords in the middle representation of the colored Cube Dance to see how they are mapped by the selected element of Aut(S). The 'add to list' button appends the selected automorphism to the list on the right, in which the current chord progression is successively transformed through automorphism composition.

The combination of SVG graphics possibilities in HTML with Javascript allows one to quickly develop user-friendly interfaces for math/music concepts, thus showing the potential of these technologies for outreach activities. It is our hope that the colored Cube Dance web interface will prove useful for students, musicians, and composers to creatively explore chord transformations via automorphisms.

References

1. (2021). https://www.youtube.com/watch?v=nz5TYob02B4&ab_channel=MatheMusic4D. Accessed 10 Jan 2022
2. (2021). https://alexpof.github.io/Web_ColoredCubeDance. Accessed 10 Jan 2022
3. Bostock, M.: D3.js - data-driven documents (2012). http://d3js.org/. Accessed 10 Jan 2022
4. Douthett, J., Steinbach, P.: Parsimonious graphs: a study in parsimony, contextual transformations, and modes of limited transposition. J. Music Theory **42**(2), 241–263 (1998). http://www.jstor.org/stable/843877
5. Guichaoua, C., Besada, J., Bisesi, E., Andreatta, M.: The tonnetz environment: a web platform for computer-aided "mathemusical" learning and research. In: Proceedings of the 13th International Conference on Computer Supported Education - Volume 1: CSME, pp. 680–689. INSTICC, SciTePress (2021)

6. Popoff, A.: On the use of relational presheaves in transformational music theory. J. Math. Music **16**(1), 51–79 (2022). https://doi.org/10.1080/17459737.2020.1825845
7. Popoff, A., Andreatta, M., Ehresmann, A.: Relational poly-Klumpenhouwer networks for transformational and voice-leading analysis. J. Math. Music **12**(1), 35–55 (2018). https://doi.org/10.1080/17459737.2017.1406011

Combinatorial Spaces

Robert W. Peck(✉) [ORCID]

Louisiana State University, Baton Rouge, LA 70803, USA
rpeck@lsu.edu

Abstract. Combinatoriality—the property that obtains when unions of corresponding subsets within tone rows comprise aggregates—takes various forms, following the canonical operations that relate the constituent rows to one another: transposition, inversion, retrograde, and/or retrograde inversion. The mathematical field of combinatorics presents tools to answer such basic questions as: How many combinatorial sets exist in a space of a given size? To how many equivalence classes do they belong? Such enumeration procedures involve various techniques that have prior connections to music theory. In the process of answering these questions, our results reveal further aspects of combinatorial sets. For instance, no combinatorial n-chords are held invariant by a translation operation with an odd index. The set of I-invariant n-chords that are P-combinatorial is equivalent to the set of those that are I-combinatorial, and this set is precisely the set of all-combinatorial n-chords. Such information sheds new light on these intriguing structures.

Keywords: Combinatoriality · Serialism · Combinatorics · Enumeration

1 Introduction

Combinatoriality in serial music takes various forms, following the canonical operations that relate constituent tone rows to one another: prime or transposition (P), inversion (I), retrograde (R), and/or retrograde inversion (RI). Inversional combinatoriality, or I combinatoriality, is of particular historical significance, as it characterizes much of Arnold Schoenberg's twelve-tone music. Among the tone rows in his forty-two twelve-tone compositions, thirty-six (85.7%) use hexachords that produce I combinatoriality. Regarding the basic set of his *Variations for Orchestra*, op. 31, Schoenberg writes [13, p. 116]: "the inversion a fifth below of the first six tones, the antecedent, should not reproduce a repetition of one of these six tones, but should bring forth the hitherto unused six tones of the chromatic scale. Thus, the consequent of the basic set...comprises the tones of this inversion, but, of course, in a different order," as shown here in Fig. 1.

Specifically, the tone row from Schoenberg's op. 31 is combinatorial under the pitch-class operation I_5. To maintain the complement relation between the hexachords, no two pitch classes that relate by I_5 can be present in the same

Basic set, antecedent: (10, 4, 6, 3, 5, 9)   Basic set, consequent: (2, 1, 7, 8, 11, 0)

Inversion, antecedent: (7, 1, 11, 2, 0, 8)   Inversion, consequent (3, 4, 10, 9, 6, 5)
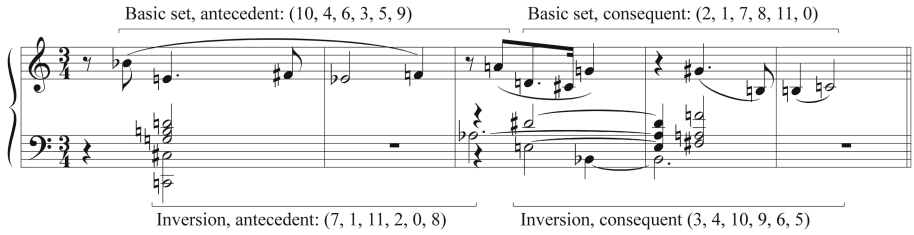
Fig. 1. Hexachordal I combinatoriality in the Thema to Schoenberg's *Variations for Orchestra*, op. 31, mm. 34–38.

hexachord, as those pitch classes map onto one another under that operation. Figure 2 depicts the members of the row's two hexachords as beads in a binary necklace; the white beads represent the pitch classes of the first hexachord and the black beads represent those of the second. We note that the necklace balances across the I_5 axis: for each pitch class c of one hexachord, a corresponding pitch class $d = 11c + 5 \pmod{12}$ from the other hexachord exists directly across the axis. We can represent any partition of the twelve-tone aggregate into I_5-combinatorial hexachords in this way. Therefore, as we find two possible positions relative to the I_5 axis for any one of the six $\{c, d\}$ pairs, we note that there exist $2^6 = 64$ hexachords that are I_5-combinatorial.
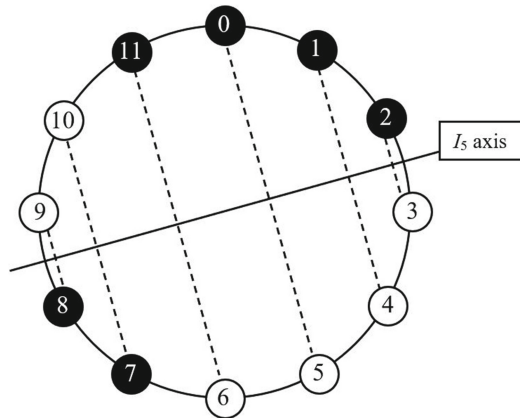
Fig. 2. The tone row of Schoenberg's *Variations for Orchestra*, op. 31, as a binary necklace, balanced across the I_5 axis (first hexachord in white, second hexachord in black).

Whereas we find sixty-four I_x-combinatorial hexachords for each one of the six odd values of x, we note that there exist fewer than $64 \times 6 = 384$ I-combinatorial hexachords in total, as some of these hexachords are combinatorial

under more than one I_x operator. Consulting a standard post-tonal textbook, such as [16], we count 348 pitch-class sets that have this property and note that those hexachords belong to nineteen set classes, but how may we arrive at these numbers computationally? Furthermore, composers in the twentieth century make use of additional types of combinatoriality. For example, Milton Babbitt incorporates "all-combinatorial" sets frequently in his compositions. These sets display all four of the canonical combinatorial types: P, I, R, and RI. How might we obtain similar results for these other sorts of combinatorial sets or for combinatorial sets in modular spaces of sizes other than twelve?

The mathematical field of combinatorics presents tools to answer such basic questions as: How many combinatorial sets of any type exist in a space of a given size? What are their symmetries? To how many equivalence classes do they belong? In the process of answering these questions, our results reveal further aspects of combinatorial sets. The general notion of combinatoriality is not limited to serial procedures or to twelve-element aggregates. The defining concepts that it brings together—complementation and equivalence under translation and reflection—are of broad musical interest, as both are interval-preserving when the integrant sets are of the same cardinality (per the Generalized Hexachord Theorem, see [17]). The concepts manifest in combinatoriality apply to numerous musical parameters in addition to pitch, such as rhythmic structure. Further, the procedures we use to study combinatorial structures incorporate various techniques that have prior connections to music theory (e.g., [3,6], and [7]), including the enumeration of serial structures, linking this inquiry with the investigation of other aspects of musical structure. In particular, [4, especially pp. 135–158] presents a detailed enumeration of tone rows in the standard 12-tone chromatic space; further, [4, p. 161] enumerates 12-tone tropes (following [5]) according to different types of combinatoriality (including P-, I-, R-, RI-, and all-combinatoriality), though using different combinatorial techniques from those in the present study.

2 Music-Theoretical and Mathematical Background

In this section, we give basic information that will apply to later sections. Further detailed information on the mathematical theory of musical serialism, particularly from the perspective of combinatorics, can be found in [4].

Let $\mathbb{Z}_{2n}$ be a modular space of elements (typically pitch classes), called the aggregate. Let $\mathscr{S}_{2n}$ be the set of all orderings of the $2n$ elements of that space. $\mathscr{S}_{2n}$ is of size $(2n)!$. Call $S \in \mathscr{S}_{2n}$ a $2n$-tone row, where $(s_0, s_1, ..., s_{2n-1})$ is the particular ordering of elements within the row. G is the canonical group of serial operations with an action on $\mathscr{S}_{2n}$, generated by unit transposition $T_1 := s_i \mapsto s_i + 1$; inversion $I_x := s_i \mapsto (2n-1)s_i + x$, where $x \in \mathbb{Z}_{2n}$; and order-position retrograde $R_x : s_i \mapsto s_{2n-1-i} + x$, where $x \in \mathbb{Z}_{2n}$. G is of order $8n$.

We call an unordered subset $N \subset \mathbb{Z}_{2n}$ an n-chord if $|N| = n$. $\mathscr{N}_{2n}$ is the set of all n-chords in $\mathbb{Z}_{2n}$. We call the orbit of N under the action of the group H of transposition-and-inversion operators, $H(N)$, a set class, following [2]. $\mathscr{N}_{2n}/H$ is

the set of all n-chordal set classes under the action of H on $\mathscr{N}_{2n}$, and $\mathscr{N}_{2n}^h$ is the set of n-chords that are stabilized by the element $h \in H$. $\bar{N}$ is the complement of N in $\mathbb{Z}_{2n}$, and we note that $\left|\bar{N}\right| = |N|$.

We are concerned with five types of n-chordal combinatoriality: P-, I-, R-, RI-, and all-combinatoriality.

Definition 1. *Given a $2n$-tone row $S \in \mathscr{S}_{2n}$, where $N = \{s_0, s_1, ..., s_{n-1}\}$, S has the property $\boldsymbol{n\text{-}chordal\ combinatoriality}$ if and only if there exists some $g \in G$ such that $\bar{N} = \{g(s_0), g(s_1), ..., g(s_{n-1})\}$. Then, we call N $\boldsymbol{P\text{-}}$ $\boldsymbol{combinatorial}$ if there exists some $x \in \mathbb{Z}_{12}$ such that $\bar{N} = T_x(N)$; N is $\boldsymbol{I\text{-}}$ $\boldsymbol{combinatorial}$ if there exists some $x \in \mathbb{Z}_{12}$ such that $\bar{N} = I_x(N)$; N is $\boldsymbol{R\text{-}}$ $\boldsymbol{combinatorial}$ if there exists some $x \in \mathbb{Z}_{12}$ such that $N = T_x(N)$; and N is $\boldsymbol{RI}\text{-}\boldsymbol{combinatorial}$ if there exists some $x \in \mathbb{Z}_{12}$ such that $N = I_x(N)$. Finally, N is $\boldsymbol{all\text{-}combinatorial}$ if all four of the preceding statements are true.*

Other forms of combinatoriality involve aggregates formed as unions of $m > 2$ n-chords; in these cases, the relevant space is of size mn. (For example, in $\mathbb{Z}_{12}$, we may use trichordal combinatoriality, in which the aggregate comprises four images of a set of cardinality $n = 3$; see [14,15]). In this study, however, we consider only the special case of n-chordal combinatoriality that results from the unions of two n-chords; hence, we observe that such n-chordal combinatoriality obtains only in spaces with even-parity size.

As all $2n$-tone rows are trivially combinatorial under order-position retrograde without transposition R_0 [1, p. 91], we note that the full set of combinatorial n-chords in $\mathbb{Z}_{2n}$ is equivalent to $\mathscr{N}_{2n}$ itself. Therefore, we distinguish between particular subsets of $\mathscr{N}_{2n}$: the subsets of P-combinatorial n-chords ($\mathscr{N}_{2n(P)}$), I-combinatorial n-chords ($\mathscr{N}_{2n(I)}$), R-combinatorial n-chords ($\mathscr{N}_{2n(R)}$), RI- combinatorial n-chords ($\mathscr{N}_{2n(RI)}$), and all-combinatorial n-chords ($\mathscr{N}_{2n(\text{all})}$).

Our enumeration incorporates various standard results from the mathematical fields of combinatorics, group theory, and number theory. Many of our formulae use powers of 2, which we use to count binary strings, as in our example of the 2^6 I_5-combinatorial hexachords as binary necklaces in Sect. 1 above. Concerning the powers of 2, certain of our results use the 2-adic order of n.

Definition 2. *Given a prime number p, the $\boldsymbol{p\text{-}adic\ order}$ of the integer n is the highest exponent ν_p such that $p^{\nu_p} \mid n$. (If $p^{\nu_p} \nmid n$, then $\nu_p = 0$, since $p^0 = 1$.)*

The formula n-choose-k counts the number of all k-subsets of an n-set.

Definition 3. *Binomial coefficient.*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

We use the Möbius μ-function, which eliminates redundancies in reckoning the sizes of various sets of combinatorial n-chords by incorporating 0 and -1 among its three coefficients as potential multipliers.

Definition 4. *Möbius μ-function.*

$$\mu(n) = \begin{cases} 1, & \text{if } n = 1 \\ 0, & \text{if } n \text{ has a square prime factor} \\ (-1)^r, & \text{if } n \text{ has } r \text{ distinct prime factors.} \end{cases}$$

Moreover, we use the Cauchy-Frobenius Lemma (Lemma 1) to determine numbers of orbits in the action of a finite group G on a finite set S (e.g., music-theoretical set classes). Here, S^g is the set of all elements in S that are stabilized by $g \in G$.

Lemma 1. *Cauchy-Frobenius*

$$|S/G| = \frac{1}{|G|} \sum_{g \in G} |S^g|$$

Finally, we introduce two theorems that relate to each of the cases in the next section.

Theorem 1. $\left| \mathscr{N}_{2n}^{T_{2x}} \right| = \left| \mathscr{N}_{2\gcd(n,x)} \right|.$

Proof. As a cyclic group, the action of the group generated by T_{2x} on $\mathbb{Z}_{2n}$, $x \in \mathbb{Z}_{2n}$, partitions $\mathbb{Z}_{2n}$ into $2\gcd(x,n)$ orbits of size $n/\gcd(x,n)$. Then, an n-set $N \subseteq \mathbb{Z}_{2n}$ is stabilized by T_{2x} if and only if N is formed by a union of $\gcd(x,n)$ of these orbits. Hence, we find $\binom{2\gcd(x,n)}{\gcd(x/n)}$ possibilities for a T_{2x}-symmetrical N, which is equivalent to $\left| \mathscr{N}_{2\gcd(n,x)} \right|$, as any $\mathscr{N}_{2n}$ contains $\binom{2n}{n}$ n-chords by definition. □

Theorem 2. $\left| \mathscr{N}_{2n}^{T_{2x+1}} \right| = 0.$

Proof. As a cyclic group, the action of the group generated by T_{2x+1} on $\mathbb{Z}_{2n}$, $x \in \mathbb{Z}_{2n}$, partitions $\mathbb{Z}_{2n}$ into $\gcd(2x+1, 2n)$ orbits of size $2n/\gcd(2x+1, 2n)$. Then, an n-set $N \subseteq \mathbb{Z}_{2n}$ is stabilized by T_{2x+1} if and only if N is formed by a union of $\gcd(2x+1, 2n)/2$ of these orbits. However, as $\gcd(2x+1, 2n)/2$ is odd, it is not possible to write N as the union of $\gcd(2x+1, 2n)/2$ orbits. □

3   Results and Applications

3.1   *I* Combinatoriality

We begin with I combinatoriality, which is the most commonly studied type because of Schoenberg's frequent incorporation of it. As such, it serves as a useful introduction to our applications. We noted in Sect. 1 that $2^6 = 64$ I-combinatorial hexachords exist for each of the six odd-indexed inversion operators in $\mathbb{Z}_{12}$, yet we find fewer than $64 \cdot 6 = 384$ I-combinatorial hexachords in total, and that the reason for this discrepancy is the fact that certain hexachords are combinatorial under several different I_x operators. The same situation

exists in any space $\mathbb{Z}_{2n}$. First, I combinatoriality is possible only under inversion operators with odd indices, as even-indexed inversion operators always hold two elements of $\mathbb{Z}_{2n}$ invariant (for instance, see [16, p. 316]). Then, we find 2^n I-combinatorial n-chords for any odd-indexed inversion operator I_{2x+1}—hence, $\left|\mathscr{N}_{2n(I)}\right| \leq 2^n n$—but, again, certain of these n-chords are combinatorial under more than one inversion operator.

Ultimately, per Lemma 1, to reckon the number of set classes to which the members of the set of $\mathscr{N}_{2n(I)}$ belong, we need to determine how many I-combinatorial n-chords are stabilized by each member of the transposition-and-inversion group H. (The set classes are orbits in the action of H on $\mathscr{N}_{2n(I)}$.) The following equation, derived from [9], which counts the number of $2n$-bead balanced binary strings that are rotationally equivalent to reversed complement, determines the number of I-combinatorial N-chords that are stabilized by even-indexed transposition operators, including T_0.

$$\left|\mathscr{N}_{2n(I)}^{T_{2x}}\right| = \sum_{j|\gcd(x,n)} \sum_{k|j} \mu(k)2^{j/k}j \tag{1}$$

As with our observation in Sect. 1 that the number of I-combinatorial n-chords that are stabilized by any one particular I_{2x+1} operator is a power of 2, a power of 2 serves also as the basis of Eq. 1. Then, the Möbius μ-function (Definition 4) eliminates redundancies from n-chords that are combinatorial under multiple values of I_{2x+1}. As an example, the following application illustrates the numbers of I-combinatorial hexachords in $\mathbb{Z}_{12}$ that are stabilized by the identity element $T_{2x=0}$.

– For $j = 6$:

$$
\begin{aligned}
k &= 1 : (1 \cdot 2^6) \cdot 6 \quad= 384 \\
k &= 2 : (-1 \cdot 2^3) \cdot 6 = -48 \\
k &= 3 : (-1 \cdot 2^2) \cdot 6 = -24 \\
k &= 6 : (1 \cdot 2^1) \cdot 6 \quad= \frac{+12}{324}
\end{aligned}
$$

– For $j = 3$:

$$
\begin{aligned}
k &= 1 : (1 \cdot 2^3) \cdot 3 \quad= 24 \\
k &= 3 : (-1 \cdot 2^1) \cdot 3 \;= \frac{-6}{18}
\end{aligned}
$$

– For $j = 2$:

$$
\begin{aligned}
k &= 1 : (1 \cdot 2^2) \cdot 2 \quad= 8 \\
k &= 2 : (-1 \cdot 2^1) \cdot 2 \;= \frac{-4}{4}
\end{aligned}
$$

– For $j = 1$:

$$k = 1 : (1 \cdot 2^1) \cdot 1 \quad = \quad 2$$

It yields 324 T_0-symmetric hexachords that are combinatorial under precisely one inversion operator, eighteen hexachords that are combinatorial under two, four hexachords that are combinatorial under three, and two hexachords that are combinatorial under all six odd-indexed inversion operators, for a total of 348, the size of $\mathscr{N}_{12(I)}$. In this way, we may determine the numbers of I-combinatorial n-chords that are stabilized by any other even-indexed transposition operator.

The next two equations determine the number of I-combinatorial n-chords that are stabilized by inversion operators with even and odd indices, respectively.

$$\left| \mathscr{N}_{2n(I)}^{I_{2x}} \right| = 2^{((n/2^{\nu_2(n)})+1)/2} \tag{2}$$

$$\left| \mathscr{N}_{2n(I)}^{I_{2x+1}} \right| = 2\alpha(n) - \left| \mathscr{N}_{2n(I)}^{I_{2x}} \right|, \tag{3}$$

$$\text{where} \quad \begin{cases} \alpha(0) & = 1 \\ \alpha(2n) & = \alpha(n) + 2^{n-1}, \text{ for } n > 0 \\ \alpha(2n+1) & = 2^n, \text{ for } n \geq 0 \end{cases}$$

The first equation incorporates the 2-adic order of n (Definition 2). The second uses the α-function [10], which determines the number of $2n$-bead balanced binary necklaces which are equivalent to their reverse, complement, and reversed complement. In this case, we note that $2\alpha(n)$ counts the total number of I-combinatorial n-chords that are stabilized by both I_{2x} and I_{2x+1} for a specific value of $x \in \mathbb{Z}_n$, so it is necessary to subtract the number of I-combinatorial n-chords that are stabilized by the even-indexed inversion operator I_{2x} to determine the number of those stabilized by an odd-indexed inversion operator. For instance, given $n = 6$, Eq. 2 yields four I-combinatorial hexachords that are stabilized by an even-indexed inversion operator I_{2x}. For odd-indexed inversions, Eq. 3 yields eight hexachords for I_{2x+1}, as $\alpha(6) = 6$; hence, $2\alpha(6) - \left| \mathscr{N}_{12(I)}^{I_{2x}} \right| = 8$.

Table 1 presents a summary of all the values for stabilized hexachords in the familiar example of $n = 6$ in $\mathbb{Z}_{12}$. Thus, by Lemma 1, the number of set classes to which the members of the set $\mathscr{N}_{12(I)}$ belong is the average number of hexachords stabilized by twenty-four members of the transposition and inversion group H, or nineteen (see Eq. 4).

$$\frac{348 + (2 \cdot 2) + (6 \cdot 2) + 20 + (4 \cdot 6) + (8 \cdot 6)}{24} = 19 \tag{4}$$

Finally, Table 2 shows the results of applying this enumeration to cases in which $n \leq 12$.

Table 1. Sizes of $\mathcal{N}_{12(I)}^{h}$ for each member $h \in H$.

$$\left|\mathcal{N}_{12(I)}^{T_0}\right| = 348 \qquad \left|\mathcal{N}_{12(I)}^{T_1}\right| = 0 \qquad \left|\mathcal{N}_{12(I)}^{I_0}\right| = 4 \qquad \left|\mathcal{N}_{12(I)}^{I_1}\right| = 8$$

$$\left|\mathcal{N}_{12(I)}^{T_2}\right| = 2 \qquad \left|\mathcal{N}_{12(I)}^{T_3}\right| = 0 \qquad \left|\mathcal{N}_{12(I)}^{I_2}\right| = 4 \qquad \left|\mathcal{N}_{12(I)}^{I_3}\right| = 8$$

$$\left|\mathcal{N}_{12(I)}^{T_4}\right| = 6 \qquad \left|\mathcal{N}_{12(I)}^{T_5}\right| = 0 \qquad \left|\mathcal{N}_{12(I)}^{I_4}\right| = 4 \qquad \left|\mathcal{N}_{12(I)}^{I_5}\right| = 8$$

$$\left|\mathcal{N}_{12(I)}^{T_6}\right| = 20 \qquad \left|\mathcal{N}_{12(I)}^{T_7}\right| = 0 \qquad \left|\mathcal{N}_{12(I)}^{I_6}\right| = 4 \qquad \left|\mathcal{N}_{12(I)}^{I_7}\right| = 8$$

$$\left|\mathcal{N}_{12(I)}^{T_8}\right| = 6 \qquad \left|\mathcal{N}_{12(I)}^{T_9}\right| = 0 \qquad \left|\mathcal{N}_{12(I)}^{I_8}\right| = 4 \qquad \left|\mathcal{N}_{12(I)}^{I_9}\right| = 8$$

$$\left|\mathcal{N}_{12(I)}^{T_{10}}\right| = 2 \qquad \left|\mathcal{N}_{12(I)}^{T_{11}}\right| = 0 \qquad \left|\mathcal{N}_{12(I)}^{I_{10}}\right| = 4 \qquad \left|\mathcal{N}_{12(I)}^{I_{11}}\right| = 8$$

Table 2. Numbers of I-combinatorial n-chords and their set classes in spaces $\mathbb{Z}_{2n}$, $n \leq 12$.

Space	$\mathbb{Z}_2$	$\mathbb{Z}_4$	$\mathbb{Z}_6$	$\mathbb{Z}_8$	$\mathbb{Z}_{10}$	$\mathbb{Z}_{12}$	$\mathbb{Z}_{14}$	$\mathbb{Z}_{16}$	$\mathbb{Z}_{18}$	$\mathbb{Z}_{20}$	$\mathbb{Z}_{22}$	$\mathbb{Z}_{24}$
$n =$	1	2	3	4	5	**6**	7	8	9	10	11	12
$\left|\mathcal{N}_{2n(I)}\right|$	2	6	20	54	152	**348**	884	1974	4556	10056	22508	48636
$\left|\mathcal{N}_{2n(I)}/H\right|$	1	2	3	6	10	**19**	36	70	136	266	528	1043

3.2 *P* Combinatoriality

P combinatoriality results when the union of two n-chords that relate by transposition form an aggregate. Hence, there exists some value(s) of $x \in \mathbb{Z}_{2n}$ for which $T_x(N) = \bar{N}$. As with I-combinatorial n-chords, our enumeration of P-combinatorial n-chords derives from the numbers of n-chords that are stabilized by various members of the transposition-and-inversion group. Equation 5 gives the number of n-chords that are stabilized by an even-indexed transposition operator T_{2x}. It incorporates the β-function [8], which determines the number of $2n$-bead balanced binary strings that are rotationally equivalent to their complement.

$$\left|\mathcal{N}_{2n(P)}^{T_{2x}}\right| = \beta(\gcd(n, x)), \tag{5}$$

$$\text{where} \quad \begin{cases} \beta(0) & = 1 \\ \beta(2n) & = \beta(n) + 2^{2n}, \text{ for } n > 0 \\ \beta(2n+1) & = 2^{2n+1}, \text{ for } n \geq 0 \end{cases}$$

For instance, as $\gcd(6, 0) = 6$ and $\beta(6) = 72$, we use Eq. 5 to determine that there exist 72 P-combinatorial hexachords in $\mathbb{Z}_{12}$ that are stabilized by the identity element T_0.

Regarding the numbers of P-combinatorial n-chords that are stabilized by inversion operators with even and odd indices, we note the following result, which accounts for both cases.

Theorem 3. $\mathscr{N}^{I_x}_{2n(P)} = \mathscr{N}^{I_x}_{2n(I)}$

Proof. Assume that $I_y(N) = \bar{N}$ for some $y \in \mathbb{Z}_{2n}$. Then, by definition, there exists some inversion operation I_{2x+1}, $x \in \mathbb{Z}_{2n}$, such that

$$\begin{aligned}
\bar{N} &= I_{2x+1}(N) \\
&= I_{2x+1}(I_y(N)) \\
&= (I_{2x+1}I_y)(N) \\
&= T_{2x+1-y}(N).
\end{aligned}$$

As I_y stabilizes N and I_{2x+1} maps N to $\bar{N}$, we observe that $2x + 1 \neq y$. Therefore, $T_{2x+1-y} \neq T_0$, so N is also P-combinatorial (by definition). By the same reasoning, the reverse is true: if N is P-combinatorial, then it is also I-combinatorial. □

Corollary 1. $\mathscr{N}_{2n(all)} = \mathscr{N}^{I_x}_{2n(I)}$

Proof. Every all-combinatorial n-chord must belong to $\mathscr{N}^{I_x}_{2n(I)}$, $x \in \mathbb{Z}_{2n}$, by definition. The members of $\mathscr{N}^{I_x}_{2n(I)}$ are I-combinatorial, also by definition. Theorem 3 determines further that they are P-combinatorial. As all n-chords are trivially R-combinatorial, we may combine these facts to ascertain that the members of $\mathscr{N}^{I_x}_{2n(I)}$ are also RI-combinatorial; hence, they are all-combinatorial. □

For example, we reckon the numbers of P-combinatorial hexachords that are stabilized by the twenty-four elements of the usual transposition-and-inversion group's action on $\mathbb{Z}_{12}$ (see Table 3).

Table 3. Sizes of $\mathscr{N}^h_{12(P)}$ for each member $h \in H$.

$\left\|\mathscr{N}^{T_0}_{12(P)}\right\| = 72$	$\left\|\mathscr{N}^{T_1}_{12(P)}\right\| = 0$	$\left\|\mathscr{N}^{I_0}_{12(P)}\right\| = 4$	$\left\|\mathscr{N}^{I_1}_{12(P)}\right\| = 8$
$\left\|\mathscr{N}^{T_2}_{12(P)}\right\| = 2$	$\left\|\mathscr{N}^{T_3}_{12(P)}\right\| = 0$	$\left\|\mathscr{N}^{I_2}_{12(P)}\right\| = 4$	$\left\|\mathscr{N}^{I_3}_{12(P)}\right\| = 8$
$\left\|\mathscr{N}^{T_4}_{12(P)}\right\| = 6$	$\left\|\mathscr{N}^{T_5}_{12(P)}\right\| = 0$	$\left\|\mathscr{N}^{I_4}_{12(P)}\right\| = 4$	$\left\|\mathscr{N}^{I_5}_{12(P)}\right\| = 8$
$\left\|\mathscr{N}^{T_6}_{12(P)}\right\| = 8$	$\left\|\mathscr{N}^{T_7}_{12(P)}\right\| = 0$	$\left\|\mathscr{N}^{I_6}_{12(P)}\right\| = 4$	$\left\|\mathscr{N}^{I_7}_{12(P)}\right\| = 8$
$\left\|\mathscr{N}^{T_8}_{12(P)}\right\| = 6$	$\left\|\mathscr{N}^{T_9}_{12(P)}\right\| = 0$	$\left\|\mathscr{N}^{I_8}_{12(P)}\right\| = 4$	$\left\|\mathscr{N}^{I_9}_{12(P)}\right\| = 8$
$\left\|\mathscr{N}^{T_{10}}_{12(P)}\right\| = 2$	$\left\|\mathscr{N}^{T_{11}}_{12(P)}\right\| = 0$	$\left\|\mathscr{N}^{I_{10}}_{12(P)}\right\| = 4$	$\left\|\mathscr{N}^{I_{11}}_{12(P)}\right\| = 8$

Using Lemma 1, we are able to determine that the 72 P-combinatorial hexachords belong to eight set classes. Accordingly, Table 4 gives the numbers of P-combinatorial n-chords and their set classes for values $n \leq 12$.

Table 4. Numbers of P-combinatorial n-chords and their set classes in spaces of size $\mathbb{Z}_{2n}$, $n \leq 12$.

Space	$\mathbb{Z}_2$	$\mathbb{Z}_4$	$\mathbb{Z}_6$	$\mathbb{Z}_8$	$\mathbb{Z}_{10}$	$\mathbb{Z}_{12}$	$\mathbb{Z}_{14}$	$\mathbb{Z}_{16}$	$\mathbb{Z}_{18}$	$\mathbb{Z}_{20}$	$\mathbb{Z}_{22}$	$\mathbb{Z}_{24}$
$n =$	1	2	3	4	5	**6**	7	8	9	10	11	12
$\left| \mathscr{N}_{2n(P)} \right|$	2	6	8	22	32	**72**	128	278	512	1056	2048	4168
$\left| \mathscr{N}_{2n(P)}/H \right|$	1	2	2	4	4	**8**	10	20	30	56	94	180

3.3 R Combinatoriality

Unlike I and P combinatoriality, in which n-chords map to their complements under some member of the transposition-and-inversion group, R and RI combinatoriality result when some $h \in H$ exists that maps N to itself. Specifically, R combinatoriality occurs when h is a transposition operator. As this situation always obtains for the identity element T_0, we observe that all n-chords are trivially R-combinatorial. Instead of powers of 2—as with Eqs. 1, 2, 3, and 5—the basis for the enumeration of R-combinatorial n-chords is the binomial coefficient. The formula in Eq. 6 for determining the numbers of n-chords stabilized by even-indexed transposition operators brings the binomial coefficient together with the results of Theorem 1.

$$\left| \mathscr{N}_{2n(R)}^{T_{2x}} \right| = \binom{2n/j}{n/j}, \text{ where } j = n/\gcd(n, x) \tag{6}$$

For example, using $x = 0$ in the familiar case of $n = 6$, we find $\binom{12}{6} = 924$ R-combinatorial hexachords that are stabilized by T_0.

  The numbers of R-combinatorial n-chords that are stabilized by inversion operators derive from the binomial coefficient as well. However, unlike the formula for determining numbers of n-chords stabilized by transposition operators, the formulae in Eqs. 7 and 8 differentiate between even and odd values of n.

$$\left| \mathscr{N}_{2n(R)}^{I_{2x}} \right| = \begin{cases} \binom{n-1}{n/2} + \binom{n-1}{(n/2)-1}, & \text{if } 2 \mid n \\ 2\binom{n-1}{(n-1)/2}, & \text{if } 2 \nmid n \end{cases} \tag{7}$$

$$\left| \mathscr{N}_{2n(R)}^{I_{2x+1}} \right| = \begin{cases} \binom{n}{n/2}, & \text{if } 2 \mid n \\ 0, & \text{if } 2 \nmid n \end{cases} \tag{8}$$

An outline of a simple proof follows. Equations 7 and 8 present four cases: (1) I_{2x} with $2 \mid n$, (2) I_{2x+1} with $2 \mid n$, (3) I_{2x} with $2 \nmid n$, and (4) I_{2x+1} with $2 \nmid n$, which we take in turn.

1. For any even-indexed inversion in a space of size $2n$, where n is even, the axis of reflection runs through two fixed points: y and $y + n$. Hence, the inversionally symmetrical n-chord may exclude both these points as members, in which case there exist $n - 1$ points on either side of the axis from which to choose one half, $n/2$, of the elements of the n-chord; or the n-chord may

include both these points as members, in which case there exist $n-1$ points on either side of the axis from which to choose one less than one half, $(n/2)-1$, of the elements of the n-chord.

2. For any odd-indexed inversion in a space of size $2n$, where n is even, the axis of reflection fixes no points. Hence, there exist n points on either side of the axis from which to choose one half, $n/2$ of the elements of the inversionally symmetrical n-chord.

3. For any even-indexed inversion in a space of size $2n$, where n is odd, the axis of reflection runs through two fixed points: y and $y+n$. Hence, the inversionally symmetrical n-chord must include one or the other—but not both—of these points as members. In either of the two cases, there exist $n-1$ points on either side of the axis from which to choose one half of the remaining points, $(n-1)/2$, of the elements of the n-chord.

4. For any odd-indexed inversion in a space of size $2n$, where n is odd, the axis of reflection fixes no points. However, for the n-chord to be inversionally symmetrical, one point must be fixed. Hence, the situation fails.

Along with the Lemma 1, the above equations enable us to determine the numbers of set-classes to which the set of R-combinatorial n-chords belong. Table 5 presents this information for cases $n \leq 12$.

Table 5. Numbers of R-combinatorial n-chords and their set classes in spaces of size $\mathbb{Z}_{2n}$, $n \leq 12$.

Space	$\mathbb{Z}_2$	$\mathbb{Z}_4$	$\mathbb{Z}_6$	$\mathbb{Z}_8$	$\mathbb{Z}_{10}$	$\mathbb{Z}_{12}$	$\mathbb{Z}_{14}$	$\mathbb{Z}_{16}$	$\mathbb{Z}_{18}$	$\mathbb{Z}_{20}$	$\mathbb{Z}_{22}$	$\mathbb{Z}_{24}$
$n =$	1	2	3	4	5	**6**	7	8	9	10	11	12
$\|\mathcal{N}_{2n(R)}\|$	2	6	20	70	252	**924**	3432	12870	48620	184756	705432	2704156
$\|\mathcal{N}_{2n(R)}/H\|$	1	2	3	8	16	**50**	133	440	1387	4752	16159	56822

3.4 *RI* Combinatoriality

RI combinatoriality results when an n-chord maps onto itself under an inversion operation. As with R combinatoriality, we determine the numbers of RI-combinatorial n-chords by using the binomial coefficient. Moreover, the formula for reckoning the number of RI-combinatorial n-chords that are stabilized by even-indexed transposition operators (equivalent to the number of $2n$-bead balanced binary necklaces that are equivalent to their reverse [11]) also incorporates the μ-function, which again eliminates redundancies.

$$\left|\mathcal{N}_{2n(RI)}^{T_{2x}}\right| = \sum_{j|\gcd(x,n)} \sum_{k|j} \mu(k)wj, \tag{9}$$

$$\text{where } w = \begin{cases} \binom{j/k}{j/2k} + \binom{(j/k)-1}{j/2k} + \binom{(j/k)-1}{(j/2k)-1}, & \text{if } 2 \mid j/k \\ 2\binom{(j/k)-1}{((j/k)-1)/2}, & \text{if } 2 \nmid j/k \end{cases}$$

As was the case with I and P combinatorialities, the set of I_x-stabilized RI-combinatorial n-chords in any particular space $\mathbb{Z}_{2n}$ is the same as it is for R-combinatoriality.

Theorem 4. $\mathcal{N}^{I_x}_{2n(RI)} = \mathcal{N}^{I_x}_{2n(R)}$

Proof. We note that any n-chord is R-combinatorial. Therefore, an n-chord is RI-combinatorial, if and only if it is stabilized by I_x for some $x \in \mathbb{Z}_{2n}$.   □

Using these results, we are now ready to apply Lemma 1 to determine the number of set classes to which the members of $\mathcal{N}_{2n(RI)}$ belong. Table 6 provides sample results for $n \leq 12$.

Table 6. Numbers of RI-combinatorial n-chords and their set classes in spaces of size $\mathbb{Z}_{2n}$, $n \leq 12$.

Space	$\mathbb{Z}_2$	$\mathbb{Z}_4$	$\mathbb{Z}_6$	$\mathbb{Z}_8$	$\mathbb{Z}_{10}$	$\mathbb{Z}_{12}$	$\mathbb{Z}_{14}$	$\mathbb{Z}_{16}$	$\mathbb{Z}_{18}$	$\mathbb{Z}_{20}$	$\mathbb{Z}_{22}$	$\mathbb{Z}_{24}$
$n =$	1	2	3	4	5	**6**	7	8	9	10	11	12
$\left\|\mathcal{N}_{2n(RI)}\right\|$	2	6	8	38	52	**216**	268	1062	1232	4956	5524	21848
$\left\|\mathcal{N}_{2n(RI)}/H\right\|$	1	2	2	6	6	**20**	20	70	70	252	252	924

4   Conclusions

In this study, we have examined combinatorial n-chords using techniques from the mathematical fields of combinatorics, number theory, and group theory. Specifically, we have enumerated the sets of P-, I-, R-, and RI-combinatorial n-chords and their set classes. In the process, our results reveal further aspects of combinatorial sets. For instance, we note that the number of T_{2x}-symmetric combinatorial n-chords in a space of size $2n$ is equivalent to the total number of combinatorial n-chords in a space of size $2\gcd(n,x)$ (Theorem 1). No combinatorial n-chords are held invariant by a translation operation with an odd index (Theorem 2). The set of I-invariant n-chords that are P-combinatorial is equivalent to the set of those that are I-combinatorial (Theorem 3), and this set is precisely the set of all-combinatorial n-chords (Corollary 1). Similarly, the set of I-invariant n-chords that are R-combinatorial is equivalent to the set of those that are RI-combinatorial (Theorem 4).

Several avenues exist for future work on combinatorial n-chords and their spaces. Whereas this study is limited to aggregates formed from unions of two n-chords, its methodology could be extended to study aggregate formation that results from unions of $m > 2$ P-, I-, R-, and RI-combinatorial n-chords. Further, we can study sets of combinatorial n-chords from other mathematical perspectives that have yielded significant music-theoretical results, such as the Discrete Fourier Transform or algebraic topology and geometry. Such investigations will continue to shed new light on these intriguing structures.

References

1. Babbitt, M.: Set structure as a compositional determinant. In: Peles, S., Dembski, S., Mead, A., Straus, J.N. (eds.) The Collected Essays of Milton Babbitt, pp. 86–108. Princeton University Press, Princeton (2003)
2. Forte, A.: The Structure of Atonal Music. Yale University Press, New Haven (1973)
3. Fripertinger, H.: Enumeration and construction in music theory. In: Feichtinger, H.G., Dörfler, M. (eds.) Diderot Forum on Mathematics and Music: Computational and Mathematical Methods in Music, pp. 179–204. Österreichische Computergesellschaft, Vienna (1999)
4. Fripertinger, H., Lackner, P.: Tone rows and tropes. J. Math. Music **9**(2), 111–172 (2015)
5. Hauer, J.M.: Zur Lehre vom atonalen Melos. Das Kunstblatt **8**(12), 353–360 (1924)
6. Hook, J.: Why Are there twenty-nine tetrachords? A tutorial on combinatorics and enumeration in music theory. Music Theory Online **13**(4) (2007) https://www.mtosmt.org/issues/mto.07.13.4/mto.07.13.4.hook.html
7. Nolan, C.: Combinatorial space in nineteenth- and early twentieth-century music theory. Music Theory Spectr. **25**(2), 205–41 (2003)
8. OEIS Foundation Inc.: Number of $2n$-bead balanced binary strings, rotationally equivalent to complement, Entry A045654 in The On-Line Encyclopedia of Integer Sequences (2022). https://oeis.org/A045654. Accessed 11 Mar 2022
9. OEIS Foundation Inc.: Number of $2n$-bead balanced binary strings, rotationally equivalent to reversed complement, Entry A045655 in The On-Line Encyclopedia of Integer Sequences (2022). https://oeis.org/A045655. Accessed 11 Mar 2022
10. OEIS Foundation Inc.: Number of $2n$-bead balanced binary necklaces which are equivalent to their reverse, complement and reversed complement, Entry A045674 in The On-Line Encyclopedia of Integer Sequences (2022). https://oeis.org/A045674. Accessed 11 Mar 2022
11. OEIS Foundation Inc.: Central binomial coefficients $C(2n, n)$ repeated, Entry A128014 in The On-Line Encyclopedia of Integer Sequences (2022). https://oeis.org/A128014. Accessed 11 Mar 2022
12. Peck, R.: Generalized commuting groups. J. Music Theory **54**(2), 143–177 (2010)
13. Schoenberg, A.: Composition with twelve tones. In: Newkin, D. (ed.) Style and Idea, pp. 102–143. Philosophical Library Inc., New York (1950)
14. Starr, D., Morris, R.: A general theory of combinatoriality and the aggregate (part 1). Perspect. New Music **16**(1), 3–35 (1977)
15. Starr, D., Morris, R.: A general theory of combinatoriality and the aggregate (part 2). Perspect. New Music **16**(2), 50–84 (1978)
16. Straus, J.N.: Introduction to Post-tonal Theory. W. W. Norton & Company Inc., New York (2016)
17. Wilcox, H.J.: Group tables and the generalized hexachord theorem. Perspect. New Music **21**(1/2), 535–539 (1982/1983)

Euler's "Tentamen": Historical and Mathematical Aspects on the Consonance Theory

Sonia Cannas[1,2(✉)] and Maria Polo[3,4]

[1] Università di Cagliari, Cagliari, Italy
`sonia.cannas@unica.it`
[2] Liceo Classico-Scientifico "Euclide", Cagliari, Italy
[3] Dipartimento di Matematica e Informatica, Università di Cagliari, Cagliari, Italy
`mpolo@unica.it`
[4] CRSEM, Università di Cagliari, Cagliari, Italy

Abstract. The *Tentamen novae theoriae musicae* is a treatise in which Euler elaborated a new music theory using mathematics. The aim of this paper is to explain his theoretical system to justify the pleasure of listening to music and to analyze differences and similarities with other consonance theories.

Keywords: Euler · Tentamen · Consonance theory

1 Introduction

Since the XVII century, important mathematicians as Euler (1707–1783) and d'Alembert (1717–1783) worked on music theories. This is obviously not the result of a "historical chance". On the contrary, it represents prolongation of a tradition in which mathematics helps to describe music from an acoustical and theoretical point of view. Probably the first ancient example is the Pythagorean scale, defined by Pythagoreans starting from the first four natural numbers and the observation of the sounds produced by the division of the string of a monochord, determining the ratios of the consonant intervals of octave, fifth and fourth. Didymus and Ptolemy developed the same idea to describe intervals as mathematical ratios defining a new musical scale, and Zarlino (1517–1590) resumed it in *Le istitutioni harmoniche* [22]. They did not know, but this idea is totally agree with harmonic series.

After composers as Bach (1685–1750), Handel (1685–1759), Rameau (1683–1764), Haydn (1732–1809), Mozart (1756–1791), music had a profound change by abandoning the medieval counterpoint in favor of a new harmony. This change had to be explained, the western music needed a new theory, and it inspired scientist with a passion for music. Several scientists dedicated workd on it: Descartes (1596–1650) with *Compendium musicae* [9], Mersenne (1588–1648) with *Harmonie universelle* [17] and Leibniz [16] in several letters.

Euler wrote severel works on music [12], whose first and best known is *Tentamen novae theoriae musicae ex certissismis harmoniae principiis dilucide expositae* [11], written in Latin in 1731 and published in 1739. The *Tentamen novae theoriae musicae* is a treatise in which Euler elaborated a new music theory using mathematics. In recent years, interest on this work by Euler has been increasing [2,3,8,13]. Probably this interest also arises from the important developments of the Mathematical Music Theory of the last thirty years linked to the Tonnetz and his generalizations [1,4,6,7,10,20]. The Tonnetz is a 2-dimensional simplicial complex which tiles the Euclidean plane with triangles representing major and minor triads. This structure today is interesting because is a model for the neo-Riemannian operations P, L and R, from which generalizations have been introduced in Mathematical Music Theory [5,15]. A graph similar to the Tonnetz appeared in a the chapter of *Tentamen* in order to represent some intervals of the just intonation in a scheme.

In this paper, we will focus on Euler's theoretical system to justify the pleasure of listening to music. More precisely, we will analyze differences and similarities with other consonance theories. We will start, in Sect. 2, with a historical-musical context in which Euler lived and his possible musical knowledges. In Sects. 3 and 4 we summarize and explained the mathematical ideas on sound and on the pleasure of consonance, described in the first four chapters of the *Tentamen*. Finally, in the last section we will analyze his consonance theory by comparing it with other theories.

2   Some Historical Aspects on Euler's Musical Interests and the Birth of the "Tentamen"

In agreement with De Piero [8], we note that Euler's interests on music began very early.

The *Dissertatio physica de sono*, written in 1727 to compete for the chair of Physics in Basel, is a demonstration of this. In this dissertation the physical foundations of music are provided; these concepts were taken up and expanded in the first chapter of *Tentamen*. The *Dissertatio* did not allow Euler to obtain the chair in Basel, but showed the mathematician's interest in the subject and laid the foundations for the next and more important essay, already completed in 1731, and which originally bore the name of *Tractatus de musica*. De Piero points out that the news of the writing of this essay is present in a letter that Euler sent to his teacher Johann Bernoulli on May 25, 1731 from St Petersburg, where he moved in 1727. In this letter the mathematician showed that he had designed the entire work and that he wrote most of it, moreover he has not yet proposed a title, but speaks indefinitely of *Systema Musicus*. But, in the reply sent on 11 August 1731, his teacher introduced for the first time the term *Tractatus Musices*, a term that led Euler to encode the first title in *Tractatus de Musica*, as we learn from the letter sent to Bernoulli on December 20, 1738. According to De Piero, we believe the title was established before 1738, but we have the first evidence of it only on that date. However, the work was published

only in 1739 in St Petersburg, printed by the Academy of Sciences with the new title *Tentamen novae theoriae musicae ex certissimis harmoniae principisis dilucidae expositae*.

Actually, the idea of writing an essay on music dates back to the period spent in Basel, this can be seen from Euler's first notebook, which the Swedish mathematician Gustav Enestroem dated to 1726[1], therefore one year before the publication of the *Dissertatio*. From the notes in the notebook the work, in its first formulation, should have been titled *Musices Theoreticae Systema*, and should have been divided into three sections: *De compositiones solius discantus*, *De compositione integrorum concertorum*, *De compositione certarum specierum*.

The real reason why Euler gave up on this project, completing another completely different one, is not known. According to Ferdinand Rudio[2] the transfer of the mathematician to St Petersburg and other kind of works would have forced him to postpone his musical projects. However, this does not convince De Piero, because in 1731 Euler already spoke to Bernoulli about his projects, claiming that he had already conceived the overall plan of the work and that he had already drawn up most of it. De Piero hypothesizes that at first Euler started from *Dissertatio physica de sono* and, once the conceptual foundations of a physical nature had been laid, he was so passionate about hypothesizing the drafting of an essay expressly dedicated to composition. However, he also hypothesizes that Euler may have known that Johann Mattheson was dedicating himself to a similar essay in the same years[3]. However, the problem remains open.

Although Euler wrote about music theory, there are no particular testimonies of meetings or discussions with musicians. But, according to De Piero [8], during the period at the Frederick the Great's court he probably met musicians and composers such as Carl Philipp Emanuel Bach[4], Johann Joachim Quantz[5] or the brothers Carl Heinrich e Johann Gottlieb Graun[6]. Moreover, in 1747 Johann Sebastian Bach[7] went to the Frederick the Great's court, who asked him to compose music on the basis of a theme composed by himself, which will then be collected in the famous *Musical offer*. Also in St Petersburg many musicians as Baldassare Galuppi[8] stayed during the reign of Catherine II. According to De Piero, although there is no direct evidence, a mathematician as interested

---

[1] G. Ernstroem, *Bericht an die Eulerkommission der Schweizerschen naturforschenden Gesellshaft iber die Eulerchen Manuskripte der Petersburg Akademie*, in "Jahresbericht der Deutchen Mathematiker-Vereinigung, 22", 1913, p. 197.

[2] See [8] p. 7.

[3] This is the essay *Grosse General-Bass-Schuule*, published in 1731 in Hamburg and also cited in *Tentamen*.

[4] Carl Philipp Emanuel Bach (1714–1788), composer, p. 45 [21].

[5] Johann Joachim Quantz (1697–1773), composer and music theorist, p. 715 [21].

[6] Carl Heinreich Graun (1701–1759), Johann Gottlieb Graun (1702 o 1703–1771), composers, p. 360 [21].

[7] Johann Sebastian Bach (1685–1750), composer, pa. 46 [21].

[8] Baldassarre Galuppi (1706–1785), known as Buranello, Master of the Ducal Chapel of San Marco in Venice, stayed there from 1756 to 1768, p. 333 [21].

in music as Euler and employed in the same court may hardly remain outside of all this. The only evidence of disquisitions with musicians of the time are correspondence exchanges. To date, four letters are known: two between Euler and Giuseppe Tartini[9] and two more between the Swiss mathematician and Jean-Philippe Rameau[10].

Reading the *Tentamen*, however, it is evident that Euler knew very well the musical theories of that time and of the past. In fact, the conceptual structure of the work starts from the Pythagorean principles of harmony, influenced by the writings of Marin Mersenne, René Descartes and Gottfried Wilhelm von Leibniz. The latter, as librarian and historian of the court of Hannover, provided consultancy for the staging of the court shows, consequently maintained daily relationships with musicians and singers and this allowed him to develop very original and interesting ideas about music. These ideas were not collected in a specific text, but can be traced in letters sent to mathematicians and theorists. Euler quoted Leibniz[11] as evidence of that any piece of music is composed of the exponents of only the numbers $2, 3$ and 5 since only the numerical ratios based on such numbers or on the respective multiples produce listening pleasure, while too complex numerical relationships cannot be perceived as pleasure.

Although he is never mentioned in the *Tentamen*, Descartes is the author who most left his mark froma methodological point of view. There are several points in common between Euler's essay and that of the French mathematician, *Compendium Musicae*, published in Utrecht in 1650. In fact, the Swiss mathematician took up the concept according to which the causes of *delectare et movere affectus* are to be found in the relationship of duration or time between sounds and in the relationship between high and low pitches[12], and according to both the perception of pleasure by the hearing is governed by simple arithmetic ratios, because they are more easily understood.

3 Sound and Hearing

The first principle from which Euler starts, mainly presented in the first chapter of the *Tentamen*, is about physical nature and concerns the science of sounds[13]. On the other hand this aspect was not addressed by Descartes and, according to De Piero, this constitutes the most original contribution provided by Euler. For the Swiss mathematician, sounds are vibrations of the air perceived as multiple

---

[9] Giuseppe Tartini (1692–1770), composer, violinist and music theorist, p. 880 [21].

[10] Jean-Philippe Rameau (1683–1764), composer and music theorist, p. 725 [21].

[11] L. Euler, *Tentamen*, chap. X, par.19.

[12] R. Descartes, emph Compendium Musicae, Utrecht, 1650, ed. mod. *Abregé de musique*, Édition nouvelle, in traduction, presentation and notes by Fr. de Buzon, Paris, Presses Universitaires de France, 1987. Cf. p. 55: "Media ad finem, vel soni affectiones duae sunt praecipue: nempe huius differentiae, in ratione durationis vel temporis, et in ratione intensionis circa acutum aut grave".

[13] We observe that throughout the work, Euler speaks of *sound* but never of *sound waves*.

by the ears[14]. Given two sounds, we understand the relation between them by the ratio of the number of vibrations carried out for one, with the number of vibrations carried out for the other in the same time. For example, if there were 3 vibrations for the first, while for the second there would be 2 of them, we know their relation and consequently their order, by observing the ratio of the numbers 3 and 2 which is 3:2. Then, we observe that Euler obtained the same ratio introduced in Pythagorean scale and in just intonation for describing the interval of fifth.

The second principle, introduced in the preface and exposed above all in the second chapter, invests in the causes of pleasure, questioning why a person likes or dislikes music. In the preface to the Tentamen, Euler argues that consonances do not depend on human habits. In support of this, he quotes Pythagoras, who identified the cause of the pleasure produced by the consonances in the mathematical relationships of the intervals, although he did not understand how these relationships are perceived by hearing. Euler notes that European music is not appreciated by barbarians[15] and vice versa. The Basel mathematician attributes this to the complexity of European musical composition, made up of various melodic lines that intertwine with each other and which are not easily distinguishable and noticeable by few trained ears. Again, he observes that in many countries it is commonly believed that the octave, fifths, fourths, thirds and sixths are consonant intervals, while tritons, sevenths, seconds and all the others that can be constituted are dissonant. Euler's aim is to investigate the causes of this judgment which he considers universal. Euler therefore proposes a classification of the degrees of pleasure of consonances based on mathematical relationships.

4   Pleasure and Consonance

Music is formed by sounds played together that Euler defines consonance. We observe that it does not distinguish between consonance and dissonance, as historically it has always been used and as it is still usual today, every compound sound is consonant and may or may not be liked. According to Euler, pleasure consists in the exact perceptibility of sounds and their relationships: two or more sounds like when one perceives the relationship that the numbers of the vibrations emitted have between them, vice versa they are dislike when no order is felt. We perceive pleasure, if from that structure we understand how all the parts intertwine with each other, and how their actions all converge. For Euler where there is order, there is perfection, and that the rule or law of the order corresponds to the goal which marks the perfection. The order of sounds primarily consists on two types: according to the pitch and according to the duration.

As already mentioned, Euler defines sound as successive beats produced in the air in a certain order, so a sound is distinctly perceived if all the beats from

---

[14] This is clearly the frequency of the sound wave.

[15] De Piero hypothesizes that what Euler defines *barbarians* are populations living outside Europe.

the hearing organs are heard and their order is recognized. In general, perception of order can occur in two ways:

1. when the rule is known;
2. when the rule is not known, but can be deduced by looking the structure.

In music the order is perceived according to this latter way, in fact it is through listening that the order that sounds have among themselves is understood. Since the perception of perfection produces pleasure, according to Euler a composition pleases if the order of the sounds is perceived. It may happen that some people perceive this order that others do not hear, which is why the same music can like some and not others.

4.1   Study of Chords with Two Sounds

Given two sounds, their relationship is perceived through the ratios that the number of hits, emitted at the same time, have between them. From this, Euler defines the degrees of pleasure.

The first and simplest degree of pleasure is the unison, represented by the numerical ratio 1 : 1. On the other hand, two sounds having a double ratio, therefore 1 : 2, are part of the second degree of pleasure. The sounds expressed with the numerical ratio 1 : 3 and the ratio 1 : 4 belong to the third degree. In fact, the first is expressed by small numbers, so it is easily perceptible, the second would apparently seem a more complex ratio, but it is obtained simply by dividing the ratio 1 : 2 by 2, therefore it is not very difficult to distinguish from the latter. For this reason Euler believes that both relationships are part of the third degree of pleasure. Similarly, the ratios 1 : 8, 1 : 16 belong to the fourth and fifth degree of pleasure, respectively. More generally, the ratio $1 : 2^n, n \in \mathbb{N}$, corresponds to the degree of pleasure $n + 1$.

If the ratio contains divisors different by 1 and 2, the degree of pleasure is less. But the degree of pleasure is also estimated by looking at the magnitude of the numbers: the ratio 1 : 5 is simpler than 1 : 7, although the latter is no simpler than 1 : 8. For the ratios 1 : p, where p is a prime number, is easy to determine the degree of pleasure. In fact since 1 : 2 belongs to the second degree and 1 : 3 to the third one, then 1 : 5 belongs to the fifth and 1 : 7 to the seventh. More generally, if p is prime, the ratio 1 : p belongs to the degree of pleasure p.

It follows that if the ratio 1 : p refers to the degree m, the ratio 1 : $2p$ belongs to the degree $m + 1$. In fact, multiplying the number p by 2, the perception of the ratio requires the perception of 1 : p and the division by 2, an operation that increases the degree of pleasure by one unit. Similarly 1 : $4p$ belongs to the degree $m + 2$. More generally, $1 : (2^n p)$ belongs to the degree $m + n$. Similarly, the degree of pleasure of the relationship 1 : (pq), with p and q prime numbers, is equal to $p + q - 1$, since 1 : (pq) is composed by 1 : p and 1 : q. This also applies to 1 : (PQ) with any positive integer P and Q. Similarly, the ratio 1 : (pqr), with p, q, r prime, being constituted by 1 : (pq) and 1 : r whose degrees of pleasure are

$p + q - 1$ and r respectively, will have degree of pleasure $p + q + r - 2$. Iterating the reasoning, the degree of the relationship $1 : (pqrs)$ will be $p + q + r + s - 3$, and so on.

So, if p is prime, the degree of pleasure of $1 : p^2$ is $2p - 1$, and for $1 : p^3$ is $3p - 2$. More generally $1 : p^n$ belongs to the degree of pleasure $np - n + 1$. Therefore, since $1 : q^m$ belongs to the degree $mq - m + 1$, the ratio $1 : (p^n q^m)$ will belong to the degree

$$np + mq - n - m + 1 \tag{1}$$

So for any number P, in order to determine the degree of pleasure of a ratio $1 : P$ we have to represent the ratio in simple factors and the degree will be obtained by subtracting from their sum the number of factors subtracted by one. To clarify, Euler also shows the following example.

Example 1. *We look for the degree of the ratio $1 : 72$.*
First we factorize 72, then $72 = 2^3 \cdot 3^2$. Using the 1, we have $3 \cdot 2 + 2 \cdot 3 + 2 - 3 + 1 = 8$.
Therefore the degree of pleasure of the ratio $1 : 72$ is 8.

In [12] it is generalized it as the following general formula that Euler does not write. Given two sounds such that their ratio is $\dfrac{1}{P}$, we factorize P as

$$P = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_n^{\alpha_n}.$$

Then, the degree $d(P)$ of pleasure of $\dfrac{1}{P}$ is

$$d(P) = \sum_{i=1}^{n} (\alpha_i p_i - \alpha_i) + 1 \tag{2}$$

4.2   Study of Chords with More Than Two Sounds

At this point, Euler continues his theory by examining relationships between more than two numbers, that is, he analyzes the degree of pleasure obtained by more than two sounds. Given two prime numbers p, q, in the ratio of three numbers such as $1 : p : q$ we also perceive $1 : p$ and $1 : q$, the perception here is equal to that of $1 : (pq)$. Similarly, the ratio of four numbers $1 : p : q : r$, where p, q, r are always prime numbers, belongs to the same degree as $1 : (pqr)$. Euler proposes another ex:

Example 2. *Suppose we have 4 sounds expressed by the following numbers: $1 : 2 : 3 : 5$. The degree of pleasure of these sounds is the same as those expressed by the ratio $1 : 30$, therefore: $2 + 3 + 5 - 3 + 1 = 8$.*
Therefore they belong to the eighth degree of pleasure.

DE CONSONANTIIS.   6 ß

Gr. II.	2:5.	Gr. IIX.	3:7.	3:64.	1:160.
I:2.	I:18.	I:14.	1:25.	I:256.	5:32.
Gr. III.	2:9.	2:7.	1:28.	Gr. X.	I:162.
I:3.	I:24.	I:30.	4:7.	I:42.	2:81.
I:4.	3:8.	2:15.	I:45.	3:14.	I:256.
Gr. IV.	I:32.	3:10.	5:9.	6:7.	8:27.
I:6.	Gr. VII.	5:6.	I:60.	I:50.	I:288.
2:3.	I:7.	I:40.	3:20.	2:25.	9:32.
I:8.	I:15.	5:8.	4:15.	I:56.	I:384.
Gr. V.	3:5.	I:54.	5:12.	7:8.	3:128.
I:5.	I:20.	2:27.	I:80.	I:90.	I:512.
I:9.	4:5.	I:72.	5:16.	2:45.	
I:12.	I:27.	8:9	I:81.	5:18.	
3:4.	I:36.	I:96.	I:108.	9:10.	
I:16.	4:9.	3:32.	4:27.	I:120.	
Gr. VI.	I:48.	I:128.	I:144.	3:40.	
I:10.	3:16.	Gr. IX.	9:16.	5:24.	
	I:64.	I:21.	I:192.	8:15.	

Fig. 1. The first ten degree of pleasure in *Tentamen*, p. 61.

The Swiss mathematician observes that these prime numbers must be all unequal. In fact, $1 : p : p$ is perceived exactly as $1 : p$. Similarly, to perceive the relationship $1 : (pr) : (qr) : (ps)$, only the ratios $1 : p, 1 : q, 1 : r, 1 : s$ are needed, it is not necessary to count twice $1 : p$ and $1 : r$. Therefore, the degree of pleasure is the same as the ratio $1 : pqrs$.

At this point Euler determine the universal rule in order to know the degree of pleasure in perceiving the ratio of several numbers proposed at the same time: the least common multiple is determined and the 1 is used. Thus, Euler shows in a table to which degrees correspond all the least common multiples.

Example 3. *Let the numbers* $72, 80, 100, 112$ *be consider. Their factorizations are respectively:* $2^3 \cdot 3^2, 2^4 \cdot 5, 2^2 \cdot 5^2, 2^4 \cdot 7$. *So, the least common multiple is* $2^4 \cdot 3^2 \cdot 5^2 \cdot 7 = 25200$, *and it belongs to the twenty-third degree.*

5   Comparison Between *Tentamen* and Other Mathematical Consonance Theories

As already mentioned, studies on consonance and dissonance has been central to music theory since ancient Greece.

But what does it mean consonance and dissonance? Several music theorists have been investigating to try and answer this question, such as Hindemith [14] and Tenney [19]. We can observe that there are semantic problems on the terms consonance and dissonance: we do not have a strict and universal definition of them. During the history many definitions and different meanings are found. For instance, only focusing on western music tradition, we observe that major

and minor thirds and sixths were considered dissonant in antiquity but, in 14th century, they were accepted as consonant. For Tenney [19], there are five different forms of the consonance-dissonance conception (CDC).

> "Before the rise of polyphonic practice they were used in an essentially melodic sense, to distinguish degrees of affinity, agreement, similarity, or relatedness between pitches sounding successively. During the first four centuries of the development of polyphony they were used to describe an aspect of the sonorous character of simultaneous dyads, relatively independent of any musical context in which they might occur. In the 14th century the CDC began to change (again) in conjunction with the newly developing rules of counterpoint, and a new system of interval-classification emerged which involved the perceptual clarity of the lower voice in a polyphonic texture) and of the text which it carried). In the early 18th century, 'consonance' and 'dissonance' came to be applied to individual tones in a chord, giving rise to a new interpretation of these terms which would eventually yield results in diametric opposition to all of the earlier forms of the CDC. Finally-in the mid-19th century-a conception f consonance and dissonance arose in which 'dissonance' was equated with "roughness", and this had implications quite different from those of earlier forms of the CDC."

Without a clear and precise definition of consonance it is difficult to developed a theory able to explain the nature of consonance and dissonance in musical perception, because it is not clear what this theory would have to explain. From this point of view, Euler's idea of not distinguishing consonance and dissonance but different levels of consonance may be interesting.

Moreover, consonance is perceived in different way in different cultures. For all these reasons, we may say that concept of consonance is not universal.

In addition to this, we know that in a musical piece the perception of consonance do not only depends from the individual chords, but also for the chord sequences. In other terms: the concept of consonance is both local and global. This problem was known also by Euler indeed, in chapter V, he studied also the succession of consonance. But he considered the preeminence of pitch over that of the duration, since that one is measured by the frequencies of vibration, Euler brings back the evaluation of the musical pleasure to the arithmetic measurement of the proportions related to the sounds. Therefore, he brings back musical science using the theory of proportions already used in the Ancient Greece with Pythagoreans and Ptolomy or by Zarlino. Despite the same starting point, there are differences stressing the originality of Euler. Initially, the theory of the basel mathematician exceeds by far the simple consideration of the ratios of the frequencies of two sounds. Then, contrary to his predecessors, who had also founded their theory on the proportions without any explanation, Euler introduces a philosophical argumentation, in which by the proportions one arrives at the musical pleasure, via the order and the perfection. As for the Leibniz source, one can undoubtedly recognize from it the influence in the distinction which Euler makes between the two modes of perception of the order.

Moreover, as we see in Fig. 1, we note that the ratios representing intervals in just intonation are all included also in Euler's theory. This is an interesting point, because the ratios of just intonation are totally agree with the physical theory of natural harmonics and harmonic series. We note that Euler's Fig. 1 does not include the ratios 15 : 16 and 32 : 45. This is why his table represents only the first ten degree of pleasure. It is not specified in the *Tentamen*, but given a ratio $P : Q$, where $P, Q \in \mathbb{N}$, $P = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_n^{\alpha_n}$, and $Q = q_1^{\beta_1} \cdot p_2^{\beta_2} \cdots q_m^{\beta_m}$, we can determine the degree of pleasure d of $\dfrac{P}{Q}$ with the following formula

$$d = \sum_{i=1}^{n} (\alpha_i p_i - \alpha_i) + \sum_{j=1}^{m} (\beta_j q_j - \beta_j) + 1 \tag{3}$$

Using the formula 3 we have that the degree of pleasure of 15 : 16 is XI, and that one of 32 : 45 is XIV.

Table 1. Relations between intervals in just intonation and Euler's degree of consonance

Interval	Ratio	Degree
Unison	1 : 1	I
Octave	1 : 2	II
Perfect fifth	2 : 3	IV
Perfect fourth	3 : 4	V
Major sixth	3 : 5	VII
Major third	4 : 5	VII
Minor third	5 : 6	VIII
Minor sixth	5 : 8	VIII
Major second	8 : 9	VIII
Minor seventh	9 : 16	IX
Major seventh	8 : 15	X
Minor second	15 : 16	XI
Tritone	32 : 45	XIV

Despite Euler's theory include the interval ratios of just intonation, we note some limits in his classification summarized in Table 1. In fact, there are ratios in the same degree that do not correspond to the same level of consonance, neither in his time nor in our day. For instance, in the VIII degree we found the ratios 5 : 6 (minor third), 5 : 8 (minor sixths) and 8 : 9 (major second), but the first two are imperfect consonance, while the last one is a dissonance.

Conclusions and Future Works

The concept of consonance is not universal in space and time, therefore every intention of analysis of the first four chapters of Euler's *Tentamen* is limited and incomplete. In this work, our aim has been to analyze the first 4 chapters of the *Tentamen*, limiting the concept of consonance as the pleasure of listening to a single chord. We have found that Euler's classification include the ratios of just intonation, but in the same degree of pleasure there are ratios representing intervals that do not have the same level of consonance. Since consonance also depends on the succession of chords, it will be interesting analyze the chapters of *Tentamen* involved on it.

References

1. Bigo, L., Andreatta, M., Giavitto, J.-L., Michel, O., Spicher, A.: Computation and visualization of musical structures in chord-based simplicial complexes. In: Yust, J., Wild, J., Burgoyne, J.A. (eds.) MCM 2013. LNCS (LNAI), vol. 7937, pp. 38–51. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39357-0_3
2. Caddeo, R., Hascher, X., Jehel, P., Papadopoulos, A., Papadopoulos, H.: Leonhard Euler-Écrits sur la musique. Hermann-Collection GREAM, Paris (2015)
3. Calinger, R.: Leonhard Euler: The First St. Petersburg Years (1727–1741). Historia Mathematica, vol. 23 (1996)
4. Cannas, S.: Geometric Representation and Algebraic Formalization of Musical Structures Ph.D. dissertation, Université de Strasbourg and Università degli Studi di Pavia e di Milano-Bicocca (2018)
5. Cannas, S., Antonini, S., Pernazza, L.: On the group of transformations of classical types of seventh chords. In: Agustín-Aquino, O.A., Lluis-Puebla, E., Montiel, M. (eds.) MCM 2017. LNCS (LNAI), vol. 10527, pp. 13–25. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-71827-9_2
6. Cannas, S., Andreatta, M.: A generalized dual of the Tonnetz for seventh chords: mathematical, computational and compositional aspects. In: Bridges Conference Proceedings, Stockholm (2018)
7. Catanzaro, M.J.: Generalized Tonnetze. J. Math. Music **5**(2), 117–139 (2011)
8. De Piero, A.: Il Tentamen novae theoriae musicae di Lehonard Euler (Pietroburgo 1739): traduzione e introduzione. Memorie della Accademia delle Scienze di Torino: Serie V, vol. 34 (2010)
9. Descartes, R.: Compendium musicae (1618)
10. Douthett, J., Steinbach, P.: Parsimonious graphs: a study in parsimony, contextual transformation, and modes of limited transposition. J. Music Theory **42**(2), 241–263 (1998)
11. Euler, L.: Tentamen novae theoriae musicae ex certissismis harmoniae principiis dilucide expositae. Opera Omnia, Series 3, vol. 1 (1739)
12. Hascher, X., Papadopoulos, A.: Leonhard Euler, mathématicien, physicien et théoricien de la musique. CNRS Édition, Paris (2015)
13. Hellegouarch, Y.: L'"Essai d'une nouvelle théorie de la musique" de Leonhard Euler, publication of l'IREM Paris Nord (1986)
14. Hindemith, P.: Craft of Musical Composition. Schott (1942)
15. Hook, J: Uniform triadic transformations. J. Music Theory **46**(1–2) (2002)

16. Leibniz, G.: Viri illustris Godefridi Guilielmi Leibnitii: epistolae ad diversos. Ed. Christian Kortholt, vol. 4, leipzig (1734–1742)
17. Mersenne, M.: Harmonie universelle (1636)
18. Parncutt, R., Hair, G.: Consonance and dissonance in music theory and psychology: disentangling dissonant dichotomies. J. Interdiscip. Music Stud. **5**(2), 119–166 (2011)
19. Tenney, J.: A History of "Consonance" and "Dissonance". Excelsior Music Publishing Company, New York (1988)
20. Tymoczko, D.: The generalized Tonnetz. J. Music Theory **56**(1), 1–52 (2012)
21. Vv. Aa.: Enciclopedia della musica. Le Garzantine, Garzanti libri (1999)
22. Zarlino, G.: Le istitutioni harmoniche. Francesco Senese, Venezia (1558)

Categorical and Algebraic Approaches to Music

A Projection-Oriented Mathematical Model for Second-Species Counterpoint

Octavio A. Agustín-Aquino[1(✉)] and Guerino Mazzola[2]

[1] Instituto de Física y Matemáticas, Universidad Tecnológica de la Mixteca,
Huajuapan de León, Oaxaca, Mexico
`octavioalberto@mixteco.utm.mx`
[2] School of Music, University of Minnesota, Minneapolis, MN, USA
`mazzola@umn.edu`

Abstract. Drawing inspiration from both the classical Guerino Mazzola's symmetry-based model for first-species counterpoint (one note against one note) and Johann Joseph Fux's *Gradus ad Parnassum*, we propose an extension for second-species (two notes against one note).

Keywords: Second-species · Counterpoint

1 Introduction

Guerino Mazzola's counterpoint model, founded on the concepts of

1. *strong dichotomy*, which encodes the notion of consonance and dissonance, and
2. *counterpoint symmetry*, which is the carrier of contrapuntal tension and allows to deduce the rules of counterpoint,

has been successful in explaining the necessity of regarding the fourth as a dissonance and obtaining the general prohibition of parallel fifths and tritone skips as a theorem. It also allows to define new understandings of consonance and dissonance, thereby leading to the concept of *counterpoint world*, i.e., paradigms for the handling of two-voice compositions represented as digraphs, whose vertices are consonant intervals and an arrow connects two of them whenever we have a valid progression. This, in turn, allows us to *morph* one world into another. See the monograph [2] and the treatise [4, Part VII] for a thorough account.

Despite these accomplishments, Mazzola's model is restricted to the case of *first-species* counterpoint, which means that only one note can be placed against another. Hence, in order to increase the potential of Mazzola's model for analysis and composition, it is indispensable to extend it to *second-species* counterpoint (i.e., two notes against one) and further. Our approach for a first step in this direction is to extend the notion of counterpoint interval to a 2-interval, i.e., one

such that two intervals are attached to a cantus firmus, the first one coming in the downbeat and the second one in the upbeat.

For our extension, the main idea is that the counterpoint symmetries in this case do not determine another 2-interval successor, but a first-species interval in the downbeat. The idea behind this is to blend the species of counterpoint more easily.

2 General Overview of Mazzola's Counterpoint Model

Here we quickly survey the key aspects of Mazzola's counterpoint model (we refer the reader to [2] and [4, Part VII] for a complete account). We consider the action of the group

$$\overrightarrow{GL}(\mathbb{Z}_{2k}) := \mathbb{Z}_{2k} \rtimes \mathbb{Z}_{2k}^{\times}$$

(which we call the group of *general affine symmetries*) on $\mathbb{Z}_{2k}$, which can be described in the following manner:

$$T^u.v(x) = vx + u;$$

here T^u is the *transposition* by u, and v is the *linear part* of the transformation.

We know [1,2] that, for any $k > 4$, there is at least one dichotomy $\Delta = (X/Y)$ of $\mathbb{Z}_{2k}$ such that there is a unique $p \in \overrightarrow{GL}(\mathbb{Z}_{2k})$ and

$$p(X) = Y \quad \text{and} \quad p \circ p = \mathrm{id}_{\mathbb{Z}_{2k}},$$

which is called the *polarity* of the dichotomy. The dichotomies with this property are called *strong*, and represent the division of intervals into generalized *consonances* X and *dissonances* Y.

Next we consider the *dual numbers*

$$\mathbb{Z}_{2k}[\epsilon] = \frac{\mathbb{Z}_{2k}[\mathcal{X}]}{\langle \mathcal{X}^2 \rangle} = \{x + \epsilon.y : x, y \in \mathbb{Z}_{2k}, \epsilon^2 = 0\}$$

in order to attach to each *cantus firmus* x the interval y that separates it from its *discantus*[1]. Thus for a strong dichotomy $\Delta = (X/Y)$ we have the consonant intervals

$$X[\epsilon] := \{c + \epsilon.x : c \in \mathbb{Z}_{2k}, x \in X\}$$

and the dissonant intervals $Y[\epsilon] = \mathbb{Z}_{2k} \setminus X[\epsilon]$. Considering the group

$$\overrightarrow{GL}(\mathbb{Z}_{2k}[\epsilon]) := \{T^{a+\epsilon.b}.(v + \epsilon.w) : a, b, w \in \mathbb{Z}_{2k}, v \in \mathbb{Z}_{2k}^{\times}\},$$

there is a canonical *autocomplementary* symmetry $p_{\Delta}^c \in \overrightarrow{GL}(\mathbb{Z}_{2k}[\epsilon])$ such that

$$p_{\Delta}^c(X[\epsilon]) = Y[\epsilon], \quad p_{\Delta}^c \circ p_{\Delta}^c = \mathrm{id}_{\mathbb{Z}_{2k}[\epsilon]},$$

and leaves the *tangent space* $c + \epsilon.\mathbb{Z}_{2k}$ invariant.

---

[1] The discantus can be understood in the *sweeping* $(x + y)$ or the *hanging* $(x - y)$ orientations, but we will only use the sweeping orientation from this point on.

With this preamble it is possible to state a classical paradox for first-species counterpoint theory: all the intervals $c + \epsilon.k$ used in a first-species counterpoint composition or improvisation are consonances. Hence, how can any tension between the voices arise, if at all? Mazzola's solution is inspired in the fact [6, pp. 33–35] that it is not that the point c which is to be confronted against $c + k$, but it is the consonant point $\xi = c_1 + \epsilon.k_1$ who will face a successor $\eta = c_2 + \epsilon.k_2$. The idea is to *deform* the dichotomy $(X[\epsilon]/Y[\epsilon])$ into $(gX[\epsilon], gY[\epsilon])$ through a symmetry $g \in \overrightarrow{GL}(\mathbb{Z}_{2k}[\epsilon])$, such that

1. the interval ξ becomes a deformed dissonance, i.e., $\xi \in gY[\epsilon]$,
2. the symmetry p_Δ^c is an autocomplementary function of

$$(gX[\epsilon], gY[\epsilon])$$

   which means that $p(gX[\epsilon]) = gY[\epsilon]$,

and thus we can transit from ξ to a consonance η which is also a deformed consonance, i.e., $\eta \in gX[\epsilon] \cap X[\epsilon]$. Since we wish to have the maximum amount of choices, we request also that

3. the set $gX[\epsilon] \cap X[\epsilon]$ is of maximum cardinality among the symmetries that satisfy conditions 1 and 2.

The elements of this latter set are the *admitted successors*.

3 Dichotomies of 2-Intervals

For the purposes of the second-species counterpoint, we need now an algebraic structure such that two intervals can be attached to a base tone. In the spirit of the model presented in the previous section, we take all the polynomials of the form[2]

$$c + \epsilon_1.x + \epsilon_2.y \in \frac{\mathbb{Z}_{2k}[\mathcal{X}, \mathcal{Y}]}{\langle \mathcal{X}^2, \mathcal{Y}^2, \mathcal{X}\mathcal{Y} \rangle} = \mathbb{Z}_{2k}[\epsilon_1, \epsilon_2]$$

where $\epsilon_1 \equiv \mathcal{X} \bmod \langle \mathcal{X}^2, \mathcal{Y}^2, \mathcal{X}\mathcal{Y} \rangle$, $\epsilon_2 \equiv \mathcal{Y} \bmod \langle \mathcal{X}^2, \mathcal{Y}^2, \mathcal{X}\mathcal{Y} \rangle$, c is the cantus firmus and x, y are the intervals (x is for the downbeat and y is for the upbeat). An element $\xi \in \mathbb{Z}_{2k}[\epsilon_1, \epsilon_2]$ is called a 2-*interval*. If $\Delta = (X/Y)$ is a strong dichotomy with polarity $p = T^u \circ v$, then

$$X[\epsilon_1, \epsilon_2] := \mathbb{Z}_{2k} + \epsilon_1.X + \epsilon_2.\mathbb{Z}_{2k}$$

is an dichotomy in $\mathbb{Z}_{2k}[\epsilon_1, \epsilon_2]$. We choose this dichotomy because the rules of counterpoint demand that the interval that comes on the downbeat to be a

---

[2] The original inspiration for using dual numbers in counterpoint was the Zariski tangent space, thus the definition of the tangent space of a morphism of schemes can be seen as a cue to use this kind of algebraic structure for second-species. See [7] for details.

consonance. A polarity for this dichotomy, which is analogous to the one for the first-species case, is

$$p^c = T^{c(1-v)+\epsilon_1.u+\epsilon_2.u} \circ v$$

because

$$
\begin{aligned}
p^c X[\epsilon_1, \epsilon_2] &= T^{c(1-v)} \circ v.\mathbb{Z}_{2k} + \epsilon_1.pX + \epsilon_2.p\mathbb{Z}_{2k} \\
&= \mathbb{Z}_{2k} + \epsilon_1.Y + \epsilon_2.\mathbb{Z}_{2k} \\
&= Y[\epsilon_1, \epsilon_2]
\end{aligned}
$$

and it is such that

$$p^c(c + \epsilon_1.\mathbb{Z}_{2k} + \epsilon_2.\mathbb{Z}_{2k}) = c + \epsilon_1.\mathbb{Z}_{2k} + \epsilon_2.\mathbb{Z}_{2k},$$

which means p^c fixes the tangent space to cantus firmus c as well.

We also check the following formula for future use:

$$
\begin{aligned}
p^{c_1+c_2} &= T^{(c_1+c_2)(1-v)+\epsilon_1.u+\epsilon_2.u} \circ v \tag{1} \\
&= T^{c_1(1-v)+c_2(1-v)+\epsilon_1.u+\epsilon_2.u} \circ v \\
&= T^{c_1} \circ T^{-vc_1} \circ T^{c_2(1-v)+\epsilon_1.u+\epsilon_2.u} \circ v \\
&= T^{c_1} \circ T^{c_2(1-v)+\epsilon_1.u+\epsilon_2.u} \circ v \circ T^{-c_1} \\
&= T^{c_1} \circ p^{c_2} \circ T^{-c_1}.
\end{aligned}
$$

4 Species Projections

If we represent the polynomial $c+\epsilon_1.x+\epsilon_2.y$ as a column vector, the candidates to (non-invertible) *species projections* are

$$
g : \mathbb{Z}_{2k}[\epsilon_1, \epsilon_2] \rightarrow \mathbb{Z}_{2k}[\epsilon_1], \tag{2}
$$

$$
\begin{pmatrix} c \\ x \\ y \end{pmatrix} \mapsto \begin{pmatrix} s & 0 & 0 \\ sw_1 & s & sw_2 \end{pmatrix} \begin{pmatrix} c \\ x \\ y \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}
$$

$$
= [sc + t_1] + \epsilon_1.[s(w_1c + x + w_2y) + t_2]
$$

for we want to keep it as simple as possible and that the upbeat of the first interval to influence the downbeat of the successor, but not its upbeat one. We do not require the transformation to be bijective for we want it to be able to swap from second-species to first-species if necessary[3].

---

[3] For the converse swap the standard rules of counterpoint suffice: we can arbitrarily define the third component of the 2-interval. This is coherent with the local application of counterpoint rules in Fux's theory, and also with the particular idea of projection that stems from the fact that, in order to analyze a fragment, we "disregard" notes on the upbeat [3, pp. 41–43].

Definition 1. *A matrix of the form that appears in a species projection is called a* projection matrix.

Let $X[\epsilon_1, \epsilon_2.y] := \mathbb{Z}_{2k} + \epsilon_1.X + \epsilon_2.y$. We might define a counterpoint projection of a 2-interval $\xi = c + \epsilon_1.x + \epsilon_2.y$ as one such that

1. the condition $c + \epsilon_1.x \notin gX[\epsilon_1, \epsilon_2.y]$ holds,
2. the square

$$
\begin{array}{ccc}
\mathbb{Z}_{2k}[\epsilon_1, \epsilon_2] & \xrightarrow{\ g\ } & \mathbb{Z}_{2k}[\epsilon_1] \\[4pt]
{\scriptstyle p^c}\big\downarrow & & \big\downarrow{\scriptstyle p^c_\Delta} \\[4pt]
\mathbb{Z}_{2k}[\epsilon_1, \epsilon_2] & \xrightarrow[\ g\ ]{} & \mathbb{Z}_{2k}[\epsilon_1]
\end{array}
\tag{3}
$$

commutes, where

$$
p^c_\Delta := T^{c(1-v)+\epsilon_1.u} \circ v
$$

is the *canonical* polarity of $(X[\epsilon_1]/Y[\epsilon_1])$, and
3. the cardinality of $gX[\epsilon_1, \epsilon_2.y] \cap X[\epsilon_1]$ is maximal among the projections with the previous properties.

The reason for the second requirement is that when it is fulfilled then

$$
p^c_\Delta(gX[\epsilon_1, \epsilon_2]) = g(p^c X[\epsilon_1, \epsilon_2]) = gY[\epsilon_1, \epsilon_2],
$$

thus p^c_Δ is an autocomplementary function of $gX[\epsilon_1, \epsilon_2]$.

5   Algorithm for the Calculation of Projections

As with the first-species case, if for a projection of the form

$$
g = T^{\epsilon_1.t_2} \circ M
$$

where M is a projection matrix, we define

$$
g^{(t_1)} = g \circ T^{\epsilon_1.s^{-1}w_1 t_1 + \epsilon_2.t_1}
$$

then the relation

$$
T^{t_1} \circ g = g^{(-t_1)} \circ T^{s^{-1}t_1 + \epsilon_2.t_1},
\tag{4}
$$

holds, and hence contrapuntal projections can be calculated with cantus firmus 0 and successors can be suitably adjusted [2, Theorem 2.2].

Remark 1. The groups

$$
T^{\mathbb{Z}_{2k}}, T^{\mathbb{Z}_{2k}+\epsilon_2\mathbb{Z}_{2k}}
$$

are subgroups of the group of automorphisms of $X[\epsilon_1]$ and $X[\epsilon_1, \epsilon_2]$, respectively.

The following identities are needed for the simplification of the calculation of contrapuntal symmetries.

Lemma 1. *For a species projection of the form $g = T^{\epsilon_1.t_2} \circ M$ the following holds:*

$$(g^{(t_1)})^{(t_2)} = g^{(t_1+t_2)},$$
$$T^t \circ g(X[\epsilon_1, \epsilon_2]) = g^{(-t)}(X[\epsilon_1, \epsilon_2]) \quad and$$
$$T^t \circ g(Y[\epsilon_1, \epsilon_2]) = g^{(-t)}(Y[\epsilon_1, \epsilon_2]).$$

Proof. The first identity is straightforward:

$$(g^{(t_1)})^{(t_2)} = (g \circ T^{\epsilon_1.s^{-1}w_1 t_1 + \epsilon_2.t_1})^{(t_2)}$$
$$= g \circ T^{\epsilon_1.s^{-1}w_1 t_1 + \epsilon_2.t_1} \circ T^{\epsilon_1.s^{-1}w_1 t_2 + \epsilon_2.t_2}$$
$$= g \circ T^{\epsilon_1.s^{-1}w_1(t_1+t_2) + \epsilon_2.(t_1+t_2)}$$
$$= g^{(t_1+t_2)}.$$

For the second identity, note that

$$(T^t \circ g)(X[\epsilon_1, \epsilon_2]) = g^{(-t)} \circ T^{s^{-1}.t + \epsilon_2.t}(X[\epsilon_1, \epsilon_2])$$
$$= g^{(-t)} X[\epsilon_1, \epsilon_2]$$

using (4) and Remark 1. The case for $Y[\epsilon_1, \epsilon_2]$ is proved mutatis mutandis. □

Remark 2. If we have a species projection of the form $g = T^{z+\epsilon_1.t} \circ M$, then we define $f = T^{\epsilon_1.t} \circ M$ and thus $g = T^z \circ f$. Using Lemma 1, we have

$$g(X[\epsilon_1, \epsilon_2]) = (T^z \circ f)(X[\epsilon_1, \epsilon_2]) = f^{(-z)}(X[\epsilon_1, \epsilon_2]).$$

This means that in our discussion it suffices to consider projections whose translational part has zero non-dual component.

The following pair of results reduce the amount of computations required to obtain counterpoint projections.

Lemma 2. *Let $\xi = x + \epsilon_1.k$, g a species projection, and $z \in \mathbb{Z}_{2k}$. If*

$$\xi \notin g(X[\epsilon_1, \epsilon_2]) \quad and \quad p_\Delta^x : g(X[\epsilon_1, \epsilon_2]) \xrightarrow{\cong} g(Y[\epsilon_1, \epsilon_2])$$

then

$$T^z(\xi) \notin (T^z \circ g)(X[\epsilon_1, \epsilon_2]) \quad and$$
$$p_\Delta^{z+x} : (T^z \circ g)(X[\epsilon_1, \epsilon_2]) \xrightarrow{\cong} (T^z \circ g)(Y[\epsilon_1, \epsilon_2]).$$

Furthermore,

$$(T^z \circ g)(X[\epsilon_1, \epsilon_2]) \cap X[\epsilon_1] = T^z(g(X[\epsilon_1, \epsilon_2]) \cap X[\epsilon_1])$$

and, in particular,

$$|(T^z \circ g)(X[\epsilon_1, \epsilon_2]) \cap X[\epsilon_1, \epsilon_2]| = |g(X[\epsilon_1, \epsilon_2]) \cap X[\epsilon_1, \epsilon_2]|.$$

Proof. Since T^z is a symmetry of $g(X[\epsilon_1, \epsilon_2])$, it follows that $T^z(\xi) \notin T^z(g(X[\epsilon_1, \epsilon_2]))$. Now, using (1),

$$
\begin{aligned}
(p_\Delta^{x+z} \circ T^z \circ g)(X[\epsilon_1, \epsilon_2]) &= (T^z \circ p_\Delta^x \circ T^{-z} \circ T^z \circ g)(X[\epsilon_1, \epsilon_2]) \\
&= (T^z \circ p_\Delta^x \circ g)(X[\epsilon_1, \epsilon_2]) \\
&= (T^z \circ g)(Y[\epsilon_1, \epsilon_2]).
\end{aligned}
$$

From Remark 1 it follows that

$$
\begin{aligned}
(T^z \circ g)(X[\epsilon_1, \epsilon_2]) \cap X[\epsilon_1, \epsilon_2] &= (T^z \circ g)(X[\epsilon_1, \epsilon_2]) \cap T^z(X[\epsilon_1, \epsilon_2]) \\
&= T^z(g(X[\epsilon_1, \epsilon_2]) \cap X[\epsilon_1])
\end{aligned}
$$

since T^z is bijective.   □

Theorem 1. *If $\xi = x + \epsilon_1.k + \epsilon_2.z \in X[\epsilon_1, \epsilon_2]$ and $g = T^{t_1 + \epsilon_1.t_2} \circ M$ is any species projection that satisfies the counterpoint conditions, then there is a species projection $h = T^{\epsilon_1.t} \circ M$ such that it also satisfies the counterpoint conditions for ξ. Moreover: in order to verify that the conditions also hold for h, it suffices to check them for the 2-interval $\epsilon_1.k + \epsilon_2.z$, the projection $h^{(x)}$ and the polarity p_Δ^0.*

Proof. The replacement of g follows from Remark 2. By Lemma 1, we have

$$
(T^{-x} \circ h)(X[\epsilon_1, \epsilon_2]) = h^{(x)}(X[\epsilon_1, \epsilon_2]).
$$

Using Lemma 2 with $z = -x$, we can verify that h is a counterpoint projection using $h^{(x)}$ with the interval $T^{-x}(\xi) = \epsilon_1.k + \epsilon_2.z$ and the polarity $p_\Delta^{-x+x} = p_\Delta^0$. From Lemma 2 it also follows that

$$
\begin{aligned}
(h^{(x)}(X[\epsilon_1, \epsilon_2]))) \cap (X[\epsilon_1] &= (T^{-x} \circ h)(X[\epsilon_1, \epsilon_2])) \cap X[\epsilon_1] \\
&= T^{-x}(h(X[\epsilon_1, \epsilon_2]) \cap X[\epsilon_1])
\end{aligned}
$$

which implies that any cardinalities computation we need to perform with h will be the same than doing them with $h^{(x)}$.   □

Therefore, we can set $t_1 = 0$ and work with intervals of the form $\xi = \epsilon_1.y + \epsilon_2.z$. For (3) to commute, it is necessary and sufficient that

$$
t_2 + su(1 + w_2) = u + vt_2. \tag{5}
$$

For $\epsilon_1.y \notin gX[\epsilon_1, \epsilon_2.z]$ we need

$$
y = sp(\ell) + t_2 + sw_2 z
$$

for some $\ell \in X$. Hence, for some $\ell \in X$ we have

$$
t_2 = y - s(p(\ell) + w_2 z). \tag{6}
$$

Remark 3. Letting $w_2 = 0$ in (5) and (6), they reduce to the first-species case. Thus, taking $s = v$ and $\ell = y$ both are satisfied and hence we conclude that there exists at least one second-species counterpoint projection.

We only need to work with the following set

$$
\begin{aligned}
gX[\epsilon_1, \epsilon_2.z] &= \bigcup_{x \in \mathbb{Z}_k} g(x + \epsilon_1.X + \epsilon_2.z) \\
&= \bigcup_{x \in \mathbb{Z}_{2k}} (sx + \epsilon_1.(sw_1 x + sw_2 z + t_2 + sX)) \\
&= \bigcup_{r \in \mathbb{Z}_{2k}} (r + \epsilon_1.(w_1 r + sX + w_2 sz + t_2)) \\
&= \bigcup_{r \in \mathbb{Z}_{2k}} (r + \epsilon_1.T^{w_1 r + w_2 sz + t_2} \circ sX)
\end{aligned}
$$

to calculate the following cardinality

$$
|gX[\epsilon_1, \epsilon_2.z] \cap X[\epsilon_1, \epsilon_2.z]| = \sum_{r \in \mathbb{Z}_{2k}} |T^{w_1 r + w_2 sz + t_2} \circ sX \cap X|.
$$

When (6) holds, this reduces to

$$
|gX[\epsilon_1, \epsilon_2.z] \cap X[\epsilon_1, \epsilon_2.z]| = \sum_{r \in \mathbb{Z}_{2k}} |T^{w_1 r + y - sp(\ell)} \circ sX \cap X|. \tag{7}
$$

From now on we only need to adapt *mutatis mutandis* Hichert's algorithm [2, Algorithm 2.1] to search projections that maximize the intersection.

We must remark that (5) and (6) are perturbations of the conditions to find the counterpoint symmetries for the first-species case. These, together with (7), show that the conditions for deducing a counterpoint theorem [2, Theorem 2.3] hold again, which yields the following result.

Theorem 2. *Given a marked strong dichotomy (X/Y) in $\mathbb{Z}_{2k}$, the 2-interval $\xi \in X[\epsilon_1, \epsilon_2]$ has at least k^2 and at most $2k^2 - k$ admitted successors given by a single counterpoint projection.*

Algorithm 3. *Here $\chi(x, y)$ is the function that returns the cardinality $T^x.yX \cap X$.*
Input: *A strong dichotomy $\Delta = (X/Y)$ and its polarity $T^u.v$.*
Output: *The set of counterpoint projections $\Sigma_{y,z} \subseteq H$ for each $\epsilon_1.y + \epsilon_2.z \in X[\epsilon_1, \epsilon_2]$.*
1: **for all** $y \in X$ *and* $z \in \mathbb{Z}_{12}$ **do**
2:     $M \leftarrow 0, \Sigma_{y,z} \leftarrow \emptyset$.
3:     **for all** $s \in GL(\mathbb{Z}_{2k})$ **do**
4:         **for all** $\ell \in X$ **do**
5:             **for all** $w_1, w_2 \in \mathbb{Z}_{2k}$ **do**
6:                 $t_2 \leftarrow y - s((v\ell + u) + w_2 z)$.

7:         **if** $t_2 + su(1 + w_2) = u + vt_2$ **then**

8:          **if** $w_1 = 0$ **then**

9:            $S \leftarrow 2k\chi(t_2, s)$.

10:        **else if** $w_1 \in GL(\mathbb{Z}_{2k})$ **then**

11:            $S \leftarrow k^2$

12:        **else**

13:            $\rho \leftarrow \gcd(w_1, 2k)$

14:            $S \leftarrow \rho \sum_{j=0}^{\frac{2k}{\rho}-1} \chi(j\rho + t_2 + w_2 z, s)$.

15:          **if** $S > M$ **then**

16:            $\Sigma_{y,z} \leftarrow \left\{ T^{\epsilon_2 . t_2} \circ \begin{pmatrix} s & 0 & 0 \\ sw_1 & s & sw_2 \end{pmatrix} \right\}$.

17:            $S \leftarrow M$.

18:          **else if** $S = M$ **then**

19:            $\Sigma_{y,z} \leftarrow \Sigma_{y,z} \cup \left\{ T^{\epsilon . t_2} \circ \begin{pmatrix} s & 0 & 0 \\ sw_1 & s & sw_2 \end{pmatrix} \right\}$.

20:    **return** $\Sigma_{y,z}$.

Example 1. The first (valid[4]) example of second-species counterpoint in the *Gradus ad Parnassum* [3, p. 45] is (see Fig. 1)

$$\xi_1 = 2 + \epsilon_1.7 + \epsilon_2.0, \ \xi_2 = 5 + \epsilon_1.4 + \epsilon_2.6, \ \xi_3 = 4 + \epsilon_1.8 + \epsilon_2.3,$$
$$\xi_4 = 2 + \epsilon_1.7 + \epsilon_2.0, \ \xi_5 = 7 + \epsilon_1.4 + \epsilon_2.5, \ \xi_6 = 5 + \epsilon_1.9 + \epsilon_2.4,$$
$$\xi_7 = 9 + \epsilon_1.3 + \epsilon_2.5, \ \xi_8 = 7 + \epsilon_1.9 + \epsilon_2.4, \ \xi_9 = 5 + \epsilon_1.9 + \epsilon_2.4,$$
$$\xi_{10} = 4 + \epsilon_1.7 + \epsilon_2.9, \ \xi_{11} = 2 + \epsilon_1.0.$$

Fig. 1. First (valid) example of second-species counterpoint in Fux's *Gradus ad Parnassum*.

---

[4] The first example is the student's attempt to write a second-species discantus by himself, but he makes two mistakes near the end of the exercise, namely the steps from the sequence $7 + \epsilon_1.7 + \epsilon_2.4$, $5 + \epsilon_1.7 + \epsilon_2.4$, $4 + \epsilon_1.7 + \epsilon_2.9$. They are also forbidden steps in the projection model!.

Some counterpoint projections for the successors are

$$g_1 = \begin{pmatrix} 7\ 0\ 0 \\ 0\ 7\ 0 \end{pmatrix}, g_2 = T^{\epsilon_1.6} \circ \begin{pmatrix} 1\ 0\ 0 \\ 6\ 1\ 0 \end{pmatrix}, g_3 = T^{\epsilon_1.6} \circ \begin{pmatrix} 7\ 0\ 0 \\ 6\ 7\ 0 \end{pmatrix}$$

$$g_4 = g_1, g_5 = g_2, g_6 = T^{\epsilon_1.8} \circ \begin{pmatrix} 5\ 0\ 0 \\ 8\ 5\ 0 \end{pmatrix},$$

$$g_7 = \begin{pmatrix} 11\ 0\ 0 \\ 0\ 11\ 8 \end{pmatrix}, g_8 = g_6, g_9 = g_6, g_{10} = g_1.$$

Let us examine in little bit more of detail the first transition. Note that $\eta = 11 + \epsilon_1.4 + \epsilon_2.11$ is a consonance, and that

$$g_1(\eta) = \begin{pmatrix} 7\ 0\ 0 \\ 0\ 7\ 0 \end{pmatrix} \begin{pmatrix} 11 \\ 4 \\ 11 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \end{pmatrix},$$

which justifies the fact that the 2-interval $5 + \epsilon_1.4 + \epsilon_2.6$ is an admitted successor.

6   Comparison with Fux's Approach

Fux states the following in relation to second-species counterpoint (emphasis is our own) [3, p. 41]:

> The second species results when two half notes are set against a whole note. The first of them comes on the downbeat and must always be consonant; the second comes on the upbeat and *it may be dissonant if it moves from the preceding note and to the following note stepwise.* However, *if it moves by a skip, it must be consonant.*

We coded[5] in Octave the calculation of counterpoint projections for the Fux-ian (K/D) dichotomy and some more to compare the performance between "restricted" Fux rules against the projection model. More explicitly, taking a second-species step

$$(0 + \epsilon_1.k_1 + \epsilon_2.t_1, c_2 + \epsilon_1.k_2)$$

such that we can proceed (in first-species) from $0 + \epsilon_1.k_1$ to $c_2 + \epsilon_1.k_2$, we verify the following cases:

1. the upbeat interval t_1 of the first 2-interval is allowed to be dissonant only when it connects a valid progression of consonances stepwise, i.e., $0 + t_1$ is between $0 + k_1$ and $c_2 + k_2$ and it is separated at most 2 semitones from them and
2. if t_1 is consonant, we duplicate the cantus firmus and check if $(0 + \epsilon.k_1, 0 + \epsilon.t_1)$, $(0 + \epsilon.t_1, c_2 + \epsilon.k_2)$ and $(0 + \epsilon.k_1, 0 + \epsilon.k_2)$ are valid first-species steps.

The results appear in Table 1 for cases 1 and 2.

Table 1. Data for comparison of Fux's model with restrictions for second species against the projection model.

Number of steps	Case 1	Case 2
Total	192	2592
Valid only for Fux model	13	107
Valid only for the projection model	50	1227
Valid in both models	129	1137

We note that the number of cases the projection model cannot explain and only Fux can is relatively small: they amount to 6.8% and 4.1% for cases 1 and 2, respectively. Thus we can conclude that the vast majority of what is allowed by Fux's rules is allowed by the projection model, or that we have successfully extended Fux's handling of dissonance and consonance for second species. Even if this could be ascribed to the fact that the projection model admits 93.229% and 91.204% of the total of transitions in cases 1 and 2, respectively, it should be kept in mind that the original one-species model admits 89.671% of the possible steps between consonant intervals [5, p. 48].

References

1. Agustín-Aquino, O.A.: Counterpoint in $2k$-tone equal temperament. J. Math. Music **3**(3), 153–164 (2009)
2. Agustín-Aquino, O.A., Junod, J., Mazzola, G.: Computational Counterpoint Worlds. Springer, Heidelberg (2015)
3. Mann, A.: The Study of Counterpoint. W. W. Norton & Company (1965)
4. Mazzola, G.: The Topos of Music, vol. I, 2nd edn. Springer, Heidelberg (2017)
5. Nieto, A.: Una aplicación del teorema de contrapunto. B.Sc. thesis (2010)
6. Sachs, K.J.: Der Contrapunctus im 14. und 15. Jahrhundert, Beihefte zum Archiv fü Musikwissenschaft, vol. 13. Franz Steiner Verlag (1974)
7. The Stacks project authors: The Stacks project, Section 0B28 (2022). https://stacks.math.columbia.edu/tag/0B28

When Virtual Reality Helps Fathom Mathemusical Hyperdimensional Models

Gilles Baroin[1,2,3(✉)] and Stéphane de Gérando[4]

[1] Laboratoire ENAC, Université Fédérale de Toulouse, Toulouse, France
`Gilles@Baroin.org`
[2] Laboratoire LLA Créatis, Université Fédérale de Toulouse, Toulouse, France
[3] Neometsys NMS Lab, Toulouse, France
[4] Conservatoire Ville de Paris, Paris, France

Abstract. Mathemusicians have always produced models for understanding, analyzing or computing music. We are used to visualize some of them on paper, in a theater or on a computer screen.

Even if they refer to multidimensional spaces (3D-4D), while displaying these models on a computer screen the viewer ends up with a 2D picture, or a movie.

Planar projection limits the perception, nowadays, in the era of virtual reality, we propose tools and solutions to better apprehend these models and give the viewer an improved immersive experience.

Taking advantage of methods used in air traffic simulations, we are developing techniques that we will apply to existing mathemusical visualizations, beginning with *Tonnetze* and *Hyperspheres*.

We herewith introduce two recently revealed mathemusical models that we have created:

2D: The *Shadow Tonnetz*, our latest extension of the Tonnetz that keeps trace of a harmonic path.

4D: The *Entangled Hyperspheres*, a combination of two *Planet-4D* models that enables us to visualize microtonal music.

The images in this paper are extracted from immersive virtual reality world; during MCM we intend to presented the movies with adapted 3D equipment.

All videos including virtual ones will be available on www.mathemusic.net.

Keywords: Virtual reality · Mathemusic · Visualization

1 Context and Definitions

1.1 Mathemusical Models

At this time, we only focus on models representing musical objects: scales, chords or pitches expressed by their frequencies and do not intend to embody other musical parameters such as interpretation, *Klangfarbe*,…

S. de Gérando—Directeur Publication 3Icar Editions.

We start the voyage by implementing virtual reality for the representations of Z_{12}, the *Tonnetze*, musical graphs and original spatial models that we have developed: from *Planet-4D* [1] to the *Hypersphere of Spectra* applied to spectral music [2] which has been premiered at Ircam Paris in 2012.

The mathematical model behind the Spheres:

All hyperspheres that we employ are topologically equivalent to S_3, they are four dimensional objects, we place the pitches on their 3D surfaces. Thus, the positions of our objects on the surface are defined by a unitary quaternion that we can interpret, depending on the usage as:

– Four spatial coordinates $q \in R^4 : q(x, y, z, w)$ for computing projections,
– two angles $q \in \mathbb{C} \times \mathbb{C} : q(e^{i\alpha}, e^{i\vartheta})$ for rotations,
– or a pure quaternion $q \in \mathcal{H}$ for calculations.

In order to achieve an additional symmetry on Euclidian distances between the 12 pitch classes, our *Spheres* are scaled differently with respect to the two complex planes: they are actually "hyper ellipsoids". Since they are topologically equivalent to S_3, we are used to call them *Hyperspheres* or *Spheres* [1]. All possible positions of elementary objects (pitches) are hence located on the following curve defined as:

$$F(x) = Q_x = \left(\frac{1}{\sqrt{3}} e^{i\frac{2x\pi}{3}}, \frac{1}{\sqrt{2}} e^{i\frac{2x\pi}{4}} \right) \quad with \quad x \in R^+, x < 12; Q_x \in \mathcal{H}.$$

This apply to equal temperament only; for different tuning systems, the *Sphere* is still topologically equivalent to S_3, but its radius varies with the pitch, according to the specific tuning system [3].

1.2 Virtual Reality

Virtual reality (VR) offers a simulated experience to the spectator who is plunged in a virtual world with the help of a specific headset. A person using virtual reality equipment is able to look around the artificial world, move inside in it, and interact with virtual features or items. The realization is commonly made up on VR headsets consisting of a head-mounted display with two small screens in front of the eyes [5].
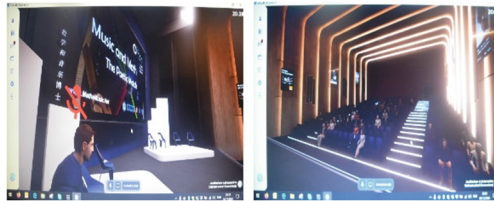

Fig. 1. Premiere of the *Entangled Hyperspheres* during virtuality.

We were honored to give a keynote lecture during the event Virtuality 2020 which, due to sanitary conditions, took place virtually at the same time in Buenos Aires and Paris [4]. During this online event, attendees were acting via their avatars, with the ability to talk to each other's, and participate to conferences in virtual auditoriums. As displayed in Fig. 1, the lecturer is on the stage while the attendees sit in a virtual theater.

For this first appearance we merely hosted a session on Mathemusic while displaying movies on the big screen in the auditorium. Users had the possibility to interact with the speakers and download the movies.

At this time, we already noticed that movies projected within a virtual environment, probably due to the context, have a different impact on the user experience. The virtual world experience provides a better awareness, when compared to a usual remote Zoom® presentation.

On the other hand, we use our experience in modelling and visualizing air traffic and airplanes trajectories for applying the virtual reality tools to the mathemusical field. As in real air traffic simulation, we have to position moving objects in a virtual space; For detailed explanations on VR features and different techniques, please refer to [5]. On the left image of See Fig. 2, we use VR to examine planned trajectories. In order to simulate drone traffic, we use a test environment structured as a virtual city [7].

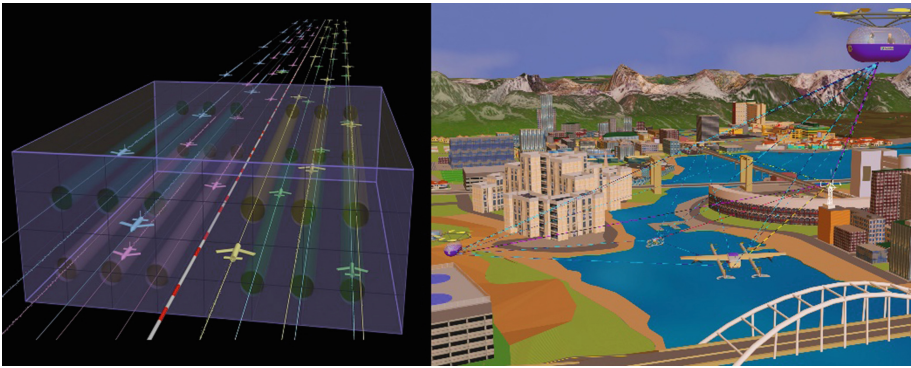**Fig. 2.** Air traffic simulation (Neometsys NMS Lab for FC2A, European project), Virtual city used for drone simulations in civil aviation (Neometsys Lab/Baroin).

1.3   Immersive World Generation

Instead of using a traditional physical camera that produces 2D images, a 360° camera generates images that may be assembled in an immersive media: In case of a movie, objects and camera are animated.

NB: The appellation 360° is a commercial denomination; we mathematicians prefer the accurate term 4π steradian.

Technically, a 4π sr image is a flat image representing the full $360 \times 180°$ 3-D sphere (S_2). It is an extension of the 360° panoramas we are familiar with and accessing with digital cameras or smartphones. The image is then projected on a sphere around the user: everything in all directions is hereafter visible to the observer: front to back and straight up and down. Please refer to Fig. 3 below.
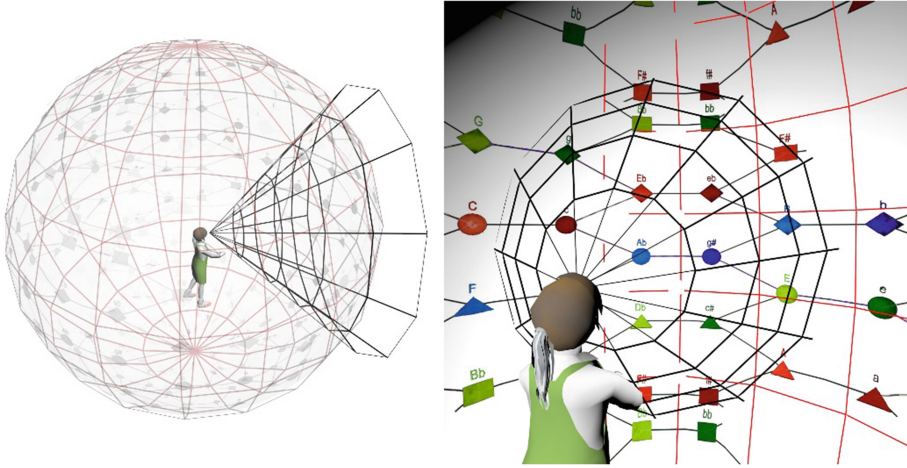

Fig. 3. A 360° projection around the observer featuring *Polarized-Tonnetz* by H. Seress and G. Baroin.

2   Methodology

2.1   Tools

As hardware, we are currently using Oculus Quest2® (now sold as "Meta Quest2 by Facebook®"), any other headset with VR capabilities would be suitable to visualize the assets.

For the software, Microsoft Excel® and Autodesk 3DSMax® perform the modeling and calculation of the images. After montage and postprocessing, the movie is finalized as a specific mp4 video file that is ready to be played in any VR environment.

For interactive applications, we are currently working with the game development environment Unity®.

As film makers need physical cameras in the real-world, we use virtual cameras for our computer-generated imagery. These cameras simulate realistically the physical cameras. They are fully configurable and have the main advantage of being able to be positioned in places that would be inaccessible in the physical world. Especially in the case we would like to observe the inside of a hypersphere!

2.2   VR Techniques

For mathemusical models we currently employ two main techniques, depending on the subject and on the mathematical dimension where the model is defined:

Monoscopic movie projection in 360°, and Stereoscopic virtual camera:

– Monoscopic, means that the video has one single channel, but the movie is actually displayed in the headset for both eyes.
– Stereoscopic refers to a video encompassing 2 channels; one for each eye, with slightly different viewing angles. Stereoscopy gives you the same depth perception as you can experience while watching a 3D-movie with specific glasses in a theater. A stereoscopic movie is produced in a form of a clip where the left eye view is separated from the right one. Movies can be viewed on a smartphone with a cheap headset that only separate eyes channels, projected in a theater featuring 3D equipment, or on a computer powered headset as we do.

2.2.1   Monoscopic Movie Projection in 360°

The monoscopic 360° format was the first to be employed and is nowadays the most widely used format for immersive video. One can buy 360° cameras for filming the environment then publish the media, for instance on YouTube 360°®, and play the movies in a headset.

Mono 360° is typically a rectilinear video container with a 2:1 aspect ratio, that is projected on a sphere. To understand spherical projection, one can imagine projecting an existing flat world map, back onto our earth.
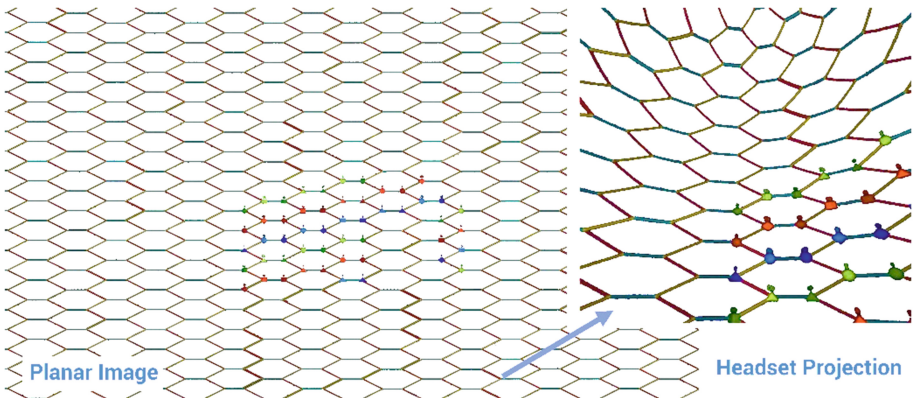

Fig. 4. From 2D to spherical environment (*Polarized Tonnetz* by H. Seress and G. Baroin).

This technique is adapted to representing planar Tonnetze: as in Fig. 4, the source image is a large representation of the dual of the Tonnetz that we project on the spherical environment; the image on the right is a simple screen capture of the headset view, the observer looks straight ahead.

Since there are theoretically no boundaries in the 2D planes, we map the Tonnetze as a spherical projection on the sphere surrounding the observer. The Tonnetz image that were used in the former animations are extended so that the region where the animation takes place is located in front of the observer. This type of projection gives the impression of being plunged into an infinite space. Moreover, the intrinsic curvature of the restorative spherical environment participates to this effect.

2.2.2 Stereoscopic Virtual Camera

As 3D objects cannot be fully depicted in a plane, but only imagined thanks to their projection, four dimensional models will never be fully graspable in our world. By projecting a 4D structure to 3D, it is then possible to fathom its 3D shadow. The 3D projection is then, with the help of VR, viewed in three dimensions, this leads to a closer perception of the reality of 4D space. Since our 4D models have to be projected to 3D, the stereoscopic technique is best adapted.

In the virtual environment, the camera is placed at the viewer's eye level, sometimes called FPV (first person view, in games phraseology). The mathemusical objects are constructed and animated in the virtual world.

The user can move freely in a predefined area (for safety reasons) and look around the animated world in 360°.

For interactive VR, we ought to structure the scenes withing a game engine, such as Unity® or Unreal Engine® which are the most advanced tools at this point, and became the standard in this industry. In a game engine, the calculation and rendering are performed in real-time on a computer connected to the headset, or within the headset if it features the possibility.

For our applications, in order to have the feeling to be within the *Hypersphere*, the virtual camera has to be set near the reference point. This reference point is the user's position as formulated in each *Planet-4*D models [1].
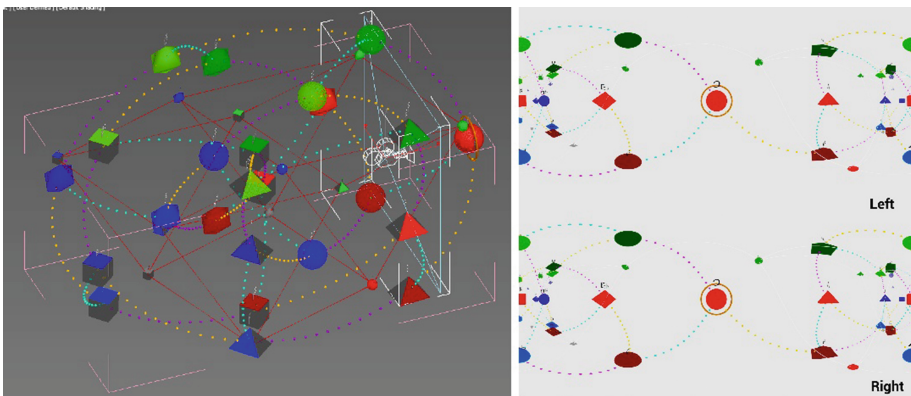

Fig. 5. Left: Virtual camera within The *Hypersphere of Chords* by G.Baroin. Right: Stereoscopic output of the scene.

Among different possible standard rendering formats for VR, we have selected the "Top/Bottom" configuration. As displayed in Fig. 5, the rendering produces a square image made of two pictures with 2:1 ratio. While the top part is destinated to the left eye, the bottom part is intended to the right eye. The decoding software is a VR movie player in the headset or a computer performing the stereoscopic vision reconstruction.

NB: For 4D objects, the projection to 3D is carried out before displaying the objects. As we are accustomed to, we perform an orthographic fixed projection from 4D [1]. This transformation is the most user-friendly conversion to 3D and is fairly intelligible, even for inexperienced viewers.

3   Applications

3.1   2D Space: Virtual Swan Lake

This is an immersive presentation of "Tchaikovsky Swan Lake" theme in the *Shadow Tonnetz.*

The *Shadow Tonnetz* is the latest extension of the Tonnetz visualizations developed by G.Baroin & I.Khannanov [8]. It has been premiered at the Conservatory of Moscow during Euromac 2020.

Usual musical graphs, Tonnetz or *Cube-Dance*, feature chords and their Neo-Riemanian (P, L, R) or other specific relations only. Our next evolution: *Polarized Tonnetz* [9], grants a unique representation to each type of relation between neighbor chords and thus links visualization to a better musical correlation. In order to enhance the existing models, the *Shadow Tonnetz* [8] introduce two quantities defined as *Speed* and *Weight*.

Speed, from a pure mathematical point of view, is the first-time derivative of a position (in our case chord coordinates).

Weight, is a concept borrowed from "Activity On Nodes" (AON) graphs, where the time spent on a node (in this situation a chord), is called *Weight*.

In order to visualize *Speed* and *Weight*, we use a foot-printing technique that emphasis the most played path and chords. A Hamiltonian harmonic path [11] played regularly [12] results in a homogeneous trace in space, whereas more usual piece's traces display the main points of inertia (Fig. 6).
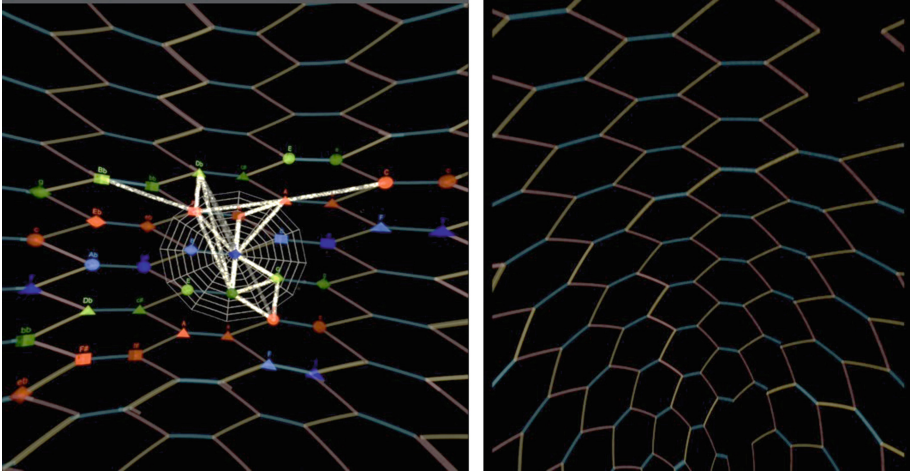
Fig. 6. Front and back view as in the headset of *Virtual Swan Lake* featuring the *Shadow-Tonnetz* by I. Khannanov and G. Baroin.

In the left image we can observe the "traces" of the harmonic path in Tchaikovsky's Swan Lake within the *Shadow Tonnetz.*

3.2   4D Space: Entangled Hypersphere

We now propose in virtual reality, the construction of the *Entangled Hyperspheres* [13], the accompanying music is [14]. Developed by G. Baroin and S. de Gérando, it enables visualization of microtonal music on the surface of two *Hyperspheres*. See below Fig. 7.

The *Entangled Hyperspheres*, as if its name was predestined, was premiered simultaneously in Buenos Aires and Paris.

For music in ¼th tone, the timbre harmonies are composed from singular pitches, a particular case of so-called "all intervals" series calculated algorithmically (world premiere) [14]. Two acronyms are invented: STIOZ or series all intervals in an octave and in zigzag, STISMI or series all intervals nested in micro-interval series.

The initial method, imagined by Ivan Wyschnegradsky for two pianos as described in [15], can be adapted for two orchestra called for instance A and B.

– Orchestra A is tuned at 440 Hz and plays the 12 usual notes (do do#,)
– Orchestra B is tuned at 440 Hz - ¼th tone (427.7 Hz), and plays the 12 additional notes (do+ do#+,…).
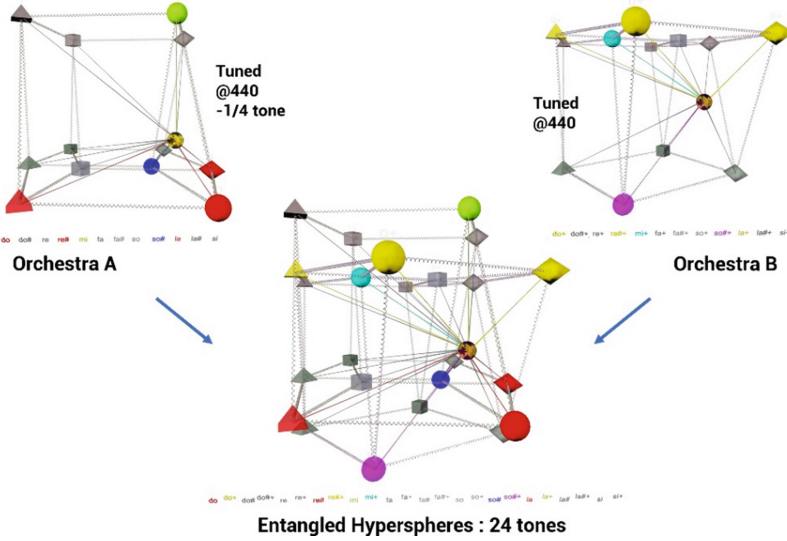
Fig. 7. Construction of the *Entangled Hyperspheres* (G. Baroin S. de Gérando 2020).

Since any aggregate of 12 tones can be displayed on the *Planet-4D* model (see *Hypersphere of Anyset*: the first representation of atonal music in hyperspace for Webern Bagatelle #6 at Ircam Paris), and since the production of 24 tones music is performed by two synchronized orchestra, the *Entangled Hyperspheres* is made of two *Planet-4D* models:

One represents the notes played by orchestra A, the other displays the music from orchestra B: Because all 24 tones are on the same curve on the *Hypersphere*, (see Sect. 1.1, Mathemusical models), when put together on the same center of view, the two *Spheres* become naturally interwoven. As personal choice, we call them *entangled*.

The challenge was to make both models share the same parameters for space and time; starting with the positions of the notes on the first sphere, then calculating the new spherical barycenter that corresponds to the user position taking into account the 24 pitch classes (for details concerning the construction, please refer to [1]).

The quaternionic coordinates of each tempered pitch class in orchestra A being:

$$Q_{nt} = Q_n f_n/f_0 \text{ with } Q_n = \left(\frac{1}{\sqrt{3}}e^{i\frac{2n\pi}{3}}, \frac{1}{\sqrt{2}}e^{i\frac{2n\pi}{4}}\right), n \in [0.11]:$$

the original coordinates of pitch class n in the Planet-4D Model. f_n, f_0.

Since orchestra B, is playing one quarter tone above orchestra A, all symbols representing its pitch classes have to be rotated 1/24th of the whole circle along the chromatic circle, that is equivalent to adding to each coordinate a constant quaternion qc:

$$qc = \left(\frac{1}{\sqrt{3}}e^{i\frac{\pi}{12}}, \frac{1}{\sqrt{2}}e^{i\frac{2\pi}{12}}\right)$$

The quaternionic coordinates of each tempered pitch class in orchestra B are therefore:

$$Q_{Bn} = Q_n + qc = \left(\frac{1}{\sqrt{3}} e^{i \frac{(8n+1)\pi}{12}} +, \frac{1}{\sqrt{2}} e^{i \frac{(6n+1)\pi}{12}} \right), n \in [0.11] :$$

the coordinates of pitch class n in the Planet-4D Model. f_n, f_0:

The new barycenter of the system is therefore the spherical barycenter of all notes played at the actual time. For detailed calculation of quaternionic spherical barycenter, please refer to "*The Entangled Hyperspheres, an innovative approach to visualize microtonal music*" [13].
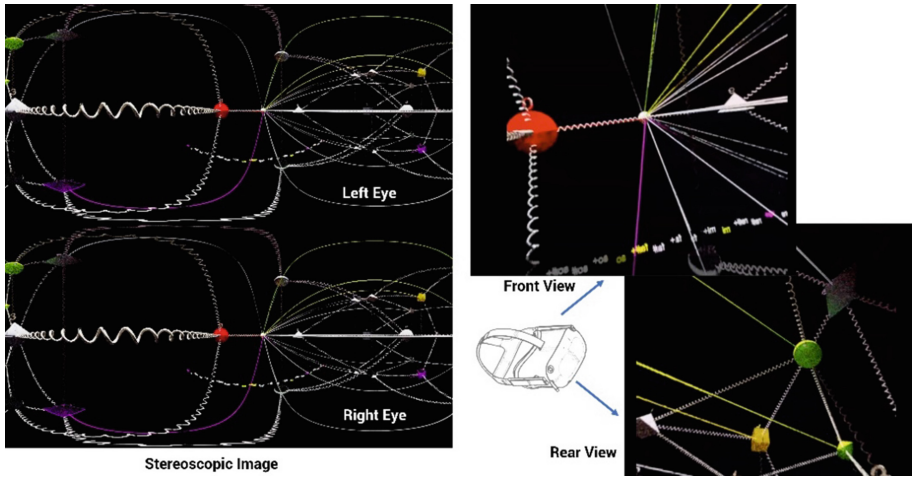
Fig. 8. Stereoscopic image and VR views for the *Entangled Hyperspheres*.

In above figure Fig. 8, the stereoscopic image that is produced by the rendering software. Front and rear views are screenshots casted from the headset onto a computer.

4 Feedback and Future

4.1 Feedback

Beside our personal feelings, we have been able to perform some tests with different types of users, child, adults, experimented mathemusicians in situation:

For experimented mathemusicians, the VR experience provides a sensation to be embedded within the model, since the display encompasses the whole viewing field, we are able to see details that were not seen on computer screens and we can focus on specific parts of the animation. Thus, the spherical aspect of Tonnetze, tori, spheres, due to the curvature of the rendering, gives reality to these objects.

To general public, adults with no specific knowledge of music or mathematics, the immersive impression is also astounding. It is usually the first time they can feel the

environment, adults have generally some questions about the technique we used while children are more inclined to let go and enjoy the performance.

Presenting/Teaching:

The observer, immersed in the virtual world, has no perception of the real one. It is then more challenging to explain or teach as we are used to practice with the help of a large screen. Moreover, unless we can operate in a real 4π Steradian theater, the observer follows its own tempo and looks in his own direction. Observers are seldom coordinated which each other's: a solution would be synchronizing headsets including the teacher's one.

Our primary tests, performed on few subjects corroborates the experiments already performed by different institutions, see for instance [17].

4.2   Future Developments

VR is a fast evolving field, we are implied in this evolution and will be able to present improved demonstrations for MCM 2022.

4.2.1   Include Other Models

We intend to employ these tools to visualize other famous models that where already embedded on the *Hypersphere* and presented during MCM 2015. For instance, mathemusical graphs, Cohn's *Regions* [18] and Douthett's *Cube Dance* [19].

We intend to create virtual images or movies for other eminent 3D and 4D models

– 3D:Mazzola *Torus of Thirds* [20], Amiot's *Torus of Phases* [21], Yust's Tonnetze and Zeitnetze [22] Chew's *Spiral* [24].
– 4D:Tymoczko's space as in the *Geometry of Music* [23].

4.2.2   Add Interactivity

We plan to make an interactive projection of the hyperspheres where the user can pick out his favorite projection from 4D to 3D. Some development is in progress with other artists to create interactive performances based on mathematics, music and graphical art.

Adding interactivity will enable the user to select elements, navigate on the model, or give directives to the software, supervising music changes consequently.

4.2.3   Go to Metaverse®

A space for Mathemusic is actually under construction in the meta universe, Metaverse by Meta® (Facebook®).

In this virtual space, we intend to host historical and actual mathemusical representations, other mathemusicians will be able to give me their models that we will integrate there, other artists will have some space to exhibit their work.

5 Conclusion

This paper covers our first experiences of embedding our hyperspaces models in a virtual environment.

We could establish new relations between the geometry and the music thanks to innovative perception of the models in hyperspaces, within an immersive environment. The encounters differ from the ones we were used to.

This work, using precalculated time, is a simple step that is to be considered in the continuity of our research on hyper-dimensional musical spaces in the last 20 years. We wish to continue collaboration with other fellow mathemusicians to explore new worlds and seek out new perceptions.

References

1. Baroin, G.: The planet-4D model: an original hypersymmetric music space based on graph theory. In: Agon, C., Andreatta, M., Assayag, G., Amiot, E., Bresson, J., Mandereau, J. (eds.) MCM 2011. LNCS (LNAI), vol. 6726, pp. 326–329. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21590-2_25

2. Baroin, G., de Gérando, S.: Sons et représentation visuelle en hyperespace: l'hypersphère des spectres, Les Cahiers de l'Institut International pour l'Innovation, la Création Artistique et la Recherche, 3icar éditions (2012)

3. Baroin, G., Calvet, A.: Visualizing temperaments: squaring the circle? In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O.A. (eds.) MCM 2019. LNCS (LNAI), vol. 11502, pp. 333–337. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_27

4. Baroin, G.: Music Mathematic and 4D, keynote during Virtuality Experience, Buenos Aires/Paris 04 December 2020. www.virtuality.io)

5. Sherman, W., et al.: Understanding Virtual Reality: Interface, Application, and Design, 2nd edn. Elsevier Science (2018)

6. Dohy, D., Mora-Camino, F., Mykoniatis, G., Raoul, J.-L.: Air traffic complexity through local covariance in the context of large areas of operations. In: 9th International Conference on Experiments/Process/System Modeling/Simulation/Optimization, Athens, Greece, July 2021 ⟨hal-03313081⟩ (2021)

7. Baroin, G.: Drone simulations in virtual environment. Neometsys NMS Lab. www.neometsys.fr

8. Baroin, G., Khannanov, I.: The Shadow-Tonnetz: visualizing speed and weight within harmonic progressions. In: Conference: 10th European Music Analysis Conference (Euromac 10), Moscow, Pre-Proceedings (2022)

9. Andretta, M.: On group-theoretical methods applied to music: some compositional and implementational aspects. Perspectives in Mathematical Music Theory (2004)

10. Seress, H., Baroin, G.: De l'Hypersphère au Spinnen Tonnetz: propositions d'adaptation pour les modèles triadiques. Musimédiane, no. 11 (2019). https://www.musimediane.com/11seressbaroin/

11. Andreatta, M., Baroin, G.: An introduction on formal and computational models in popular music analysis and generation. In: Kapoula, Z., Vernet, M. (eds.) Aesthetics and Neuroscience, pp. 257–269. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46233-2_16

12. Albini, G., Antonini, S.: Hamiltonian cycles in the topological dual of the Tonnetz. In: Chew, E., Childs, A., Chuan, C.-H. (eds.) MCM 2009. CCIS, vol. 38, pp. 1–10. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02394-1_1

13. Baroin, G., de Gérando, S.: The Entangled Hyperspheres, an innovative approach to visualize microtonal music. Les Cahiers de l'Institut International pour l'Innovation, la Création Artistique et la Recherche, icareditions (2022, to appear)
14. de Gérando, S.: Music from «le Labyrinthe du Temps», Performances since 2020
15. Jedrzejewski, F.: Ivan Wyschnegradsky et la musique microtonale. Musique, musicologie et arts de la scène. Université de Paris 1 Panthéon-Sorbonne (2000)
16. Bigo, L., de Gérando, S.: Invention algorithmique du matériau compositionnel: les séries tous intervalles dans une octave et en zigzag (STIOZ), les séries tous intervalles imbriqués dans une série micro-intervallique (STISMI), Rapport de recherche, icarEditions (2017)
17. Villena-Taranilla, R., Tirado-Olivares, S., Cózar-Gutiérrez, R., González-Calero, J.: Effects of virtual reality on learning outcomes in K-6 education: a meta-analysis. Educ. Res. Rev. **35**, 100434 (2022). https://doi.org/10.1016/j.edurev.2022.100434
18. Cohn, R.: Weitzmann's regions, my cycles, and douthett's dancing cubes. Music Theory Spectr. **22**(1), 89–103 (2000)
19. Douthett, J., Steinbach, P.: Parsimonious graphs: a study in parsimony contextual transformations, and modes of limited transposition. J. Music Theory **42**(2), 241–265 (1998)
20. Mazzola, G.: The topos of music: Birkhäuser Basel (2005)
21. Amiot, E.: The Torii of phases. In: Yust, J., Wild, J., Burgoyne, J.A. (eds.) MCM 2013. LNCS (LNAI), vol. 7937, pp. 1–18. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39357-0_1
22. Yust, J.: Generalized Tonnetze and Zeitnetze, and the topology of music concepts. J. Math. Music **14**, 170–203 (2020)
23. Tymoczko, D.: A Geometry of Music. Oxford Studies in Music Theory. Oxford University Press, Oxford (2011)
24. Chew, E.: The spiral array: an algorithm for determining key boundaries. In: Anagnostopoulou, C., Ferrand, M., Smaill, A. (eds.) ICMAI 2002. LNCS (LNAI), vol. 2445, pp. 18–31. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45722-4_4

SUM Classes and Quotient Generalized Interval Systems

David Orvek[1] and David Clampitt[2(✉)]

[1] Indiana University, Bloomington, IN, USA
`dorvek@iu.edu`
[2] Ohio State University, Columbus, OH, USA
`clampitt.4@osu.edu`

Abstract. The present paper develops algebraic properties of the SUM-class system first developed by Richard Cohn and explored by Robert Cook and Joseph Straus, in the context of David Lewin's Generalized Interval System (GIS) concept. Motivated by his observation that harmonic triads whose pitch classes sum to a given value modulo 12 share certain voice-leading properties, Cohn defined SUM classes for the 24 consonant (major and minor) triads, and defined transformations on these equivalence classes. We present the SUM-class system as a quotient GIS structure, and explore the dual quotient GIS implied by Lewin's theory for non-commutative GISs, and we generalize to other types of pitch-class sets (other set-classes).

Keywords: Generalized Interval System · Group homomorphism · Quotient group · SUM class

1 Introduction

In the context of consonant triads as pitch-class sets (i.e., major and minor triads, set-class 3-11), Cohn [1,2], and Cook [3] observed that the total pitch-class voice-leading interval between triads X and Y remains unchanged when either is transposed by a major third (4 semitones) in either direction (i.e., by T_4 or T_8). Cohn also noted that the pitch classes of major-third-related triads sum to the same value modulo 12. Generalizing to pitch-class sets of a given cardinality, the following definitions and proposition unify the voice-leading interval and SUM concepts and motivate the SUM-class definition. In this paper we assume the usual 12 pitch-class universe and all values are reduced modulo 12.

Definition 1. *Let* $X = \{x_1, \ldots, x_n\}$ *and* $Y = \{y_1, \ldots, y_n\}$ *be pitch-class sets of cardinality* $1 \leq n \leq 12$. *Let* $VL(X, Y) = \sum_{i=1}^{n}(y_i - x_i) \bmod 12$. *We call* $VL(X, Y)$ *the* voice-leading interval *from* X *to* Y.

While the sets in this definition are indexed, it turns out that the voice-leading interval does not depend on this ordering:

Definition 2. *Let* X *be a pitch-class set.* $SUM(X) = \sum_{x \in X}(x) \bmod 12$.

Proposition 1. *Let $card(X) = card(Y)$. $VL(X,Y) = SUM(Y) - SUM(X)$.*

Proof. $VL(X,Y) = \sum_{i=1}^{n}(y_i - x_i) = (y_1 - x_1) + \cdots + (y_n - x_n) \bmod 12 = (y_1 + \cdots + y_n) - (x_1 + \cdots + x_n) = SUM(Y) - SUM(X)$ (adapted from [1], 286).

Since the proposition shows that the voice-leading interval depends only on the SUMs, defined in terms of commutative mod 12 addition, the voice-leading interval from X to Y is a property of the unordered sets. It is evident that $VL(X,Y) = -VL(Y,X)$.

Motivated by the above observation, we define SUM classes. For a given n, let S_n be the set of all pitch-class sets of cardinality n. By Definition 2, for all sets $X \in S_n$ we have a function SUM: $S_n \mapsto \mathbb{Z}_{12} : X \mapsto \text{SUM}(X)$. The image of S_n under this function is a subset of $\mathbb{Z}_{12}$, $\text{im}(S_n)$. Define the SUM classes of S_n to be the inverse images of elements in $\text{im}(S_n)$. It follows that the SUM classes partition S_n and are the equivalence classes of the equivalence relation $\equiv_{S_n}$ where $X \equiv_{S_n} Y$ if and only if $\text{SUM}(X) = \text{SUM}(Y)$. This is the natural equivalence relation defined by the function SUM.

More often, one is interested in a refinement of this equivalence relation, where SUM is restricted to a given set-class (an orbit of the group T_n/I_n acting on pitch-class subsets). The set-classes partition the sets S_n, and SUM restricted to set-class α defines an equivalence $\equiv_\alpha$ on α. For example, if we take the canonical example of the 24 consonant triads, Forte set-class 3-11, the image of 3-11 under SUM is the set $\{1, 2, 4, 5, 7, 8, 10, 11\}$. The inverse images of these eight elements are the 3-11 SUM classes, which we label with square brackets, [1], [2], [4], [5], [7], [8], [10], [11]. For example, here SUM class [1] = {C-sharp minor, F minor, A minor}. E.g., SUM(F minor) = $0 + 5 + 8 = 13 \equiv 1 \bmod 12$.

2   SUM-Class Transformation Groups and Quotient Generalized Interval Systems

We recall the definition of the T_n/I_n group: $T_n : \mathbb{Z}_{12} \mapsto \mathbb{Z}_{12} : z \mapsto n + z$; $I_n : \mathbb{Z}_{12} \mapsto \mathbb{Z}_{12} : z \mapsto n - z$. The group product is composition of functions, and $T_n T_m = T_{n+m}, T_n I_m = I_{n+m}, I_m T_n = I_{m-n}, I_m I_n = T_{m-n}$. It follows that the defining relations in terms of generators are $(T_1)^{12} = T_0$, $(I_j)^2 = T_0$, and $T_j I_0 = I_0 T_{12-j}$; that is, the group is dihedral of order 24. Extending the definition of these operations on individual pitch classes to pitch-class sets, we have a group action on the power set $2^{\mathbb{Z}_{12}}$ of pitch-class subsets: for all subsets x, we have $T_0(x) = x$, and for all $f, g \in T_n/I_n, f(g(x)) = fg(x)$. The orbits are the set-classes, and therefore the group acts transitively on each set-class: if x, y are in a given set-class, there exists $f \in T_n/I_n$ such that $f(x) = y$; for set-classes with 24 elements (i.e., set-classes of sets of cardinality k, $2 < k < 10$, with only the trivial symmetry), the action is simply transitive: f mapping x to y is unique.

From Lewin [4], 157–158, a simply transitive action of a group G on a set S is equivalent to a Generalized Interval System (GIS). The triadic GIS for $S = $ 3-11, $G = T_n/I_n$ is well known. We seek a GIS structure for the triadic

SUM-classes. Recalling Cohn's observation that voice-leading intervals between triads are invariant under pitch-class transpositions by 4 semitones, consider the subgroup $H = \{T_0, T_4, T_8\}$. H is a normal subgroup of $G = T_n/I_n$: since H commutes with all the transpositions, all that is needed to check is that for all k, $I_k(H) = (H)I_k$. $(H)I_k = (\{T_0, T_4, T_8\})I_k = \{I_k, I_{k+4}, I_{k+8}\}$ and $I_k(H) = I_k(\{T_0, T_4, T_8\}) = \{I_k, I_{k-4}, I_{k-8}\} = \{I_k, I_{k+8}, I_{k+4}\}$. Thus, left and right cosets coincide, and H is normal in G. By the principal theorems of group theory, the cosets form the quotient group $\frac{G}{H}$, isomorphic with the image of G under a homomorphism h with kernel H. The group product for $\frac{G}{H}$ is inherited from the parent group $G = T_n/I_n$ in the natural way (demonstrated below). Since for finite G $o(\frac{G}{H}) = o(G)/o(H)$, the order of the quotient group here is $24/3 = 8$. The equivalence relation $\equiv_{3-11}$ partitioned the 24 harmonic triads into 8 SUM-classes, each with 3 members, and we assert that the quotient group and the set $\mathcal{S}$ of SUM classes form a GIS, a quotient GIS of the original (defined below).

We reduce the $\equiv_{3-11}$ equivalence classes to the respective sums x, symbolized $[x]$, but for present purposes in the triadic case it is useful to tabulate them (writing lower-case letters for minor, upper-case for major): $[1] = \{c\sharp,\ f,\ a\}$, $[2] = \{C\sharp,\ F,\ A\}$, $[4] = \{d,\ f\sharp,\ b\flat\}$, $[5] = \{D,\ F\sharp,\ B\flat\}$, $[7] = \{e\flat,\ g,\ b\}$, $[8] = \{E\flat, G,\ B\}$, $[10] = \{e,\ g\sharp,\ c\}$, $[11] = \{E,\ A\flat,\ C\}$.

We similarly tabulate and name the elements of $\frac{G}{H}$, that is, the cosets of normal subgroup $\{T_0, T_4, T_8\}$ in T_n/I_n. To simplify the typography we name the cosets $\mathcal{T}_j = T_j H$, and $\mathcal{I}_j = I_j H, j = 0, 1, 2, 3$.

$$\mathcal{T}_0 = T_0 H = \{T_0, T_4, T_8\}$$

$$\mathcal{T}_1 = T_1 H = \{T_1, T_5, T_9\}$$

$$\mathcal{T}_2 = T_2 H = \{T_2, T_6, T_{10}\}$$

$$\mathcal{T}_3 = T_3 H = \{T_3, T_7, T_{11}\}$$

$$\mathcal{I}_0 = I_0 H = \{I_0, I_4, I_8\}$$

$$\mathcal{I}_1 = I_1 H = \{I_1, I_5, I_9\}$$

$$\mathcal{I}_2 = I_2 H = \{I_2, I_6, I_{10}\}$$

$$\mathcal{I}_3 = I_3 H = \{I_3, I_7, I_{11}\}$$

The general theory already tells us that $\frac{G}{H}$ is a group with product induced from the definition of operations in T_n/I_n, but let us explicitly state the composition rules. Given $\mathcal{T}_i, \mathcal{T}_j, \mathcal{I}_k, \mathcal{I}_l \in \frac{G}{H}$, we have $\mathcal{T}_i \mathcal{T}_j = \mathcal{T}_{(i+j) \bmod 4}, \mathcal{T}_i \mathcal{I}_k = \mathcal{I}_{(i+k) \bmod 4}, \mathcal{I}_k \mathcal{T}_i = \mathcal{I}_{(k-i) \bmod 4}, \mathcal{I}_k \mathcal{I}_l = \mathcal{T}_{(k-l) \bmod 4}$. Thus, we have relations $(\mathcal{T}_1)^4 = \mathcal{T}_0, (\mathcal{I}_j)^2 = \mathcal{T}_0$, and $\mathcal{T}_j \mathcal{I}_0 = \mathcal{I}_0 \mathcal{T}_{4-j}$, that is, $\frac{G}{H}$ is dihedral of order 8 (the homomorphic image of a dihedral group).

Next we tabulate the group action of $\frac{G}{H}$ on the set $\mathcal{S}$ of $\equiv_{3-11}$ SUM classes. Each element defines a permutation of $\mathcal{S}$, expressed as a product of cycles:

$\mathcal{T}_0 : \mathcal{S} \to \mathcal{S} : (\ )$

$\mathcal{T}_1 : \mathcal{S} \to \mathcal{S} : ([1]\ [4]\ [7]\ [10])\ ([2]\ [5]\ [8]\ [11])$

$\mathcal{T}_2 : \mathcal{S} \to \mathcal{S} : ([1]\ [7])\ ([4]\ [10])\ ([2]\ [8])\ ([5]\ [11])$

$\mathcal{T}_3 : \mathcal{S} \to \mathcal{S} : ([1]\ [10]\ [7]\ [4])\ ([2]\ [11]\ [8]\ [5])$

$\mathcal{I}_0 : \mathcal{S} \to \mathcal{S} : ([1]\ [11])\ ([2]\ [10])\ ([4]\ [8])\ ([5]\ [7])$

$\mathcal{I}_1 : \mathcal{S} \to \mathcal{S} : ([1]\ [2])\ ([4]\ [11])\ ([5]\ [10])\ ([7]\ [8])$

$\mathcal{I}_2 : \mathcal{S} \to \mathcal{S} : ([1]\ [5])\ ([2]\ [4])\ ([7]\ [11])\ ([8]\ [10])$

$\mathcal{I}_3 : \mathcal{S} \to \mathcal{S} : ([1]\ [8])\ ([2]\ [7])\ ([4]\ [5])\ ([10]\ [11])$

This tabulation shows that the group action of $\frac{G}{H}$ on $\mathcal{S}$ is simply transitive: for every ordered pair of elements (s_1, s_2) in $\mathcal{S} \times \mathcal{S}$, there exists exactly one $g \in \frac{G}{H}$ such that $g(s_1) = s_2$. As noted above, a simply transitive group action on a set S determines a GIS, so we have $\text{GIS}_1 = (\frac{G}{H}, \mathcal{S})$. It will later be significant to note that $\frac{G}{H}$ is non-commutative (non-abelian), so GIS_1 is a *non-commutative GIS*. This procedure holds generally, as demonstrated below, permitting the inclusion of GIS_1 in the class of *quotient GISs*.

Proposition 2. *Let $(\mathcal{G}, S)$ be a GIS, with $\mathcal{H}$ a normal subgroup in $\mathcal{G}$. Then the quotient group $\frac{\mathcal{G}}{\mathcal{H}}$ induces a set S' of equivalence classes on S, such that $\frac{\mathcal{G}}{\mathcal{H}}$ acts simply transitively on S'. That is, $(\frac{\mathcal{G}}{\mathcal{H}}, S')$ is a GIS, and one may call it a quotient GIS.*

Lemma 1. *The quotient group $\frac{\mathcal{G}}{\mathcal{H}}$ induces a set S' of equivalence classes on S.*

Proof. Consider the restriction to $\mathcal{H}$ of the action of $\mathcal{G}$ on S, $\mathcal{H}(S) = \{h(s)|h \in \mathcal{H}, s \in S\}$. For all $s \in S$, let $\sigma_s = \mathcal{H}(s)$. Let $S' = \{\sigma_s | \forall s \in S\}$. S' defines a partition of S: All elements of S belong to some σ_s, since $e \in \mathcal{H}, e(s) = s \in \sigma_s$. Suppose we have $s_1 \in \sigma_{s_1}$ and $s_1 \in \sigma_{s_2}$. Then for some $h \in \mathcal{H}, s_1 = h(s_2)$. Let t be any element in σ_{s_1}, then for some $h_1 \in \mathcal{H}$, $t = h_1(s_1) = h_1 h(s_2)$, and $t \in \sigma_{s_2}$. Thus $\sigma_{s_1} = \sigma_{s_2}$. Suppose $s_1 \notin \sigma_{s_2}$, and $t \in \sigma_{s_1} \cap \sigma_{s_2}$. Then there exist $h_1, h_2 \in \mathcal{H}$ such that $h_1(s_1) = t = h_2(s_2)$. Then $h_1^{-1} h_2(s_2) = s_1$, so $s_1 \in \sigma_{s_2}$, contradiction, and σ_{s_1} and σ_{s_2} are disjoint. Thus S' partitions S, and the members of S' are equivalence classes of the equivalence relation defined by the partition.

Lemma 2. $\frac{\mathcal{G}}{\mathcal{H}}$ *acts simply transitively on S'.*

Proof. Let $x\mathcal{H}, y\mathcal{H}$ be distinct cosets $\in \frac{\mathcal{G}}{\mathcal{H}}$ (i.e., $y^{-1}x \notin \mathcal{H}$), and $\sigma_s \in S'$. For any $\sigma_s \in S', \mathcal{H}(\sigma_s) = \mathcal{H}(\mathcal{H}(s)) = \mathcal{H}\mathcal{H}(s) = \mathcal{H}(s) = \sigma_s$, by definition, and

$x\mathcal{H}(y\mathcal{H}(\sigma_s)) = x\mathcal{H}y\mathcal{H}(\sigma_s)$, so $\frac{\mathcal{G}}{\mathcal{H}}$ acts transitively on S'. Suppose $x\mathcal{H}(\sigma_s) = y\mathcal{H}(\sigma_s)$. Then $(y\mathcal{H})^{-1}x\mathcal{H}(\sigma_s) = \sigma_s$, but also $(y\mathcal{H})^{-1}x\mathcal{H}(\sigma_s) = y^{-1}\mathcal{H}x\mathcal{H}(\sigma_s) = y^{-1}x\mathcal{H}\mathcal{H}(\sigma_s) = y^{-1}x\mathcal{H}(\sigma_s) = y^{-1}x(\sigma_s) = \sigma_s$, implies $y^{-1}x \in \mathcal{H}$, contrary to assumption. For every $s, t \in S'$, there exists $x\mathcal{H} \in \frac{\mathcal{G}}{\mathcal{H}}$ such that $x\mathcal{H}(s) = t$ (by definition of $\mathcal{S}'$), and $x\mathcal{H}$ is unique (by the above demonstration). Thus, $\frac{\mathcal{G}}{\mathcal{H}}$ acts simply transitively on S'.

$(\frac{\mathcal{G}}{\mathcal{H}}, S')$ is therefore a (quotient) GIS.

From Proposition 1 and the definition of the SUM classes, the corollary follows that the voice-leading interval from any member of $[x]$ to any member of $[y]$ is $y - x$ (modulo 12). Slightly abusing the notation, we have $VL([x], [y]) = y - x$. Recall that the voice-leading intervals are not the intervals of Generalized Interval System GIS$_1$—those are defined by the group action. From Lewin [4], 157, the interval i from $[x]$ to $[y]$, $i = \mathrm{int}([x], [y])$, is identified with the unique $g \in \frac{G}{H}$ such that $g([x]) = [y]$, whereas by Definition 1 a voice-leading interval is an integer modulo 12. We are, however, concerned here with voice-leading intervals, the motivation for $\mathcal{S}$, the set of SUM classes.

For a given GIS (G, S), Lewin ([4], 48) defines an *interval-preserving transformation* to be a mapping $X : S \to S$ such that for all $s, t \in S$, $\mathrm{int}(X(s), X(t)) = \mathrm{int}(s, t)$. Here, *int* refers to the interval function of the GIS just discussed. Let's extend his definition to voice-leading intervals:

Definition 3. *If for all SUM classes* $[x], [y] \in \mathcal{S}$ *a transformation* X *on* $\mathcal{S}$ *has the property* $VL(X([x]), X([y])) = VL([x], [y])$, *then* X *is said to be a voice-leading-interval-preserving transformation (VL-preserving transformation).*

Since $VL([x], [y]) = y - x$, X is a VL-preserving transformation if and only if $VL(X([x]), X([y])) = y - x$. The elements of $\frac{G}{H}$ are transformations of $\mathcal{S}$; are they VL-preserving?

Proposition 3. *The elements of the* $\frac{G}{H}$ *subgroup* $\mathcal{T} = \{\mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3\}$ *are VL-preserving.*

Proof. For $j = 0, 1, 2, 3$, $\mathcal{T}_j : \mathcal{S} \mapsto \mathcal{S} : [x] \mapsto [(x + 3j) \bmod 12]$. Then for all $[x], [y] \in \mathcal{S}$, $VL(\mathcal{T}_j([x]), \mathcal{T}_j([y])) = VL([(x + 3j) \bmod 12], [(y + 3j) \bmod 12]) = ((y + 3j) - (x + 3j)) \bmod 12 = (y - x) \bmod 12 = VL([x], [y])$. Therefore, the $\mathcal{T}_j$ are VL-preserving.

The elements of the coset of $\mathcal{T}$ as a (normal) subgroup of $\frac{G}{H}$, $\mathcal{I} = \{\mathcal{I}_0, \mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3\}$, are not VL-preserving; rather, they are VL-reversing.

Definition 4. *If for all SUM classes* $[x], [y] \in \mathcal{S}$ *a transformation* X *on* $\mathcal{S}$ *has the property* $VL(X([x]), X([y])) = VL([y], [x])$, *then* X *is said to be a voice-leading-interval-reversing transformation (VL-reversing transformation).*

Since $VL([x], [y]) = y - x, VL([y], [x]) = x - y$, and X is a VL-reversing transformation if and only $VL(X([x]), X([y])) = x - y$.

Proposition 4. *The elements of $\mathcal{I}$ are VL-reversing.*

Proof. For $j = 0, 1, 2, 3$, $\mathcal{I}_j : \mathcal{S} \mapsto \mathcal{S} : [x] \mapsto [(3j - x) \bmod 12]$. Then for all $[x], [y] \in \mathcal{S}$, $VL(\mathcal{I}_j([x]), \mathcal{I}_j([y])) = VL([(3j - x) \bmod 12], [(3j - y) \bmod 12]) = ((3j - y) - (3j - x)) \bmod 12 = (x - y) \bmod 12 = VL([y], [x])$. Therefore, the $\mathcal{I}_j$ are VL-reversing.

3   The Dual Quotient Generalized Interval System

Those familiar with Lewin's theory know that every non-commutative GIS (G, S) has its dual, with a different group G', isomorphic to G but with a distinct simply transitive action on S. The groups G and G' are subgroups of the group of all permutations of S, the symmetric group $Sym(S)$, so elements of the two groups compose with each other within $Sym(S)$. As Lewin demonstrates, the groups are each other's commuting groups; as subgroups of $Sym(S)$, each other's centralizer subgroup: every element of G commutes with every element of G'. It follows that they play the role of each other's group of interval-preserving transformations, as seen in the commutative diagram in Fig. 1. Let f belong to G and f' belong to G' and let s_1, s_2, s_3, s_4 be members of S. The diagram shows that f and f' commute: $f'f(s_1) = f'(s_2) = s_4$ and $ff'(s_1) = f(s_3) = s_4$. But this is equivalent to f' preserving the intervals in (G, S) and f preserving the intervals in (G', S): $\text{int}(s_1, s_2) = \text{int}(s_3, s_4)$ in (G, S) since $f(s_1) = s_2$ and $f(s_3) = s_4$, so in (G, S) $\text{int}(f'(s_1), f'(s_2)) = \text{int}(s_1, s_2)$, and thus f' satisfies Lewin's definition of an interval-preserving transformation. The same demonstration shows that in (G', S), $\text{int}'(f(s_1), f(s_3)) = \text{int}'(s_1, s_3)$; f is an interval-preserving transformation of (G', S).

$$
\begin{array}{ccc}
 & f & \\
s_1 & \longrightarrow & s_2 \\
f' \downarrow & & \downarrow f' \\
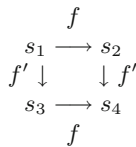s_3 & \longrightarrow & s_4 \\
 & f &
\end{array}
$$

Fig. 1. Commutativity is equivalent to interval preservation

So much is known, but we are interested here also in voice-leading intervals, and in the dual of a (non-commutative) quotient GIS. It is known that the dual to the GIS $(T_n/I_n, \text{3-11})$ is the GIS of the action of the full neo-Riemannian group on 3-11 ([5], 180–181). This group is often referred to as the *P-L-R* group, because it is generated by the three contextual inversions: P, which sends a triad to its Parallel, as C to c, c to C; L, which sends a triad to its *Leittonwechsel*, as C to e, e to C; and R, which sends a triad to its Relative, as C to a, a to C (see Cohn, [6]). One of these generators is redundant: composing RL and taking it as one generator, of order 12, and taking one of the involutions as the other

generator, we have a presentation of the group that shows it to be isomorphic to the order 24 dihedral group, that is, to the T_n/I_n group: $(RL)^{12} = e, L^2 = e$ (where e is the identity operation), and $(RL)^j L = L(RL)^{12-j}$. We may therefore refer to this as the R/L group, and to the $(R/L, 3\text{-}11)$ GIS. In this notation, the parallel structures with the T_n/I_n and its subgroups are made evident. Music theorists, who compute relatively easily with P, L, and R, conventionally express these transformations as compositions of them. In all but one case the minimal compositions are of length at most 4.

The distinguishing character of the three neo-Riemannian involutions is the parsimonious voice-leading they entail: all three preserve two common tones, the remaining tone moving by semitone under P and under L and by two semitones under R. One might imagine that this GIS is therefore well-suited to addressing voice-leading considerations, and indeed Cohn [1] treats SUM classes from this side of the duality (and moreover dives right into the quotient group). See also Cook [3], chapter 2. Since this is a true duality, one may logically begin from either side of it. The path chosen here is heuristic, considering that the T_n/I_n is perhaps better known and more convenient computationally.

For notational simplicity, set $Q = RL$: Q transposes major triads by T_7, minor triads by T_5. (NB: Music theorists often use right functional orthography in neo-Riemannian contexts. Since operations from the dual GISs will be composed, left orthography will be followed here.) The construction of the dual quotient group and dual quotient GIS_2 to $\text{GIS}_1 = (\frac{G}{H}, \mathcal{S})$ again begins with an order 3 cyclic subgroup $H' = \{Q^0, Q^4, Q^8\}$. H' is normal in $G' = R/L$, so there exists the quotient group $\frac{G'}{H'}$, of order 8. The group mirrors its dual, reflected in the notation for the cosets, which are expressed in terms of elements Q^j or $Q^j L$, for $j = 0, 1, 2, 3$, and as compositions of P, L, R. Following Cohn [1], we use $\mathcal{X}$ (for eXchange) for the sets of contextual inversions, $\mathcal{Y}$ for the others.

$$\mathcal{Y}_0 = \{Q^0, Q^4, Q^8\} = \{E, PL, LP\}$$

$$\mathcal{Y}_1 = \{Q^1, Q^5, Q^9\} = \{RL, RP, RPLP\}$$

$$\mathcal{Y}_2 = \{Q^2, Q^6, Q^{10}\} = \{RPRL, (RP)^2, LRPR\}$$

$$\mathcal{Y}_3 = \{Q^3, Q^7, Q^{11}\} = \{PLPR, LR, PR\}$$

$$\mathcal{X}_0 = \{L, Q^4 L, Q^8 L\} = \{L, P, PLP\}$$

$$\mathcal{X}_1 = \{QL, Q^5 L, Q^9 L\} = \{R, PLR, LPR\}$$

$$\mathcal{X}_2 = \{Q^2 L, Q^6 L, Q^{10} L\} = \{RLR, RPR, RPLPR\}$$

$$\mathcal{X}_3 = \{Q^3 L, Q^7 L, Q^{11} L\} = \{PRL, PRP, LRL\}$$

The composition rules are the same as in $\frac{G}{H}$, replacing $\mathcal{T}_j$ by $\mathcal{Y}_j$ and $\mathcal{I}_k$ by $\mathcal{X}_k$. The quotient groups are thus isomorphic as groups, but the action of $\frac{G'}{H'}$ on $\mathcal{S}$ is very different. It is tabulated below, again operations as products of cycles.

$\mathcal{Y}_0 : \mathcal{S} \mapsto \mathcal{S} : (\ )$

$\mathcal{Y}_1 : \mathcal{S} \mapsto \mathcal{S} : ([1]\ [4]\ [7]\ [10])\ ([2]\ [11]\ [8]\ [5])$

$\mathcal{Y}_2 : \mathcal{S} \mapsto \mathcal{S} : ([1]\ [7])([2]\ [8])\ ([4]\ [10])\ ([5]\ [11])$

$\mathcal{Y}_3 : \mathcal{S} \mapsto \mathcal{S} : ([1]\ [10]\ [7]\ [4])\ ([2]\ [5]\ [8]\ [11])$

$\mathcal{X}_0 : \mathcal{S} \mapsto \mathcal{S} : ([1]\ [2])\ ([4]\ [5])\ ([7]\ [8])\ ([10]\ [11])$

$\mathcal{X}_1 : \mathcal{S} \mapsto \mathcal{S} : ([1]\ [11])\ ([2]\ [4])\ ([5]\ [7])\ ([8]\ [10])$

$\mathcal{X}_2 : \mathcal{S} \mapsto \mathcal{S} : ([1])\ [8])\ ([2]\ [7])\ ([4]\ [11])\ ([5]\ [10])$

$\mathcal{X}_3 : \mathcal{S} \mapsto \mathcal{S} : ([1]\ [5])\ ([2]\ [10])\ ([11]\ [7])\ ([8]\ [4])$

Comparing the actions on $\mathcal{S}$ of the respective dual groups, one sees that, of course, the identity operations are the same—$\mathcal{Y}_0 = \mathcal{T}_0 = 1_\mathcal{S}$—and also $\mathcal{Y}_2 = \mathcal{T}_2$. These two operations commute with everything in both GIS$_1$ and GIS$_2$ (central in both quotient groups), so are interval-preserving transformations (moreover, operations) in both. $\mathcal{Y}_0$ and $\mathcal{Y}_2$ are also both obviously VL-preserving in GIS$_2$, but no other elements in $\frac{G'}{H'}$ are VL-preserving, as inspection of the tabulation makes clear. In GIS$_1$ the subgroup $\mathcal{T}$ was VL-preserving. In GIS$_2$, in the analogous subgroup $\mathcal{Y}$, the operations $\mathcal{Y}_1$ and $\mathcal{Y}_3$ are not VL-preserving. In the coset of $\mathcal{Y}$ the eXchange operations do interact cogently with the voice-leading intervals defined by VL, even though none are VL-preserving: each exchanges pairs $[x], [y]$ such that for a fixed n, VL$([x], [y]) = \pm n \mod 12$. The tabulation shows that $\mathcal{X}_0$ exchanges by ± 1, $\mathcal{X}_3$ by ± 4, $\mathcal{X}_2$ by ± 7, and $\mathcal{X}_1$ by ± 10. This fact motivated Cohn's notation for his sum-class transformation group, X_1, X_4, X_7, X_{10}. His Y/X group was not explicitly defined as a quotient structure, but is precisely the same group as the one derived here: $\mathcal{Y}_j = Y_{3j}, j = 0, 1, 2, 3, \mathcal{X}_0 = X_1, \mathcal{X}_1 = X_{10}, \mathcal{X}_2 = X_7, \mathcal{X}_3 = X_4$.

In [1], Cohn thematizes his X_1 and X_{10} operations because, just as the four-fold repetition of the alternation of R with P sends any given triad through an octatonic cycle, the four-fold repetition of the alternation of X_1 with X_{10} sends the SUM classes in a cycle matching an octatonic scale: $X_1([1]) = [2], X_{10}([2]) = [4], X_1([4]) = [5]$, etc. More significantly, the union of each pair of SUM classes related by X_1 forms a hexatonic region, the four pairs cover the four regions, and the subgroup formed by the union of the cosets associated with the identity Y_0 and with X_1 is the hexatonic subgroup, $\{E, PL, LP, P, L, PLP\}$. The union of each pair of SUM classes related by X_{10} forms a Weitzmann region, the four pairs cover the four regions, and the union of the cosets associated with the identity

Y_0 and with X_{10} is the Weitzmann subgroup $\{E, PL, LP, R, LPR, PLR\}$ (cf. [2,7]). In [2], Cohn refers to "voice-leading zones" in lieu of SUM classes, and presents many musical analyses employing them.

Applying Lewin's general theory, all the elements of $\frac{G'}{H'}$ commute with all of those of its dual $\frac{G}{H}$, all are interval-preserving transformations for GIS_1, and vice versa for elements of $\frac{G}{H}$ with respect to GIS_2. In [8], Orvek presents an analysis of passages from Charles Villiers Stanford's setting of Keats's poem *La belle dame sans merci*, employing a commutative diagram that relates two chromatic sequences. The sequences appear during a description of the protagonist's dream. The reduction to SUM classes employs a VL-preserving $\mathcal{T}$ transformation from GIS_1 commuting with an $\mathcal{X}$ transformation from GIS_2 that always exchanges triads at a fixed VL-interval distance. The reduction is shown in Fig. 2, adapted from ([8], fig. 2.4). Orvek presents a similar analysis, with set-class 3-3, of music from *Nacht* in Schoenberg's *Pierrot Lunaire*, op. 21. See also Cook's triadic analyses of music by César Franck [3].
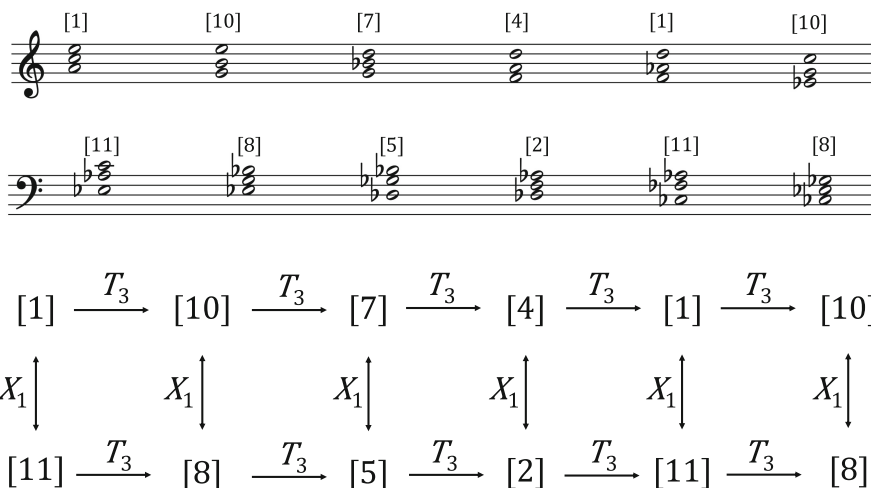

Fig. 2. A commutative transformational network relating mm. 98–108 (top) and mm. 122–128 (bottom) from Charles V. Stanford, *La belle dame sans merci*

4   Other Set-Classes

At the outset of this paper, the SUM-class equivalence relation was defined very generally, and then stated at the level of the asymmetric pitch-class set-classes, those with 24 members, and that is the level of generality at which we continue. The question is: is there always a quotient group that acts simply transitively on the SUM classes for a given asymmetric pitch-class set? The answer, sketched

below, is yes, but it is not always a quotient of the T_n/I_n group. The problem is that, as we will see, some groups acting simply transitively on the SUM classes must be cyclic (generated by a single element). The quotient group of a dihedral group is the homomorphic image of that group, so is itself dihedral (except in the trivial case where the whole group is mapped to the group with one element, which only applies to the empty set-class and the aggregate $\mathbb{Z}_{12}$, where [0] or [6], respectively, is the only SUM class). We are also not interested in the other trivial case, the quotient by the identity subgroup, because there are at most 12 SUM classes. There are three normal subgroups of T_n/I_n of order 12, the cyclic subgroup of twelve transpositions, and two dihedral subgroups that contain the 6 even transpositions, one with the 6 inversions of even index, the other with the 6 inversions of odd index. In general, any subgroup H such that $o(H) = \frac{1}{2}o(G)$ is normal: you're either on the bus or off the bus, the bus here being H; there are just 2 cosets, H and $G \setminus H$, and $\forall g \in G, gH = Hg$, since $gH = H$ if and only if $g \in H$, (if and only if $g \notin H, gH = G \setminus H$). The quotient of T_n/I_n by any of the three subgroups is therefore a subgroup of order 2, potentially applicable to a set-class with just two SUM classes.

The cyclic subgroups are normal, by the earlier argument for $\{T_0, T_4, T_8\}$. We omit the demonstration, but the only dihedral subgroups that are normal are the two of index 2 discussed above.

Orvek [8] has shown that in the cases of all but one asymmetric trichordal set-class, the situation is the same as for 3-11, and the order 3 cyclic subgroup $\{T_0, T_4, T_8\}$ effects the appropriate equivalence relations on the parent group and parent set-class. This is because transposing a trichord by any multiple n of 4 semitones adds $3 \times 4n = 12n \equiv 0 \bmod 12$ to the sum of the constituent pitch classes, thus leaves SUM classes fixed. The exception is the 3-4: (015) set class, which can be seen to admit only SUM classes [0], [3], [6], [9]. (Straus [9] refers to the exceptions as *maverick sets.*) The only order 4 quotient group is modulo the only order 6 normal subgroup, the transpositions of even index. This non-cyclic (dihedral) four-group has no element that holds SUM classes fixed, thus fails.

The appropriate simply transitive group on 3-4 from which to construct a quotient that succeeds is an abelian group T_n/J of order 24, isomorphic to $\mathbb{Z}_{12} \times \mathbb{Z}_2$, with the usual order 12 subgroup of transpositions and one contextual inversion that commutes with all the transpositions. The appropriate inversion J must hold all four SUM classes globally fixed, since it must be in the subgroup which is the identity element in the quotient subgroup. There are three possible choices for J; whichever choice is made, the other two appear as T_4J and T_8J. We may define the contextual inversion to be J such that it inverts about the pitch-class that is the endpoint of the semitone dyad that is 5 semitones away from the isolated pitch-class: thus J exchanges $\{0, 1, 5\} \longleftrightarrow \{7, 11, 0\}$. J is an involution by definition. It holds SUM classes fixed: a general element of 3-4 in prime form is $\{x, x+1, x+5\}, x \in \mathbb{Z}_{12}$, in SUM class $3x+6$; $J(\{x, x+1, x+5\}) = \{x, x-1, x-5\}$, and $x + (x-1) + (x-5) = x + (x+11) + (x+7) = 3x + 18 \equiv 3x + 6 \bmod 12$. By symmetry, involution J equally preserves the SUM class of inverted forms in 3-4.

For all n, T_n commutes with J : $T_n J(\{z, z\pm 1, z\pm 5\}) = T_n(\{z, z\mp 1, z\mp 5\}) = \{n+z, n+z\mp 1, n+z\mp 5\})$, and $JT_n(\{z, z\pm 1, z\pm 5\}) = J(\{n+z, n+z\pm 1, n+z\pm 5\}) = \{n+z, n+z\mp 1, n+z\mp 5\}$. It follows that one may define $T_n J := J_n$. Then $T_0 J = J_0$, and $(J_0)^2 = T_0$. $T_n J_m = T_n(T_m J) = T_{m+n} J = J_{m+n}$; $J_m J_n = T_m J T_n J = T_m T_n J J = T_m T_n = T_{m+n}$. Since T_n/J is abelian, all its subgroups are normal.

The quotient of Tn/J by normal subgroup $H = \{T_0, T_4, T_8, J_0, J_4, J_8\}$ is isomorphic to group $\mathbb{Z}_4$, gives rise to the 3-4 SUM classes, and acts simply transitively on them. $\frac{T_n/J}{H} = \{H, T_1 H, T_2 H, T_3 H\}$ and the mapping $i : \frac{T_n/J}{H} \leftrightarrow \mathbb{Z}_4 : T_z H \leftrightarrow z, z = 0, 1, 2, 3$ is clearly an isomorphism.

The 3-4 SUM classes are:

$[0] = \{\{2, 3, 7\}, \{6, 7, 11\}, \{10, 11, 3\}, \{9, 1, 2\}, \{1, 5, 6\}, \{5, 9, 10\}\}$
$[3] = \{\{3, 4, 8\}, \{7, 8, 0\}, \{11, 0, 4\}, \{10, 2, 3\}, \{2, 6, 7\}, \{6, 10, 11\}\}$
$[6] = \{\{0, 1, 5\}, \{4, 5, 9\}, \{8, 9, 1\}, \{7, 11, 0\}, \{11, 3, 4\}, \{3, 7, 8\}\}$
$[9] = \{\{1, 2, 6\}, \{5, 6, 10\}, \{9, 10, 2\}, \{8, 0, 1\}, \{0, 4, 5\}, \{4, 8, 9\}\}$.

The simply transitive action of $\frac{T_n/J}{H}$ on $S = \{[0], [3], [6], [9]\}$ is:

$H : S \mapsto S : (\quad)$
$T_1 H : S \mapsto S : ([0][3][6][9])$
$T_2 H : S \mapsto S : ([0][6])([3][9])$
$T_3 H : S \mapsto S : ([0][9][6][3])$.

Most (ten) asymmetric tetrachordal set-classes admit quotients by the cyclic subgroup $\{T_0, T_3, T_6, T_9\}$, because transposing a tetrachord by any multiple n of 3 semitones adds $4 \times 3n = 12n \equiv 0 \bmod 12$ to the sum of the constituent pitch classes, so leaves SUM classes fixed. But there are four examples (mavericks), identified in [8], where the quotient group must be cyclic, of order 3 (4-4, 4-14, 4-18: SUM classes $\{[0], [4], [8]\}$; 4-13: $\{[2], [6], [10]\}$). It is left as an exercise for the reader to carry out the case of the 4-4: (0125) set-class. In this case one may take the contextual inversion J to be inversion about the isolated pitch class, as in $J : \{0, 1, 2, 5\} \leftrightarrow \{5, 8, 9, 10\}$, and the normal subgroup is $\{T_0, T_3, T_6, T_9, J_0, J_3, J_6, J_9\}$.

In the case of pentachords, all asymmetric pentachords admit all 12 possible sum-classes, (since 5 and 12 are coprime). An example, again left to the reader, is 5-4: (01236), the pentachord well known from the analysis by Lewin of Stockhausen's *Klavierstück III* in [11]. In this case a commutative GIS isomorphic to that employed by Lewin suffices, with a contextual inversion J about a cluster endpoint furthest from the isolated pitch class, analogous to that for the 3-4 case, and a quotient of order 12 by the subgroup $\{T_0, J\}$.

For the hexachords, transposing by any multiple of 2 semitones adds $6 \times 2n = 12n \equiv 0 \bmod 12$ to the sum of the constituent pitch classes, thus the subgroup $H = \{T_0, T_2, T_4, T_6, T_8, T_{10}\}$ leaves SUM classes fixed. The quotient fourgroup acts simply transitively on the SUM classes of all the asymmetric hexachords except for 6-9, 6-16 (SUM classes [0], [6]) and 6-14, 6-22 (SUM classes

[3], [9]). These exceptional cases may be realized as quotients of T_n/I_n by normal subgroups of order 12: for 6-9 (012357), 6-14 (013458), 6-22 (012468) we take $H = \{T_0, T_2, T_6, T_8, T_{10}, I_1, I_3, I_5, I_7, I_9, I_{11}\}$, and the quotient group is $\{H, T_1 H\}$. For 6-9, the action of the quotient group on the SUM-classes S is $H : S \mapsto S : (\ \ )$; $T_1 H : S \mapsto S : ([0][6])$. For 6-14 and 6-22, the action is $H : S \mapsto S : (\ \ )$; $T_1 H : S \mapsto S : ([3][9])$. For 6-16 (014568), we take $H = \{T_0, T_2, T_4, T_6, T_8, T_{10}, I_0, I_2, I_4, I_6, I_8, I_{10}\}$, and the quotient group is $\{H, T_1 H\}$, with action $H : S \mapsto S : (\ \ )$; $T_1 H : S \mapsto S : ([0][6])$.

The treatment of the exceptional cases may be proven generally, including the commutativity of the ordinary transpositions with the appropriate contextual inversion, with recourse to Kochavi's study [10], in which he defines contextual inversions with respect to an indexing function for the members of the set class. Since both the parent GIS and the quotient GIS are commutative, there is no dual involved in the maverick cases.

The set-classes of larger cardinality than 6 may be treated in the same way as their complements with respect to SUM classes, quotient groups, and quotient GISs. In the case of a symmetrical set-class, the group acting on its SUM classes is a cyclic group, but the parent GIS already consists of a cyclic group (the T_n group or one of its subgroups) acting on the set-class. The quotient group falls out easily in such cases (see [8]). For example, the GIS for the augmented triad, set-class 3-12: (048), requires only the cyclic group $\{T_0, T_3, T_6, T_9\}$ acting simply transitively on the four members. There are just the four SUM classes, [0], [3], [6], [9], so the required quotient group for the quotient GIS is the trivial one by the identity subgroup $\{T_0\}$. Similarly, the GIS for the usual diatonic scale, 7-35: (013568t), requires the cyclic order 12 T_n group acting simply transitively on the 12 members of the inversionally symmetric set-class. The usual diatonic set admits the maximum 12 SUM classes, so again, the trivial quotient of T_n by its identity subgroup suffices for the quotient GIS.

The focus in this paper has been on mod 12 pitch-class sets, the groups that define their set-classes, and their quotients, because of the musical application. Most of these results could be extended to any dihedral group. It may be recalled that Proposition 2 is completely general: the Generalized Interval Systems and their quotients may be finite or infinite, and their groups may be of any structure. Many of Lewin's examples in [4] are effectively quotient GISs. Hook [12] defines a GIS homomorphism, equivalent to a quotient GIS, and proves its construction. His treatment is different in that he takes Lewin's initial intervallic GIS definition (perhaps more familiar to music theorists), rather than Lewin's equivalent formulation in terms of a simply transitive group action on a set, as has been done here.

References

1. Cohn, R.: Square dances with cubes. J. Music Theory **42**(2), 283–296 (1998)
2. Cohn, R.: Audacious Euphony: Chromaticism and the Triad's Second Nature. Oxford University Press, Oxford (2012)

3. Cook, R.: Transformational approaches to romantic harmony and the late works of César Franck. Ph.D. dissertation, University of Chicago (2001)
4. Lewin, D.: Generalized Musical Intervals and Transformations. Yale University Press New Haven (1987). Reprinted Oxford University Press, Oxford (2007)
5. Clampitt, D.: Pairwise well-formed scales: structural and transformational properties. Ph.D. dissertation, SUNY at Buffalo (1997)
6. Cohn, R.: Introduction to Neo-Riemannian theory: a survey and a historical perspective. J. Music Theory **42**(2), 167–180 (1998)
7. Cohn, R.: Weitzmann's regions, my cycles, and Douthett's dancing cubes. Music Theory Spectr. **22**(1), 89–103 (2000)
8. Orvek, D.: Generalized transformational voice-leading systems. M.A. thesis, Ohio State University (2019)
9. Straus, J.: Sum class. J. Music Theory **62**(2), 279–338 (2018)
10. Kochavi, J.: Some structural features of contextually-defined inversion operators. J. Music Theory **42**(2), 307–320 (1998)
11. Lewin, D.: Musical Form and Transformation: 4 Analytical Essays. Yale University Press, New Haven (1993). Reprinted Oxford University Press, Oxford (2007)
12. Hook, J.: Cross-type transformations and the path consistency condition. Music Theory Spectr. **29**(1), 1–39 (2007)

Extended Vuza Canons

Greta Lanzarotto[1,2,3,4] and Ludovico Pernazza[1(✉)]

[1] Department of Mathematics, University of Pavia, Via Ferrata, 5, 27100 Pavia, Italy
[2] Department of Mathematics and its Applications, University of Milano-Bicocca,
Piazza dell'Ateneo Nuovo, 1, 20126 Milan, Italy
g.lanzarotto@campus.unimib.it
[3] INdAM, Città Universitaria - P.le Aldo Moro, 5, 00185 Rome, Italy
greta.lanzarotto@math.unistra.fr
[4] University of Strasbourg, IRMA UMR 7501, 7 rue René-Descartes,
67084 Strasbourg Cedex, France
ludovico.pernazza@unipv.it
https://matematica.unipv.it/en/homepage-english/,
https://www.matapp.unimib.it/en, https://www.altamatematica.it/en/,
https://irma.math.unistra.fr/?lang=en

Abstract. Starting from well-known constructions of aperiodic tiling rhythmic canons by G. Hajós, N.G. de Bruijn and D.T. Vuza, several generalisations are given. In this way, it is possible to find new aperiodic canons, that we call *extended Vuza canons*.

Keywords: Mathematical models for music · Vuza canons · Aperiodic factorisations of cyclic groups

1 Prelude

Canons in music have a very long tradition; among these, a few cases of *tiling rhythmic canons* (i.e. canons such that, given a fixed tempo, at every beat exactly one voice is playing) have emerged. Only in the last century, stemming from the analogous problem of factorizing finite abelian groups, *aperiodic tiling rhythmic canons* have been studied: these are canons that tile a certain interval of time in which each voice (*inner voice*) plays at an aperiodic sequence of beats, and the sequence of starting beats of every voice (*outer voice*) is also aperiodic. From the musical point of view the seminal paper was probably the four-parts article written by D.T. Vuza between 1991 and 1993 [14–17], while the mathematical counterpart of the problem was studied also before, e.g. by de Bruijn [5], Sands [13], etc., and after, e.g. by Coven and Meyerowitz [4], Jedrzejewski [9], Amiot [1], Andreatta [3], etc.

A thorough theory of the conditions of existence and the structure of aperiodic tiling rhythmic canons has not been established yet. In this paper we try to give a contribution to this fascinating field.

2 Aperiodic Tiling Canons

We begin fixing some notations and giving the main definitions. In the following, we conventionally denote the cyclic group of remainder classes modulo n by $\mathbb{Z}_n$ and its elements with the integers $\{0, 1, \ldots, n-1\}$, i.e., identifying each class with its least non-negative member.

Definition 1. *Let $A, B \subset \mathbb{Z}_n$. Let us define the application*

$$\sigma : A \times B \to \mathbb{Z}_n, (a, b) \mapsto a + b.$$

We set $A + B \doteq Im(\sigma)$; if σ is bijective, we say that A and B are in direct sum, and we write

$$A \oplus B \doteq Im(\sigma).$$

If $\mathbb{Z}_n = A \oplus B$, we call (A, B) a tiling rhythmic canon *of period n; A is called the* inner voice *and B the* outer voice *of the canon.*

Remark 1. It is easy to see that the tiling property is invariant under translations, i.e., if A is a tiling complement of some set B, also any translate $A + z$ of A is a tiling complement of B (and any translate of B is a tiling complement of A). Thus, without loss of generality, we shall limit our investigation to rhythms containing 0 and consider equivalence classes under translation.

Definition 2. *A rhythm $A \subset \mathbb{Z}_n$ is* periodic *(of period z) if and only if there exists an element $z \in \mathbb{Z}_n$, $z \neq 0$, such that $z + A = A$. In this case, A is also called* periodic modulo $z \in \mathbb{Z}_n$. *A rhythm $A \subset \mathbb{Z}_n$ is* aperiodic *if and only if it is not periodic.*

Denote by $\Phi_d(x)$ the *cyclotomic polynomial of index d*. Then, tiling rhythmic canons can be characterised as follows.

Lemma 1. *Let A be a rhythm in $\mathbb{Z}_n$ and let $A(x)$ be the* characteristic polynomial *of A, that is, $A(x) = \sum_{k \in A} x^k$. Given $B \subset \mathbb{Z}_n$ and its characteristic polynomial $B(x)$, we have that*

$$A(x) \cdot B(x) \equiv \sum_{k=0}^{n-1} x^k = \frac{x^n - 1}{x - 1} = \prod_{d \mid n, d \neq 1} \Phi_d(x) \qquad \mathrm{mod} \ (x^n - 1) \quad (1)$$

if and only if $A(x)$ and $B(x)$ are polynomials with coefficients in $\{0, 1\}$ and $A \oplus B = \mathbb{Z}_n$.

As a consequence, for each $d \mid n$, with $d > 1$, we have

$$\Phi_d(x) \mid A(x) \ \text{or} \ \Phi_d(x) \mid B(x).$$

Definition 3. *A tiling rhythmic canon* (A, B) *in* $\mathbb{Z}_n$ *is an* aperiodic tiling rhythmic canon *if both* A *and* B *are aperiodic.*

For an extensive discussion on tiling problems, we refer the reader to Amiot [2]. If we indicate the set $\{d \in \mathbb{N} : d \mid n\}$ by $\mathsf{div}(n)$, the following proposition establishes a polynomial criterion for the aperiodicity of a given rhythm.

Proposition 1. *A set* $A \subset \mathbb{Z}_n$ *is aperiodic if and only if for all* $k \mid n$, *with* $k \neq n$, *we have*

$$\frac{x^n - 1}{x^k - 1} \nmid A(x),$$

that is, if and only if for all $k \in \mathsf{div}(n) \setminus \{n\}$ *there exists* $d \in \mathsf{div}(n) \setminus \mathsf{div}(k)$ *such that* $\Phi_d(x) \nmid A(x)$.

The following result, in conjunction with Theorem 2, identifies which are the periods of aperiodic tiling rhythmic canons.

Theorem 1 (Vuza). *Let*

- $\mathcal{V} \doteq \{n \in \mathbb{N} : n = p_1 n_1 p_2 n_2 n_3 \text{ with } \gcd(p_1 n_1, p_2 n_2) = 1 \text{ and } p_1, n_1, p_2, n_2, n_3 > 1\}$, *and*
- $\mathcal{H} \doteq \{p^\alpha, p^\alpha q, p^2 q^2, pqr, p^2 qr, pqrs : \alpha \in \mathbb{N}, \ p, q, r, s \text{ distinct primes}\}$,

then $\mathbb{N}^* = \mathcal{V} \sqcup \mathcal{H}$.

The minimum period necessary for an aperiodic canon is 72, and the corresponding p_i and n_i are:

$$(p_1, n_1, p_2, n_2, n_3) = (2, 2, 3, 3, 2).$$

3   Extended Vuza Canons

The canons with periods 72, 108, 120, 144 and 168 have been completely enumerated by Vuza [14], Fripertinger [7], Amiot [1], Kolountzakis and Matolcsi [11].

An exhaustive construction method for aperiodic tiling rhythmic canons is not known to date; the first method to find some of them was provided by the following result (see [8] by Hajós, Theorem 1 in [5] by de Bruijn, and Proposition 2.2 in [14] by Vuza).

Theorem 2. *Let* $n = p_1 n_1 p_2 n_2 n_3 \in \mathbb{N}$ *such that*

1. $p_1, n_1, p_2, n_2, n_3 > 1$ *and*
2. $\gcd(p_1 n_1, p_2 n_2) = 1$.

Then $\mathbb{Z}_n$ *admits an aperiodic tiling rhythmic canon.*

Example 1. In the hypotheses of Theorem 2, an example of tiling canon of $\mathbb{Z}_n$ with two aperiodic subsets is given by the following construction by F. Jedrze-jewski (see Theorem 227 in [9]). Indicating with $\mathbb{I}_k$ the set $\{0, 1, \ldots, k - 1\}$, let us call:

$$
\begin{aligned}
A_1 &= n_3 p_1 n_1 \mathbb{I}_{n_2} & A_2 &= n_3 p_2 n_2 \mathbb{I}_{n_1} \\
U_1 &= n_3 p_1 n_1 n_2 \mathbb{I}_{p_2} & U_2 &= n_3 p_2 n_2 n_1 \mathbb{I}_{p_1} \\
V_1 &= n_3 n_2 \mathbb{I}_{p_2} & V_2 &= n_3 n_1 \mathbb{I}_{p_1} \\
K_1 &= \{0\} & K_2 &= \{1, 2, \ldots, n_3 - 1\}.
\end{aligned}
$$

Then taking

$$
\begin{aligned}
A &= A_1 \oplus A_2 \\
B &= (U_1 \oplus V_2 \oplus K_1) \sqcup (U_2 \oplus V_1 \oplus K_2),
\end{aligned}
$$

we have the canon $\mathbb{Z}_n = A \oplus B$.

Remark 2. From now on, given p_1, n_1, p_2, n_2, and n_3, we will denote by A_1, A_2, U_1, U_2, V_1, and V_2 the sets so called in Example 1.

Many other ways of constructing aperiodic tiling canons are possible, see for example de Bruijn [5], Vuza [14], Fidanza [6], and Jedrzejewski [9]. These meth-ods fall into a category treated by F. Jedrzejewski (Theorem 14 in [10]). We refine his result lifting the hypothesis that p_1 and p_2 are prime and proving that B is aperiodic if n_3 satisfies a simple arithmetic constraint.

Theorem 3. *Let $n = p_1 n_1 p_2 n_2 n_3 \in \mathbb{N}$ such that:*

1. *$p_1, n_1, p_2, n_2, n_3 > 1$;*
2. *$\gcd(p_1 n_1, p_2 n_2) = 1$;*
3. *there is no prime q such that $q \mid n_3$, but $q \nmid p_1 n_1 p_2 n_2$.*

Let H be the subgroup $H = n_3 \mathbb{I}_{p_1 n_1 p_2 n_2}$ of $\mathbb{Z}_n$ and let K be a complete set of cosets representatives for $\mathbb{Z}_n$ modulo H such that K is the disjoint union $K = K_1 \sqcup K_2$. Then the pair (A, B) defined by

$$
\begin{aligned}
A &= A_1 \oplus A_2 \\
B &= (U_1 \oplus V_2 \oplus K_1) \sqcup (U_2 \oplus V_1 \oplus K_2)
\end{aligned}
$$

is an aperiodic tiling rhythmic canon of $\mathbb{Z}_n$.

Proof. The proof that $A \oplus B = \mathbb{Z}_n$ and that the set A is aperiodic is the same as in Vuza (Proposition 2.2 in [14]). We are left to prove that B is aperiodic. Consider the characteristic polynomial $B(x)$:

$$
B(x) = \frac{x^{n_3 p_1 n_1} - 1}{x^{n_3 n_1} - 1} \frac{x^n - 1}{x^{n_3 p_1 n_1 n_2} - 1} K_1(x) + \frac{x^{n_3 p_2 n_2} - 1}{x^{n_3 n_2} - 1} \frac{x^n - 1}{x^{n_3 p_2 n_2 n_1} - 1} K_2(x).
$$

Given any $h \in \operatorname{div}(n) \setminus \{n\}$, we look for a $d \in \operatorname{div}(n) \setminus \operatorname{div}(h)$ such that $\Phi_d(x) \nmid B(x)$. Let us consider the cases:

1. if $n_3 p_2 n_2 n_1 \nmid h$, then $\Phi_{n_3 p_2 n_2 n_1}(x) \nmid B(x)$ since

$$\Phi_{n_3 p_2 n_2 n_1}(x) \mid \frac{x^n - 1}{x^{n_3 p_1 n_1 n_2} - 1}$$

but

$$\Phi_{n_3 p_2 n_2 n_1}(x) \nmid \frac{x^{n_3 p_2 n_2} - 1}{x^{n_3 n_2} - 1} \frac{x^n - 1}{x^{n_3 p_2 n_2 n_1} - 1} K_2(x).$$

In particular, $\Phi_{n_3 p_2 n_2 n_1}(x) \nmid K_2(x)$ by Lemma 4 of Rédei's paper [12].

2. if $n_3 p_1 n_1 n_2 \nmid h$, then $\Phi_{n_3 p_1 n_1 n_2}(x) \mid B(x)$ (symmetrically to the previous case).

There are no other possibilities: in fact, if we had $n_3 p_2 n_2 n_1 \mid h$ and $n_3 p_1 n_1 n_2 \mid h$, then $h = \alpha n_3 p_2 n_2 n_1 = \beta n_3 p_1 n_1 n_2$ and therefore $\alpha p_2 = \beta p_1$. Since $\gcd(p_1, p_2) = 1$, it would follow $\alpha = p_1$ and $\beta = p_2$ and so $h = n$, which is a contradiction. $\square$

Example 2. Consider $n = 216$; let $p_1 = 2$, $n_1 = 2$, $p_2 = 3$, $n_2 = 3$, and $n_3 = 6$. Theorem 3 ensures that, defining

$$A = 24\mathbb{I}_3 \oplus 54\mathbb{I}_2$$
$$B = (72\mathbb{I}_3 \oplus 12\mathbb{I}_2 \oplus \{0, 106\}) \sqcup (108\mathbb{I}_2 \oplus 18\mathbb{I}_3 \oplus \{21, 43, 122, 167\}),$$

$A \oplus B = \mathbb{Z}_{216}$ and (A, B) is an aperiodic tiling rhythmic canon.

In a first generalization of Theorem 3, rhythm B is the disjoint union of three sets, one being periodic both modulo n/p_1 and modulo n/p_2.

Theorem 4. *Let $n = p_1 n_1 p_2 n_2 n_3 \in \mathbb{N}$ such that:*

1. $p_1, n_1, p_2, n_2, n_3 > 1$;
2. $\gcd(p_1 n_1, p_2 n_2) = 1$;
3. *there is no prime q such that $q \mid n_3$, but $q \nmid p_1 n_1 p_2 n_2$.*

Let H be the subgroup $H = n_3 \mathbb{I}_{p_1 n_1 p_2 n_2}$ of $\mathbb{Z}_n$ with $n = p_1 n_1 p_2 n_2 n_3$, K be a complete set of cosets representatives for $\mathbb{Z}_n$ modulo H such that K is the disjoint union $K = K_1 \sqcup K_2 \sqcup K_3$ with $K_1, K_2 \neq \varnothing$, and $W = n_3 n_1 n_2 \mathbb{I}_{p_1 p_2}$. Then the pair (A, B) defined by

$$A = A_1 \oplus A_2$$
$$B = (U_1 \oplus V_2 \oplus K_1) \sqcup (U_2 \oplus V_1 \oplus K_2) \sqcup (W \oplus K_3)$$

is an aperiodic tiling rhythmic canon of $\mathbb{Z}_n$.

Proof. The only case we need to consider is $K_3 \neq \varnothing$ (notice that this is possible only if $n_3 > 2$). We already know, from Theorem 2 that A is aperiodic; B is aperiodic too, since

$$B(x) = U_1(x)V_2(x)K_1(x) + U_2(x)V_1(x)K_2(x) + W(x)K_3(x)$$

and the cyclotomic polynomials $\Phi_{n_3 p_2 n_2 n_1}$ and $\Phi_{n_3 p_1 n_1 n_2}$ divide exactly 2 of the summands on the right hand side.

We now prove that $A \oplus B = \mathbb{Z}_n$: to this aim we make use of the following facts, proven by F. Jedrzejewski (Theorem 14 in [10]):

$$A_1 + U_1 + V_2 = A_1 + U_1 + U_2$$
$$A_2 + U_2 + V_1 = A_2 + U_2 + U_1.$$

By an easy check, we see that

$$U_1 + U_2 = n_3 n_1 n_2 \left(p_1 \mathbb{I}_{p_2} + p_2 \mathbb{I}_{p_1} \right) = n_3 n_1 n_2 \mathbb{Z}_{p_1 p_2} = W,$$

and $|U_1||U_2| = p_2 p_1 = |W|$. This means that

$$U_1 \oplus U_2 = W.$$

We obtain that

$$
\begin{aligned}
A + B &= (A_1 + A_2) + ((U_1 + V_2 + K_1) \sqcup (U_2 + V_1 + K_2) \sqcup (W + K_3)) \\
&= (A_1 + A_2 + U_1 + V_2 + K_1) \sqcup (A_1 + A_2 + U_2 + V_1 + K_2) \\
&\qquad\qquad\qquad\qquad\qquad \sqcup (A_1 + A_2 + W + K_3) \\
&= (A_1 + A_2 + U_1 + U_2 + K_1) \sqcup (A_1 + A_2 + U_2 + U_1 + K_2) \\
&\qquad\qquad\qquad\qquad\qquad \sqcup (A_1 + A_2 + U_1 + U_2 + K_3) \\
&= A_1 + A_2 + U_1 + U_2 + (K_1 \sqcup K_2 \sqcup K_3) \\
&= A_1 + U_1 + A_2 + U_2 + K.
\end{aligned}
$$

Again, an easy computation shows that

$$
\begin{aligned}
(A_1 + U_1) + (A_2 + U_2) &= n_3 p_1 n_1 \mathbb{I}_{p_2 n_2} + n_3 p_2 n_2 \mathbb{I}_{p_1 n_1} \\
&= n_3 \mathbb{I}_{p_1 n_1 p_2 n_2} \\
&= H
\end{aligned}
$$

and so

$$A + B = H + K = \mathbb{Z}_n.$$

Moreover, since $|A||B| = n = |H||K|$, the sum $A + B$ is direct. $\square$

Example 3. Let us go back to $n = 216$ with the same choices of p_1, n_1, p_2, n_2, and n_3. By Theorem 4, we find a new aperiodic tiling rhythmic canon (A, B) defining

$A = 24\mathbb{I}_3 \oplus 54\mathbb{I}_2$

$B = (72\mathbb{I}_3 \oplus 12\mathbb{I}_2 \oplus \{0, 106\}) \sqcup (108\mathbb{I}_2 \oplus 18\mathbb{I}_3 \oplus \{21, 43\}) \sqcup (36\mathbb{I}_6 \oplus \{122, 167\})$.

The second generalization of Theorem 3 widens the definitions of sets A_1, A_2, V_1, and V_2. We precede it with a useful lemma.

Lemma 2. *Suppose that a subset $S \subseteq \mathbb{Z}_n$ is periodic of period $m \mid n$, i.e. $S + m = S$, and for $i = 0, \dots, k-1$ let $S_i = \{a \in S : a \equiv i \mod k\}$ where k is a divisor of m. Then also the sets S_i are periodic of period m for every i.*

Proof. It is sufficient to observe that since m is a multiple of k the remainder classes modulo k are invariant by the translation by m, hence also $S_i + m = S_i$. □

Theorem 5. *Let $n = p_1 n_1 p_2 n_2 n_3 \in \mathbb{N}$ such that:*

1. *$p_1, n_1, p_2, n_2, n_3 > 1$;*
2. *$\gcd(p_1 n_1, p_2 n_2) = 1$;*
3. *there is no prime q such that $q \mid n_3$, but $q \nmid p_1 n_1 p_2 n_2$.*

Let H be the subgroup $H = n_3 \mathbb{I}_{p_1 n_1 p_2 n_2}$ of $\mathbb{Z}_n$, and $K = K_1 \sqcup K_2$ (with $K_1, K_2 \neq \varnothing$) be a complete set of cosets representatives for $\mathbb{Z}_n$ modulo H. Take

- *$\tilde{A}_1$ as a complete aperiodic set of coset representatives for $\mathbb{Z}_{p_2 n_2}$ modulo $n_2 \mathbb{I}_{p_2}$;*
- *$\tilde{A}_2$ as a complete aperiodic set of coset representatives for $\mathbb{Z}_{p_1 n_1}$ modulo $n_1 \mathbb{I}_{p_1}$;*
- *$\tilde{V}_1^1, \dots, \tilde{V}_1^j$ as complete aperiodic sets of coset representatives for $\mathbb{Z}_{p_2 n_1}$ modulo $p_2 \mathbb{I}_{n_1}$;*
- *$\tilde{V}_2^1, \dots, \tilde{V}_2^h$ as complete aperiodic sets of coset representatives for $\mathbb{Z}_{p_1 n_2}$ modulo $p_1 \mathbb{I}_{n_2}$.*

Set $K_1 = K_1^1 \sqcup \cdots \sqcup K_1^j$ and $K_2 = K_2^1 \sqcup \cdots \sqcup K_2^h$, where $K_\alpha^s = \{k_\alpha^{j_{s-1}+1}, \dots, k_\alpha^{j_s}\}$ are non-empty subsets of K_α for $\alpha = 1, 2$. Then the pair (A, B) defined by

$$A = n_3 p_1 n_1 \tilde{A}_1 \oplus n_3 p_2 n_2 \tilde{A}_2$$
$$B = \left(\left(U_1 \oplus n_3 n_1 \tilde{V}_2^1 \oplus \left\{ k_1^1, \dots, k_1^{l_1} \right\} \right) \sqcup \cdots \right.$$
$$\left. \cdots \sqcup \left(U_1 \oplus n_3 n_1 \tilde{V}_2^j \oplus \left\{ k_1^{l_{j-1}+1}, \dots, k_1^{|K_1|} \right\} \right) \right)$$
$$\sqcup \left(\left(U_2 \oplus n_3 n_2 \tilde{V}_1^1 \oplus \left\{ k_2^1, \dots, k_2^{m_1} \right\} \right) \sqcup \cdots \right.$$
$$\left. \cdots \sqcup \left(U_2 \oplus n_3 n_2 \tilde{V}_1^h \oplus \left\{ k_2^{m_{h-1}+1}, \dots, k_2^{|K_2|} \right\} \right) \right)$$

is an aperiodic tiling rhythmic canon of $\mathbb{Z}_n$.

Proof. We have

- $n_3 p_1 n_1 \tilde{A}_1 + U_1 = n_3 p_1 n_1 \left(\tilde{A}_1 \oplus n_2 \mathbb{I}_{p_2} \right) = n_3 p_1 n_1 \mathbb{I}_{p_2 n_2} = A_1 + U_1$;
- $n_3 p_2 n_2 \tilde{A}_2 + U_2 = n_3 p_2 n_2 \left(\tilde{A}_2 \oplus n_1 \mathbb{I}_{p_1} \right) = n_3 p_2 n_2 \mathbb{I}_{p_1 n_1} = A_2 + U_2$;
- $A_1 + n_3 n_1 \tilde{V}_2 = n_3 n_1 \left(p_1 \mathbb{I}_{n_2} + \tilde{V}_2 \right) = n_3 n_1 \mathbb{I}_{p_1 n_2} = A_1 + V_2$;
- $A_2 + n_3 n_2 \tilde{V}_1 = n_3 n_2 \left(p_2 \mathbb{I}_{n_1} + \tilde{V}_1 \right) = n_3 n_2 \mathbb{I}_{p_2 n_1} = A_2 + V_1$.

For the sake of simplicity, we now give the proof in the case $j = 1$ and $h = 1$. The general case is completely analogous. We compute

$$
\begin{aligned}
A + B &= \left(n_3 p_1 n_1 \tilde{A}_1 + n_3 p_2 n_2 \tilde{A}_2 \right) \\
&\quad + \left(\left(U_1 + n_3 n_1 \tilde{V}_2 + K_1 \right) \sqcup \left(U_2 + n_3 n_2 \tilde{V}_1 + K_2 \right) \right) \\
&= \left(n_3 p_1 n_1 \tilde{A}_1 + n_3 p_2 n_2 \tilde{A}_2 + U_1 + n_3 n_1 \tilde{V}_2 + K_1 \right) \\
&\quad \sqcup \left(n_3 p_1 n_1 \tilde{A}_1 + n_3 p_2 n_2 \tilde{A}_2 + U_2 + n_3 n_2 \tilde{V}_1 + K_2 \right) \\
&= \left(A_1 + n_3 p_2 n_2 \tilde{A}_2 + U_1 + n_3 n_1 \tilde{V}_2 + K_1 \right) \\
&\quad \sqcup \left(n_3 p_1 n_1 \tilde{A}_1 + A_2 + U_2 + n_3 n_2 \tilde{V}_1 + K_2 \right) \\
&= \left(A_1 + n_3 p_2 n_2 \tilde{A}_2 + U_1 + V_2 + K_1 \right) \\
&\quad \sqcup \left(n_3 p_1 n_1 \tilde{A}_1 + A_2 + U_2 + V_1 + K_2 \right) \\
&= \left(A_1 + n_3 p_2 n_2 \tilde{A}_2 + U_1 + U_2 + K_1 \right) \\
&\quad \sqcup \left(n_3 p_1 n_1 \tilde{A}_1 + A_2 + U_2 + U_1 + K_2 \right) \\
&= A_1 + A_2 + U_1 + U_2 + (K_1 \sqcup K_2) \\
&= A_1 + U_1 + A_2 + U_2 + K \\
&= \mathbb{Z}_n.
\end{aligned}
$$

A cardinality argument analogous to that used in Theorem 4 shows that the sum is direct.

The proof that A is aperiodic follows from Vuza's argument (Proposition 2.2 in [14]), as above. Assume now that B is periodic of period a: we can assume without loss of generality that $a = n/p$ where p is a prime number. Hypothesis 3. now implies that a is a multiple of n_3: but then by Lemma 2 also the sets $B_i = B \cap (\{i\} + n_3 \mathbb{Z}_n)$ must be periodic of period a. However, the sets B_i are simply translates of $U_1 \oplus n_3 n_1 \tilde{V}_2$ by elements of K_1 or of $U_2 \oplus n_3 n_2 \tilde{V}_1$ by elements of K_2 (remember that also the elements of U_1 and U_2 are multiple of n_3): on their turn, $U_1 \oplus n_3 n_1 \tilde{V}_2$ and $U_2 \oplus n_3 n_2 \tilde{V}_1$ are indeed periodic resp. of period n/p_1 and n/p_2, but since p_1 and p_2 are coprime no common period smaller than n is possible. A contradiction follows since we assumed both K_1 and K_2 to be non-empty. □

Remark 3. Note that Theorems 3–5 hold trivially if hypothesis 3. is replaced by the condition that n_3 is prime.

Example 4. This time we choose $n = 252$; let $p_1 = 2$, $n_1 = 7$, $p_2 = 3$, $n_2 = 3$, and $n_3 = 2$. We can take e.g.

$$\tilde{A}_1 = \{0, 2, 7\} \qquad \tilde{A}_2 = \{0, 1, 3, 4, 9, 12, 13\}$$
$$\tilde{V}_1 = \{0, 10, 17\} \qquad \tilde{V}_2 = \{0, 1\} = \mathbb{I}_{p_1}$$
$$K_1 = \{0\} \qquad K_2 = \{1\}$$

obtaining a new canon (A, B) where

$$A = 28\tilde{A}_1 \oplus 18\tilde{A}_2$$
$$= \{0, 56, 196\} \oplus \{0, 18, 54, 72, 162, 216, 234\}$$
$$B = \left(U_1 \oplus 14\tilde{V}_2 \oplus K_1 \right) \sqcup \left(U_2 \oplus 6\tilde{V}_1 \oplus K_2 \right)$$
$$= \left(\{0, 84, 168\} \oplus \{0, 14\} \oplus \{0\} \right) \sqcup \left(\{0, 126\} \oplus \{0, 60, 102\} \oplus \{1\} \right).$$

Definition 4. *We call* Vuza canons *all the canons obtained using the constructions described in Theorems 2, 3, 4, 5.*

It is possible to stretch this type of constructions even further. With the following theorem, we improve the result of Jedrzejewski (Theorem 21 in [10]).

Theorem 6. *Let $n = p_1 n_1 p_2 n_2 n_3 \in \mathbb{N}$ such that:*

1. $p_1, n_1, p_2, n_2, n_3 > 1$;
2. $\gcd(p_1 n_1, p_2 n_2) = 1$;
3. there is no prime q such that $q \mid n_3$, but $q \nmid p_1 n_1 p_2 n_2$.

Let H be the subgroup $H = n_3 \mathbb{I}_{p_1 n_1 p_2 n_2}$ of $\mathbb{Z}_n$. Suppose that L and K are proper subsets of $\mathbb{Z}_{n_3}$ such that $L \oplus K = \mathbb{Z}_{n_3}$ and $K = K_1 \sqcup K_2$, with $K_1, K_2 \neq \varnothing$. Then the pair (A, B) defined by

$$A = A_1 \oplus A_2 \oplus L$$
$$B = (U_1 \oplus V_2 \oplus K_1) \sqcup (U_2 \oplus V_1 \oplus K_2)$$

is an aperiodic tiling rhythmic canon of $\mathbb{Z}_n$.

Proof.

$$A + B = (A_1 + A_2 + L) + ((U_1 + V_2 + K_1) \sqcup (U_2 + V_1 + K_2))$$
$$= (A_1 + A_2 + L + U_1 + V_2 + K_1) \sqcup (A_1 + A_2 + L + U_2 + V_1 + K_2)$$
$$= (A_1 + A_2 + L + U_1 + U_2 + K_1) \sqcup (A_1 + A_2 + L + U_2 + U_1 + K_2)$$
$$= A_1 + A_2 + L + U_1 + U_2 + (K_1 \sqcup K_2)$$
$$= A_1 + U_1 + A_2 + U_2 + L + K.$$

The sum is direct because the computation of the cardinality leads to

$$|A_1||A_2||U_1||U_2||L \oplus K| = n.$$

Aperiodicity of A is immediate from Lemma 2, since $A_1 + A_2$ is aperiodic, and B is the union of the subsets B_i contained in different remainder classes modulo n_3, some of which have a period coprime with the period of the other ones (exactly as in the previous theorem).

Example 5. Choosing again $n = 216$ and the same values for p_1, n_1, p_2, n_2, and n_3 as in Example 3, we set $L = \{0,1\}$, $K_1 = \{2\}$, and $K_2 = \{0,4\}$. By Theorem 6, we get that

$$A = 24\mathbb{I}_3 \oplus 54\mathbb{I}_2 \oplus L$$
$$B = (72\mathbb{I}_3 \oplus 12\mathbb{I}_2 \oplus K_1) \sqcup (108\mathbb{I}_2 \oplus 18\mathbb{I}_3 \oplus K_2)$$

define an aperiodic tiling rhytmic canon.

To prove our next result we take advantage of the equivalent polynomial formulation of tilings. Using it, in [4] E.M. Coven, and A. Meyerowitz introduced two sufficient conditions for a rhythm A to be a factor of a tiling rhythmic canon. To state them we need the following definitions.

Definition 5. $R_A \doteq \{d : \Phi_d(x) \mid A(x)\}$ *and* $S_A \doteq \{p^\alpha \in R_A : p \ prime\}$.

The *Coven-Meyerowitz conditions* are the following:

T1 $|A| = \prod_{p^\alpha \in S_A} p$;
T2 for all $p^\alpha, q^\beta, r^\gamma, \ldots \in S_A$, $p^\alpha q^\beta r^\gamma \cdots \in R_A$, where $p^\alpha, q^\beta, r^\gamma, \ldots$ are powers of distinct primes.

The polynomial approach provides a few new important properties.

Lemma 3. *Let* $A(x), B(x) \in \mathbb{N}[x]$ *and* $n \in \mathbb{N}^*$. *Then*

$$A(x)B(x) \equiv \sum_{k=0}^{n-1} x^k \quad \mathrm{mod} \ (x^n - 1) \tag{T0}$$

if and only if

1. $A(x), B(x) \in \{0,1\}[x]$, *so they are the characteristic polynomials of sets A and B, and*
2. $A \oplus B = \{r_1, \ldots, r_n\} \subset \mathbb{Z}$, *with* $r_i \neq r_j \ \mathrm{mod} \ n$ *for all* $i, j \in \{1, \ldots, n\}$ *with* $i \neq j$.

Lemma 4. *Let* $f(x) \in \mathbb{Z}[x]$ *and* $n \in \mathbb{N}^*$. *The following are equivalent:*

1. $f(x) \equiv \sum_{k=0}^{n-1} x^k \ \mathrm{mod} \ (x^n - 1)$;
2. (a) $f(1) = n$ *and*
   (b) *for every* $d \mid n$, *with* $d > 1$, *we have* $\Phi_d(x) \mid f(x)$.

Definition 6. *Let A be a subset of $\mathbb{Z}_n$ and let $S_A = \{p^\alpha, q^\beta, \ldots, r^\gamma\}$. We call the extension of A any rhythm $\overline{A}$ whose characteristic polynomial is*

$$\overline{A}(x) = \Phi_{p^\alpha}\left(x^{\frac{n}{p^\alpha k_p}}\right) \Phi_{q^\beta}\left(x^{\frac{n}{q^\beta k_q}}\right) \cdots \Phi_{r^\gamma}\left(x^{\frac{n}{r^\gamma k_r}}\right).$$

where $k_p, k_q, \ldots, k_r$ *are divisors of n such that* $p \nmid k_p, q \nmid k_q, \ldots, r \nmid k_r$.

Note that by definition clearly $S_A = S_{\overline{A}}$.

Proposition 2. *Let $A \oplus B = \mathbb{Z}_n$ and let B satisfy condition (T2). Then $\overline{A} \oplus B = \mathbb{Z}_n$, too.*

Proof. Since p^α is a prime power, then

$$\Phi_{p^\alpha} \left(x^{\frac{n}{p^\alpha k_p}} \right) \in \{0,1\}\,[x],$$

and so $\overline{A}(x) \in \mathbb{N}[x]$. Moreover,

- $\overline{A}(1)B(1) = n$ and
- $\Phi_d(x) \mid \overline{A}(x)B(x)$ for all $d \mid n$, with $d > 1$.

By Lemma 4, this means that

$$\overline{A}(x)B(x) \equiv \sum_{k=0}^{n-1} x^k \mod (x^n - 1),$$

that is, condition (T0) in Lemma 3 holds. Therefore $\overline{A}(x) \in \{0,1\}\,[x]$ and $\overline{A} \oplus B = \mathbb{Z}_n$, that is, $\overline{A}$ tiles with B. □

Combining Theorem 6 and Proposition 2, we are able to find new Vuza canons where L is not a subset of $\mathbb{Z}_{n_3}$.

Theorem 7. *Let $n = p_1 n_1 p_2 n_2 n_3 \in \mathbb{N}$ such that:*

1. $p_1, n_1, p_2, n_2, n_3 > 1$;
2. $\gcd(p_1 n_1, p_2 n_2) = 1$;
3. *there is no prime q such that $q \mid n_3$, but $q \nmid p_1 n_1 p_2 n_2$.*

Let H be the subgroup $H = n_3 \mathbb{I}_{p_1 n_1 p_2 n_2}$ of $\mathbb{Z}_n$. Suppose that L and K are proper subsets of $\mathbb{Z}_{n_3}$ such that $L \oplus K = \mathbb{Z}_{n_3}$ and $K = K_1 \sqcup K_2$, with $K_1, K_2 \neq \varnothing$. Let $\tilde{L}$ be an extension of L; then the pair (A, B) defined by

$$A = A_1 \oplus A_2 \oplus \tilde{L}$$
$$B = (U_1 \oplus V_2 \oplus K_1) \sqcup (U_2 \oplus V_1 \oplus K_2)$$

is an aperiodic tiling rhythmic canon of $\mathbb{Z}_n$.

Proof. Since, by definition, A_1 and A_2 coincide with their own extensions, the extension of $A_1 \oplus A_2 \oplus L$ is A. By Theorem 6, $A_1 \oplus A_2 \oplus L \oplus B = \mathbb{Z}_n$, therefore Proposition 2 implies that $A \oplus B = \mathbb{Z}_n$.

We already know from Theorem 6 that B is aperiodic. To show that A is aperiodic, consider $\tilde{L}(x)$. By hypothesis 3 $S_{\tilde{L}}$ does not contain any maximal prime power dividing n, as S_{A_1} and S_{A_2}. As a consequence, $S_A = S_{A_1} \cup S_{A_2} \cup S_{\tilde{L}}$ does not contain any such prime power, either. By Proposition 1, A can not be periodic. □

Definition 7. *We call* extended Vuza canons *all the canons obtained using the constructions of Theorems 6 and 7, possibly combined with those of Theorems 2, 3, 4 and 5.*

Example 6. We show now an extended Vuza canon with period $n = 240$ ($p_1 = 2, n_1 = 2, p_2 = 5, n_2 = 3, n_3 = 4$). Set $L = \mathbb{I}_2$; then $\tilde{L} = 15\mathbb{I}_2$. Choosing $K_1 = \{2\}$ and $K_2 = \{0\}$, we obtain the canon

$$
\begin{aligned}
A &= A_1 \oplus A_2 \oplus \tilde{L} \\
&= 16\mathbb{I}_3 \oplus 60\mathbb{I}_2 \oplus 15\mathbb{I}_2 \\
B &= (U_1 \oplus V_2 \oplus K_1) \sqcup (U_2 \oplus V_1 \oplus K_2) \\
&= (48\mathbb{I}_5 \oplus 8\mathbb{I}_2 \oplus \{2\}) \sqcup (120\mathbb{I}_2 \oplus 12\mathbb{I}_5 \oplus \{0\}) .
\end{aligned}
$$

It is worth noting that it would not be possible to obtain such a canon without applying Theorem 7.

We include below a table showing the number of Vuza canons and extended Vuza canons for all the periods n with values between 72 and 280 (Table 1).

As a final comment, one could say that the recipes by Hajós, de Bruijn and Vuza to generate aperiodic tiling rhythmic canons are deceivingly simple.

Table 1. The number of aperiodic rhythms for non-Hajós values of n from 72 to 280, generated with the constructions described in Theorems 2–7.

n	p_1	n_1	p_2	n_2	n_3	L	#K	#A			#B				
			Theorem:					(2)	(6)	(7)	(2)	(3)	(4)	(5)	(6)
72	2	2	3	3	2	{0}	1	3	0	0	6	0	0	0	0
108	2	2	3	3	3	{0}	1	3	0	0	180	0	72	0	0
120	2	2	3	5	2	{0}	1	16	0	0	20	0	0	0	0
	2	2	5	3	2	{0}	1	8	0	0	18	0	0	0	0
144	2	2	3	3	4	{0}	1	3	0	0	2808	1944	3888	0	0
	2	2	3	3	4	{0, 1}	2	0	312	0	0	0	0	0	6
	2	2	3	3	4	{0, 9}	2	0	0	12	0	0	0	0	6
	2	2	3	3	4	{0, 2}	4	0	156	0	0	0	0	0	12
	2	4	3	3	2	{0}	1	6	0	0	12	0	0	48	0
	4	2	3	3	2	{0}	1	6	0	0	6	0	0	30	0
168	2	2	3	7	2	{0}	1	104	0	0	14	0	0	28	0
	2	2	7	3	2	{0}	1	16	0	0	6	0	0	48	0
180	2	5	3	3	2	{0}	1	9	0	0	15	0	0	105	0
	5	2	3	3	2	{0}	1	6	0	0	6	0	0	90	0
	3	5	2	2	3	{0}	1	16	0	0	500	0	200	1100	0
	5	3	2	2	3	{0}	1	8	0	0	252	0	72	1728	0
	2	2	3	3	5	{0}	1	3	0	0	45360	77760	158112	0	0

(*continued*)

Table 1. (*continued*)

n	p_1	n_1	p_2	n_2	n_3	L	$\#K$	$\#A$			$\#B$				
				Theorem:				(2)	(6)	(7)	(2)	(3)	(4)	(5)	(6)
200	2	2	5	5	2	{0}	1	125	0	0	10	0	0	50	0
216	2	4	3	3	3	{0}	1	6	0	0	180 + 540	72 + 216	0	12672	0
	2	2	3	3	6	{0, 3}	8	0	156	0	0	0	0	0	180 + 540 + 72 + 216
	2	2	3	3	6	{0, 1}	2	0	324	0	0	0	0	0	180 + 72
	2	2	3	3	6	{0}	1	3	0	0	754272	2449440	5832000	0	0
	2	2	3	3	6	{0, 1, 2}	3	0	34992	0	0	0	0	0	6
	2	2	3	3	6	{0, 2, 4}	9	0	10935	0	0	0	0	0	6 + 12
	2	2	3	9	2	{0}	1	729	0	0	6 + 12	0	0	54	0
	2	2	9	3	2	{0}	1	27	0	0	6	0	0	162	0
	4	2	3	3	3	{0}	1	6	0	0	252	0	72	5940	0
240	2	4	3	5	2	{0}	1	32	0	0	20	0	0	20 + 160	0
	2	2	3	5	4	{0, 6}	4	0	0	588	0	0	0	0	20 + 20
	2	2	3	5	4	{0, 2}	4	0	7252	0	0	0	0	0	20 + 20
	2	2	3	5	4	{0, 15}	2	0	0	64	0	0	0	0	20
	2	2	3	5	4	{0, 3}	2	0	1176	0	0	0	0	0	20
	2	2	3	5	4	{0, 1}	2	0	14504	0	0	0	0	0	20
	2	2	3	5	4	{0}	1	16	0	0	13000	9000	18000	94000	0
	2	2	5	3	4	{0}	1	8	0	0	6264	3240	5184	197856	0
	2	2	5	3	4	{0, 1}	2	0	4016	0	0	0	0	0	18
	2	2	5	3	4	{0, 5}	2	0	0	112	0	0	0	0	18
	2	2	5	3	4	{0, 15}	2	0	0	32	0	0	0	0	18
	2	2	5	3	4	{0, 2}	4	0	2008	0	0	0	0	0	12 + 24
	2	2	5	3	4	{0, 10}	4	0	0	56	0	0	0	0	12 + 24
	2	4	5	3	2	{0}	1	16	0	0	12	0	0	24 + 576	0
	4	2	3	5	2	{0}	1	32	0	0	10	0	0	290	0
	4	2	5	3	2	{0}	1	16	0	0	6	0	0	102	0
252	2	7	3	3	2	{0}	1	27	0	0	21	0	0	315	0
	7	2	3	3	2	{0}	1	9	0	0	6	0	0	618	0
	3	7	2	2	3	{0}	1	104	0	0	980	0	392	5096	0
	7	3	2	2	3	{0}	1	16	0	0	324	0	72	21312	0
	2	2	3	3	7	{0}	1	3	0	0	12830400	71383680	206126208	0	0
264	2	2	3	11	2	{0}	1	5368	0	0	22	0	0	88	0
	2	2	11	3	2	{0}	1	40	0	0	6	0	0	552	0
270	3	3	2	5	3	{0}	1	9	0	0	1125	0	450	48825	0
	3	3	5	2	3	{0}	1	6	0	0	288	0	72	48600	0

(*continued*)

Table 1. (*continued*)

n	p_1	n_1	p_2	n_2	n_3	L	$\#K$	$\#A$			$\#B$				
			Theorem:					(2)	(6)	(7)	(2)	(3)	(4)	(5)	(6)
280	2	2	5	7	2	{0}	1	2232	0	0	14	0	0	112	0
	2	2	7	5	2	{0}	1	480	0	0	10	0	0	170	0

Note: In each column only the rhythms that can be generated by the corresponding theorem, but not by previous ones are counted. Grey numbers correspond to rhythms that can be generated also by the choice of parameters in the previous line. When there is no column (e.g., #A for Theorem 3) all the possible rhythms already appear in previous columns.

Their basic mechanism can be (and has indeed been) generalised in several ways; this paper gives a generalisation on its own, but Theorem 7 can certainly still be improved. Further studies should follow, aiming at lifting the hypotheses used in the present results and (hopefully) at establishing a systematic theory of aperiodic tiling rhythmic canons given by all the known constructions, and eventually of all aperiodic tiling rhythmic canons straightaway.

References

1. Amiot, E.: New perspectives on rhythmic canons and the spectral conjecture. J. Math. Music **3**(2), 71–84 (2009)
2. Amiot, E.: Structures, algorithms, and algebraic tools for rhythmic canons. Perspect. New Music **49**(2), 93–142 (2011)
3. Andreatta, M.: Constructing and formalizing tiling rhythmic canons: a historical survey of a "mathemusical" problem. Perspect. New Music **49**(2), 33–64 (2011)
4. Coven, E.M., Meyerowitz, A.: Tiling the integers with translates of one finite set. J. Algebra **212**(1), 161–174 (1999)
5. de Bruijn, N.G.: On the factorization of finite abelian groups. Proc. Koninklijke Nederlandse Akademie van Wetenschappen Ser. A Math. Sci. **61**(3), 258–264 (1953)
6. Fidanza, G.: Canoni ritmici a mosaico. Università degli Studi di Pisa, Tesi di Laurea (2007)
7. Fripertinger, H., Reich, L.: Remarks on rhythmical canons. Colloquium on mathematical music theory. Grazer Math. Ber. **347**, 73–90 (2005)
8. Hajós, G.: Sur le problème de factorisation des groupes cycliques. Acta Mathematica Academiae Scientiarum Hungarica **1**(2), 189–195 (1950)
9. Jedrzejewski, F.: Mathematical Theory of Music. IRCAM/Delatour, Sampzon (2006)
10. Jedrzejewski, F.: Enumeration of Vuza Canons. arXiv preprint arXiv: 1304.6609 (2013)
11. Kolountzakis, M.N., Matolcsi, M.: Algorithms for translational tiling. J. Math. Music **3**(2), 85–97 (2009)
12. Rédei, L.: Ein Beitrag zum Problem der Faktorisation von endlichen abelschen Gruppen. Acta Mathematica Academiae Scientiarum Hungarica **1**(2–4), 197–207 (1950)

13. Sands, A.D.: On the factorisation of finite abelian groups. III. Acta Mathematica Academiae Scientiarum Hungarica **25**(3–4), 279–284 (1974)
14. Vuza, D.T.: Supplementary sets and regular complementary unending canons (part one). Perspect. New Music **29**(2), 22–49 (1991)
15. Vuza, D.T.: Supplementary sets and regular complementary unending canons (part two). Perspect. New Music **30**(1), 184–207 (1992)
16. Vuza, D.T.: Supplementary sets and regular complementary unending canons (part three). Perspect. New Music **30**(2), 102–124 (1992)
17. Vuza, D.T.: Supplementary sets and regular complementary unending canons (part four). Perspect. New Music **31**(1), 270–305 (1993)

Some Mathematical and Computational Relations Between Timbre and Color

Maria Mannone[1,2(✉)] 🔟 and Juan Sebastián Arias-Valero[3] 🔟

[1] Department of Engineering, University of Palermo, Palermo, Italy
`mariacaterina.mannone@unipa.it`
[2] Dipartimento di Scienze Ambientali, Informatica e Statistica (DAIS)
and European Centre for Living Technology (ECLT),
Ca' Foscari University of Venice, Venezia, Italy
`maria.mannone@unive.it`
[3] Bogotá, Colombia

Abstract. In physics, timbre is a complex phenomenon, like color. Musical timbres are given by the superposition of sinusoidal signals, corresponding to longitudinal acoustic waves. Colors are produced by the superposition of transverse electromagnetic waves in the domain of visible light. Regarding human perception, specific timbre variations provoke effects similar to color variations, for example, a rising tension or a relaxation effect. We aim to create a computational framework to modulate timbres and colors. To this end, we consider categorical groupoids, where colors (timbres) are objects and color variations (timbre variations) are morphisms, and functors between them, which are induced by continuous maps. We also sketch some gestural variations of this scheme. Thus, we try to soften the differences and focus on the similarity of structures.

Keywords: Color · Timbre · Topology · Category theory · Gestures

2010 Mathematics Subject Classification: 00A65 · 00A66 · 54-02 · 18-02

1 Introduction

Timbres and colors fascinated musicians, artists, and scientists across centuries [9,20]. In physics, the complexity of timbre is due to the superposition of simple components (sinusoidal waves), which can be separated with Helmholtz resonators [16]. Timbres can be computationally investigated with Fourier transforms and sonograms, which show the strength of each component (partial) of the superposition. Colors are also related to the idea of superposition, as proved by [32] for white light, which can be decomposed in colors through a prism. The physics involved is quite different: sound involves mechanical longitudinal waves, while light is made of electromagnetic transverse waves. However, timbres and colors have a main similarity: they are complex signals, made of simple

---

J.S. Arias-Valero—Independent researcher.

superposed wave signals [37]. This suggests a correspondence based on the relation between sound frequency and spatial frequency of light, but according to [5], absolute correspondences between these domains are difficult to establish, so the relativity and the obstructions of this problem could be softened in the categorical context.

In fact, the point of view of precise measurement can be enriched in several ways. Scholars such as Goethe pointed out the importance of perception to understand colors in the framework of nature and the arts [11]. In addition, both colors and timbres can be qualitatively rated as, for example, *cold*, *strong*, or *delicate*. Even though different cultures can associate a different (symbolic) meaning to each color, we can find aspects with certain universality, related to human perception. Some colors are more instinctively associated with higher or lower tension: red or yellow raise more attention than light blue or gray. Similarly, specific orchestral timbres are more awakening than others: a loud[1] trumpet sound is a more effective alarm than a soft flute melody. Some recent studies point out the importance of a "shared emotion" to associate colors and musical sequences [33], which also occurs in the framework of classical music listening [8]. On the other hand, both colors and timbres can be mixed or shaded—as it happens for painting and orchestration, respectively, transforming a *delicate sound (or color)* into a *strong sound (or color)*. In this way, we can draw upon the idea of *superposition* and *similarity of perception* to imagine how we can investigate colors and timbres, focusing on common aspects through abstraction.[2] These aspects are intensities, mixing, and shadows/nuances. In particular, harmonic choices, which influence timbre, are also ruled by the idea of superposition.

In this article we introduce fundamental groupoids of color and timbre spaces and functors between them. These functors could be induced by some classical (possibly) continuous maps suggested in [5]. This categorical framework [22, 23] could be adequate to express the superposition and similarity principles to understand the color/timbre relation, complementing analytical approaches. Categories have already been used to investigate processes and phenomena in the arts from a bird's-eye perspective [19, 21, 30].

This article is structured as follows. In Sect. 2, we review some color spaces, timbre spaces, and maps between them. In Sect. 3, we offer a categorical enrichment of previous approaches to relate color and timbre. In particular, in Sect. 3.3, we include a computational example of interaction between color and timbre paths. Then, Sect. 4 is devoted to a gestural extension of the previous enrichment. In Sect. 5 some conclusions and further possible applications are discussed. In the *Glossary* (Sect. 6) we provide definitions of some specialized mathematical concepts that we mention. We use **boldface** for these terms.

As a general disclaimer: colors, timbres, and their relationships constitute a vast topic. This is a position paper (or rather, a working one) aiming to open the way toward further studies in this field.

---

[1] Loudness in music performance can affect timbre characteristics [6].

[2] A first experiment, where participants were asked to associate colors, color bands, and timbres, confirmed a non-negligible perceptive correlation [26].

2   Spaces and Mappings: An Overview

2.1   The CIE 1931 Color Space

The CIE model [10] connects the visible spectrum with human perception. It assigns to each spectrum wavelength $\lambda \in [380, 780]$, measured in nanometres, three sensitivity level values $\overline{x}(\lambda)$, $\overline{y}(\lambda)$, and $\overline{z}(\lambda)$ corresponding to the kinds of human cone cells under certain standard conditions. Thus, a spectral distribution yields, by integration of its product with each color matching function ($\overline{x}$, $\overline{y}$, or $\overline{z}$), a triple (X, Y, Z) of color coordinates. All these triples amount to the unit cube $[0, 1]^3$, after normalizing units. We embed this cube in $\mathbb{R}^3$, regarding the latter as a vector space and a topological space. The vector sum in the cube, whenever defined, corresponds to color mixing (superposition of light beams).[3] If the sum is not in $[0, 1]^3$, one can take an average of vector components to represent a mixture (with average intensity) for computational purposes.

On the other hand, the standard RGB space is used for screens and photography, so we need it for experiments. It has red, green, and blue as primary colors, which give white if superposed. The standard RGB model does not cover the CIE gamut in principle, for instance, a spectral violet. However, we can transform CIE to standard RGB by means of an appropriate conversion of CIE to linear RGB followed by electro-optical transfer. The RGB space has already been considered for mathematical modeling [34,35]. In particular, [35] proposed a three-dimensional space of perceived colors, where equivalence classes correspond to perceptual match.

2.2   Timbre Space

As a possible representation of timbres, we can consider the space proposed by Grey [13], based on the dissimilarity between pairs of musical instrument sounds.

On the other hand, we have the set of all continuous periodic maps. These maps represent continuous sound waves that can be recovered from their Fourier series according to Fourier's and Fejér's theorems [4, Section 2.4]. It is embedded in the space of all continuous maps $\mathbb{R} \longrightarrow \mathbb{R}$, which has the **compact-open topology** and is a vector space. Superposition of waves corresponds to addition of the associated periodic functions, although the result need not be periodic. In what follows, we take the topological space of continuous periodic maps as our timbre space, given the structural analogy with the CIE space in the sense that color/wave superpositions correspond to vector sums.

2.3   Maps Between Timbre and Color

According to [5], a possible correspondence between color and sound can be based on the idea that a musical octave should match a color octave. A musical

---

[3] We mix colors in printing and painting with the subtractive model, a sort of dual of the additive one.

octave is a closed interval of the form $[f, 2f]$, where f is a fixed sound frequency in Hertz. Human vision barely ranges through color octave, namely the interval of wavelengths in nanometres $[380, 760]$, which corresponds to the interval of spatial frequencies $[(1/2)(1/380), 1/380]$ by means of the assignment $\lambda \mapsto 1/\lambda$. Thus, the map $\lambda \mapsto 760f/\lambda$ is a continuous bijection from the color octave $[380, 760]$ to the musical octave $[f, 2f]$. Note that under this logic, the color order violet-blue-green-yellow-orange-red corresponds to a decreasing pitch frequency.

Since human hearing ranges frequencies in the Hertz interval $[20, 20000]$, and therefore several octaves, there is not a perfect correspondence between wavelengths and frequencies. This suggests reducing the interval $[20, 20000]$ modulo a chosen octave and then using the previous correspondence. The resulting map is continuous under the assumption that we identify the endpoints of $[380, 760]$.

There are other possibilities for a correspondence between color and sound. Some scholars focus on perceived correspondences of pitch classes with classes of colors [18]. The use of classes can be formalized by means of quotient spaces. Classes take into account perceptive similarities but not perfect one-to-one associations. Other continuous correspondences could associate the transition from violet to red with an increasing pitch frequency.

The following construction is a possible way to get a continuous[4] map from the timbre space to the CIE color space. First, let us consider the case of a timbre given by simple FM synthesis [4, Sect. 8.8], namely a periodic[5] wave corresponding to

$$\sin[\omega_c t + I \sin(\omega_m t)], \tag{1}$$

where $\omega_c = 2\pi f_c$, $\omega_m = 2\pi f_m$, f_c is the carrier frequency, f_m is the modulator frequency, and I is the modulation index. An associated convergent series is

$$\sum_{n=-\infty}^{\infty} J_n(I) \sin[(\omega_c + n\omega_m)t], \tag{2}$$

where J_n is the nth Bessel function of the first kind. This series expresses the wave in terms of simple harmonics with frequencies $f_c + nf_m$ for $n \in \mathbb{Z}$. By factorizing the sign of each negative value of $f_c + nf_m$ outside of $\sin[(f_c + nf_m)t]$ we obtain:

$$\sum_{n=0}^{\infty} a_n \sin(2\pi f_n t). \tag{3}$$

Thus, given a continuous map h from frequencies to color wavelengths, we construct (by linearity) the series in the CIE space

$$\sum_{n=0}^{\infty} a_n XYZ(h(f_n)), \tag{4}$$

---

[4] We do not have a proof of this continuity.

[5] The wave is periodic if the quotient between carrier and modulator frequencies is a rational number.

where $XYZ(\lambda)$ gives the CIE coordinates of the wavelength λ, whenever the series converges in the CIE space. In general, one could use the Fourier series [30, p. 1019]:

$$a_0 + \sum_{n=1}^{\infty} a_n \sin(2\pi n f t + \phi_n) \qquad (5)$$

of the given continuous periodic wave and associate the series (if it converges in the CIE space)

$$a_0 + \sum_{n=1}^{\infty} a_n XYZ(h(nf)), \qquad (6)$$

but it is to be determined whether (1) this procedure coincides with that used for FM synthesis and (2) the phase ϕ_n affects the color quality. These are open questions. In Sect. 3.3 we exemplify computationally the procedure for the FM case.

3   Categorical Enrichment

Color and timbre, and their relations, can be recast in a categorical framework, where we emphasize the color and timbre transitions, rather than the objects *color* and *timbre* themselves.

Each topological space X (like the CIE and timbre space) has an associated category whose morphisms are invertible, that is, a *groupoid*. Its objects are the elements of X and its morphisms are homotopy classes of paths in X. The composition $[\tau] \circ [\sigma]$ of two classes $[\sigma] : x \longrightarrow y$ and $[\tau] : y \longrightarrow z$ is the class of the concatenation $\sigma\tau$. The identity on x is the class of the associated constant map and the inverse of a path σ sends $t \in [0, 1]$ to $\sigma(1 - t)$. This construction can be generalized to yield higher relations between paths as follows.

3.1   Induced Infinity-Groupoids

Let us consider the **singular complex** $\mathrm{Sing}(X)$, which is a **simplicial set** and an ∞-**groupoid**, under the definitions in Sect. 6. According Proposition 1.9 and Remark 1.10 from [14], ∞-categories have n-morphisms for each $n \geq 0$ and composition of them, which is associative up to homotopy. Thus, 1-*morphism* of $\mathrm{Sing}(X)$ is a path in X, and a 2-*morphism* is a homotopy between two paths with the same endpoints. Note that the groupoid of X comes from homotopy classes of 1-morphisms and hence the concatenation of them is associative up to homotopy equivalence. On the other hand, a 2-morphism can be seen as a band of intermediate paths between two given ones that connect the same points. Figure 1 shows examples of 1-morphisms and 2-morphisms in the cases of the CIE and timbre spaces. More generally, we can define n-morphisms of the singular complex, which describe the evolution of a single color (timbre), of a path of colors (timbres), of a homotopy of paths, and so on.

We emphasize the need for higher relations and bands. For example, we can map the transition *light blue→dark blue* into the transition *light green→dark*

green, creating a band that connects, as different shades, light green with light blue, and dark green with dark blue.[6] If the initial and final points of the band coincide, we can have the situation described in Fig. 1, where the dark blue becomes a light blue through different paths: some paths remain in the blue area, while other ones cross the violet area [27].
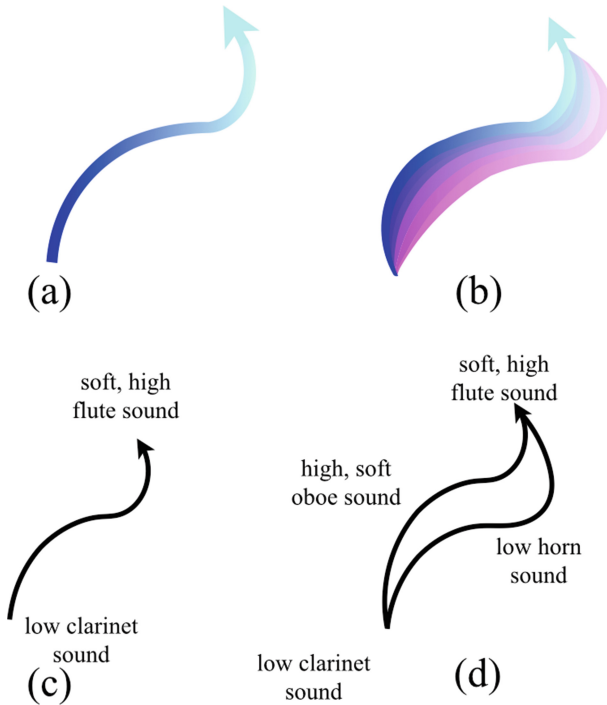

Fig. 1. (a) A 1-morphism in the space of colors, a path between two colors, (b) a 2-morphism in the same space, a band between two color paths, (c) a 1-morphism in the timbre space, and (d) a 2-morphism in the same space.

3.2   Induced Functors

Given two topological spaces X and Y, which can be the timbre and the CIE color space respectively, and a continuous map $f : X \longrightarrow Y$ there is an induced natural transformation $F : \mathrm{Sing}(X) \longrightarrow \mathrm{Sing}(Y)$ that sends a *singular n-simplex* $\sigma : \Delta^n \longrightarrow X$ to $f\sigma : \Delta^n \longrightarrow Y$. According to the definition in Sect. 1.2.7 of

---

[6] This relation is not a proper morphism but it can be achieved by pasting two suitable 2-simplices in $\mathrm{Sing}(X)$, which is a higher gesture according to [2]. It is also a hypergesture in the sense of [31].

[22], which says that a functor between **infinity-categories** is a natural transformation between the respective simplicial sets, F is a functor from $\mathrm{Sing}(X)$ to $\mathrm{Sing}(Y)$.

Note that F coincides with f on objects and sends a 1-morphism σ in X to the path $f\sigma$ in Y.

As any functor between infinity-categories, F preserves the usual categorical structure (up to homotopy), in the sense that

$$F([id_x]) = [id_{f(x)}]$$

whenever $x \in X$ and

$$F([\tau] \circ [\sigma]) = F([\tau]) \circ F([\sigma])$$

whenever $\sigma : x \longrightarrow y$ and $\tau : y \longrightarrow z$ are paths in X. More generally, F preserves the compositions of higher morphisms in an appropriate sense, but we omit these technical details. Next, a computational sketch of a functor from timbre to color.

3.3   A Computation of Colors from Timbres

As an example of associations between a timbre path and a color path, let us consider the progressive enrichment of a simple 440 Hz sine wave with harmonics, using FM synthesis, and the associated color transition.

More formally, take $f_c = 440$ and $f_m = 2f_c$. By regarding the modulation index I as a parameter in the interval $[0, 20]$, we obtain a continuous[7] path in the timbre space with parametrization (Eq. 1):

$$\sin[\omega_c t + I \sin(\omega_m t)].$$

The result is a fluctuation in the brilliance of a sort of clarinet sound since only odd harmonics are present.[8] Figure 2 is the corresponding spectrogram of the timbre path.

To obtain a color path (Fig. 3) we use the procedure in Sect. 2.3 for each value of I, see Eq. (6). For each new value of the index modulation I, harmonics vary, reaching a new timbre in Fig. 2. For each new value of I, and thus, for each timbre point reached, there is a color point reached in Fig. 3. In fact, each color bar represents a color point in the space of colors. This could mean that we are using the functor induced by any of the continuous maps from timbres to colors (Sect. 2.3), according to Sect. 3.2. In Fig. 3, the color squares correspond to the modulation index I values $n/10$ for integers n from 0 to 200. There, the modulation index increases from left to right and from top to bottom. Then one uses conversion to RGB for screen representation (Sect. 2.1).

---

[7] This defines a continuous map $[0, 20] \times \mathbb{R} \longrightarrow \mathbb{R}$, so the exponential transpose $[0, 20] \longrightarrow \mathbb{R}^{\mathbb{R}}$, which is a path, is continuous.

[8] The audio file, in which we identify the increasing modulation index with time (in seconds), can be accessed from the link https://soundcloud.com/maria-mannone/fm-path/s-cFJ7kNrqJjs?.

The Python codes for the FM path and color path are available at https://github.com/medusamedusa/color_gesture.

The results agree with Caivano's reflections [5]: the closer the sound to a white noise, the closer the color to white light, with additive color mixing. The inverse choice could associate the richness of harmonics (especially in a low-register orchestral range) with a darker color, more like in painting, with subtractive color mixing. In the first case, primary colors are red, blue, and green, and their sum gives white; in the second case, primary colors are red, yellow, and blue, and their sum gives black. In gestural chromo-similarity (Sect. 4.1), in analogy with painting we may use the second option (subtractive), see an example in [27].
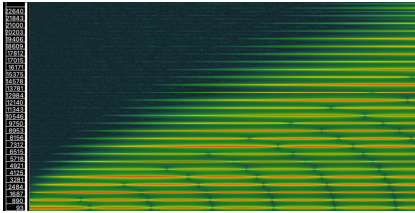
Fig. 2. An example of timbre path. The spectrogram is obtained with SonicVisualiser. The darker the color, the closer the sound to silence.
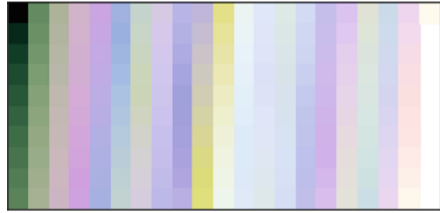

Fig. 3. Visual color gradient corresponding to the timbre path of Fig. 2. Each color corresponds to a value of the modulation index I.

4   Gestural Considerations

We close this paper with some gestural reflections that may enrich the color and timbre relation theory.

4.1   From Paths to Gestures and Gestural Similarity

Color and timbre paths (or 1-morphisms) are particular cases of **gestures** [1,2, 7,17,25,28,31], which are informally diagrams (shaped by a digraph) of paths in a topological space. Continuous maps induce new ones between respective spaces of gestures, as we explain in Sect. 4.2, so there are correspondences between color gestures and timbre gestures.

We can talk of *gestural similarity* if musical sequences (auditory domain) and simple sketches (visual domain) appear as being produced by the same generator gesture [24]. This possible definition is supported by the hypothesis of a *supramodal brain* [36]. Thus, when gestures in the space of colors and gestures in the space of timbres show perceptive analogies, we can talk of *chromo-gestural similarity*.

4.2   Induced Maps Between Spaces of Gestures

Let Γ be a digraph. A continuous map $f : X \longrightarrow Y$ induces a new one[9] between **topological spaces of Γ-gestures**, namely

$$\Gamma \pitchfork F : \Gamma \pitchfork S_X \to \Gamma \pitchfork S_Y : ((c_a)_{a \in A}, (x_v)_{v \in V}) \mapsto ((fc_a)_{a \in A}, (f(x_v))_{v \in V}),$$

where $\Gamma \pitchfork S_X$ ($\Gamma \pitchfork S_Y$) is the **space of Γ-gestures in X (Y)** (respectively).

As an example, the Attack-Delay-Sustain-Release (ADSR) envelope of a sound is a gesture shaped by the digraph $\bullet \to \bullet \to \bullet \to \bullet \to \bullet$ in the amplitude-time space. The envelope has a main role in timbre perception. We can transfer the envelope to the color space by regarding it as an intensity gesture of a single color. In fact, this remark may raise new questions regarding color envelopes, and transitions effects from a color to another one.

Color and timbre ramifications, which are gestures, are interesting objects to study and apply to composition. Shaping the orchestral colors, in particular, is a distinctive mark of a composer's style, of a genre, of an epoch. Thus, the proposed ideas can be developed in terms of machine learning as exploited in music information retrieval. Vice versa, a creative interface may be developed starting from the proposed theoretic tools.

Note that the objects and 1-morphisms of the ∞-groupoid $\mathrm{Sing}(\Gamma \pitchfork S_X)$ are Γ-gestures in X and paths between gestures, respectively. This groupoid allows one to generalize the idea of timbre paths to transformations between timbres with different ADSR envelopes (loudness profile over time). We may, for example, keep the timbre of a musical instrument while changing its envelope, or keep the envelope and change the timbre, thus performing separate transformations of the envelope and timbre in terms of spectral superposition. As a final abstraction, $\Gamma \pitchfork F$ induces a functor between ∞-groupoids $\mathrm{Sing}(\Gamma \pitchfork S_X) \longrightarrow \mathrm{Sing}(\Gamma \pitchfork S_Y)$, which would help transfer envelope transitions between the color and timbre domains.

5   Conclusion

The proposed categorical framework could be a way to understand the relation between color and timbre, complementing classical approaches from physics. This framework is based on structural analogies between the perceptual domains of hearing and vision. It is interesting to ask to what extent categorical models could be independent from perception and classical models, taking into account the computational advantages of the latter.

We also proposed a gestural extension of the categorical framework to capture gestural similarities between the musical and auditory domains.

As a possible, alternative structure to look at, we could consider the Moore paths as 1-cells, in order to have a strictly associative composition, taking homotopies of homotopies for the 2 cells [12]. Given that we are interested in invertible

---

[9] In essence, it is continuous because each component (composition with f for arrows and f for vertices) is.

arrows, another suitable structure appears to be the **bigroupoid** [15], that is, a weakly-invertible **bicategory**. Concerning the spaces, we could also consider the Euclidean space of colors (as RGB) and the Euclidean space of timbres as defined by Grey [13]. In a (bi)groupoid, all arrows are invertible. In this way, the points (single colors, single timbres) are 0-cells; the color gestures and timbre gestures are the paths, the 1-cells; the bands (hypergestures in the sense of [31]) are the 2-cells. Path associativity is verified for equivalence classes of homotopies. The model of bigroupoid for color and timbre gestures is discussed in detail in [27].

However, we stress the fact that ∞-categories simplify the involved axioms and computations in higher category, 2- and bi-categories included.

This research could lead to signal processing practical implementations, and it could provide a theoretical framework to analyze experiments in the domain of musical timbre. On the creative side, other possible directions may involve the development of interfaces for composers to manipulate timbres through symbols and/or color references, and for visual artists to do the inverse.

The possibility of translating structures from one domain to another one, provided that some cognitive conditions are verified [29], can open scenarios also for disability studies, where people with visual impairment can benefit from auditory-accessible interfaces, and people with auditory impairment can benefit from visually-accessible interfaces [5, pp. 128–129]. The reference to gesture and touch regarding intensity, organization, and time distribution of stimuli can inspire even more audacious applications for touch-based interfaces for deaf-blind people.

Thus, a simple question such as "can we join timbres and colors?" can open the way to striking applications to improve people's lives.

6   Glossary

Bicategory. In a bicategory, the morphism composition is not associative, but only associative up to an isomorphism. This notion has been introduced by Bénabou in 1967 [3]. The objects are the 0-cells, the morphisms are the 1-cells, and the morphisms between morphisms are the 2-cells.

Bigroupoid. A bigroupoid is a bicategory whose "2-cells are strictly invertible, and the 1-cells are invertible up to coherent isomorphism" [15].

Compact-Open Topology. The subbasic opens of the *compact-open topology* on the space of continuous maps $\mathbb{R}^{\mathbb{R}}$ are those of the form

$$\{f : \mathbb{R} \longrightarrow \mathbb{R} \text{ continuous } \mid f(K) \subseteq U\},$$

where K is compact (closed and bounded) in $\mathbb{R}$ and U is open in $\mathbb{R}$. This makes $\mathbb{R}^{\mathbb{R}}$ an exponential in the category of topological spaces. The fact that Top is not Cartesian closed does not imply the non-existence of $\mathbb{R}^{\mathbb{R}}$.

Simplicial Category. Denote by $[n]$ the ordered set (ordinal) $\{0, 1, \ldots, n\}$ for $n \in \mathbb{N}$. The *simplicial category* Δ has as objects all $[n]$ for $n \in \mathbb{N}$ and as morphisms all order-preserving maps between them.

Standard Simplex (Functor). For each $n \in \mathbb{N}$, we define the *standard n-simplex* Δ^n as the set

$$\{(t_1, \ldots, t_n) \mid 0 \le t_1 \le \cdots \le t_n \le 1\}.$$

The standard n-simplex is a subspace of $\mathbb{R}^n$ and this construction defines a *standard simplex functor* $\Delta^{(-)}$ from the **simplicial category** to the category of topological spaces, which sends an order-preserving map $\alpha : [n] \longrightarrow [m]$ to the appropriate continuous map $\Delta^\alpha : \Delta^n \longrightarrow \Delta^m$ sending the ith vertex (with $n - i$ zeros) to the $\alpha(i)$th one. *Examples*: Δ^0 is a singleton, Δ^1 is the interval $[0, 1]$ in $\mathbb{R}$; Δ^2 is the triangle with vertices $(0, 0)$, $(0, 1)$, and $(1, 1)$ in $\mathbb{R}^2$; and Δ^3 is the tetrahedron with vertices $(0, 0, 0)$, $(0, 0, 1)$, $(0, 1, 1)$, and $(1, 1, 1)$ in $\mathbb{R}^3$.

Simplicial Set. Functor from the opposite Δ^{op} of the **simplicial category** to the category **Set** of sets. *Example*: The **singular complex** $\text{Sing}(X)$ of a topological space X.

Singular Complex. The *singular complex of a topological space X*, denoted by $\text{Sing}(X)$, is the **simplicial set Top**$(\Delta^{(-)}, X)$, where $\Delta^{(-)}$ is the **standard simplex functor**. *Examples*: a 0-simplex of $\text{Sing}(X)$ is a point of X, a 1-simplex of $\text{Sing}(X)$ is a path in X.

Infinity-Category. A **simplicial set** S is a set such that given $n \in \mathbb{N}$ and k with $0 < k < n$, for each subset $\{a_i \mid 0 \le i \le n; \ i \ne k\}$ of $S([n-1])$ satisfying the identities

$$d_i(a_j) = d_{j-1}(a_i) \ \ (i < j; \ i, j \ne k),$$

there is an element $a \in S([n])$ such that $d_i(a) = a_i$ for $i \ne k$. If this property also holds for $k = 0$ and $k = n$, then we say that S is an ∞-**groupoid**. *Example*: The **singular complex** of a topological space is an ∞-groupoid.

Topological Space of Gestures. Let Γ be a digraph (A, V, d_0, d_1) and X a topological space. The *space of Γ-gestures in X*, denoted by $\Gamma \pitchfork S_X$ (where $\pitchfork$ stands for transversality), is the subspace of the product space (compact-open topology on X^I)

$$\left(X^I\right)^A \times X^V$$

consisting of all sequences $((c_a)_{a \in A}, (x_v)_{v \in V})$ such that $c_a(i) = x_{d_i(a)}$ for $i = 0, 1$. We say that such a sequence is a Γ-**gesture in** X.

References

1. Arias, J.S.: Spaces of gestures are function spaces. J. Math. Music **12**, 89–105 (2018). https://doi.org/10.1080/17459737.2018.1496489
2. Arias-Valero, J.S., Lluis-Puebla, E.: Simplicial Sets and Gestures: Mathematical Music Theory, Infinity-Categories, Homotopy, and Homology. Under review (2020)
3. Bénabou, J.: Introduction to bicategories, part I. In: Bénabou, J. (ed.) Reports of the Midwest Category Seminar. Lecture Notes in Mathematics, vol. 47, pp. 1–77. Springer, Heidelberg (1967). https://doi.org/10.1007/BFb0074299
4. Benson, D.: Music: a mathematical offering (2008). https://homepages.abdn.ac.uk/d.j.benson/pages/html/music.pdf
5. Caivano, J.: Color and sound: physical and psychophysical relations. Color. Res. Appl. **19**, 126–133 (1994)
6. Castellengo, M.: Écoute Musicale et Acoustique. Eyrolles, Paris (2015)
7. Clark, T.: On the topological characterization of gestures in a convenient category of spaces. J. Math. Music **15**(1), 37–61 (2020)
8. Crnjanski, N., Tomaš, D.: Musical perception and visualization. In: Paper Read at Music and Spatiality. 13th Biennale International Conference on Music Theory and Analysis (2019)
9. da Vinci, L.: Trattato della pittura. Unione cooperativa editrice, reprint (1890). https://archive.org/details/trattatodellapit00leon_0
10. Fairman, H.S., Brill, M.H., Hemmendinger, H.: How the CIE 1931 color-matching functions were derived from wright-guild data. Color. Res. Appl. **22**(11), 11–23 (1997)
11. Goethe, J.W.V.: Theory of Colours (Zur Farbenlehre). Cotta'schen Buchhandlung (1810)
12. Grandis, M.: Higher fundamental groupoids for spaces. Topol. Appl. **129**(3), 281–299 (2003). https://www.sciencedirect.com/science/article/pii/S0166864102001852
13. Grey, J.: Multidimensional perceptual scaling of musical timbres. J. Acoust. Soc. Am. **61**, 1270–1277 (1877)
14. Groth, M.: A short course on ∞-categories. In: Handbook of Homotopy Theory, chapter 14. Chapman and Hall (2020). https://people.math.rochester.edu/faculty/doug/otherpapers/groth_scinfinity.pdf
15. Hardie, K.A., Kamps, K.H., Kieboom, R.W.: A homotopy bigroupoid of a topological space. Appl. Categ. Struct. **9**, 311–327 (2001)
16. Helmholtz, H.v.: On the Sensations of Tone as a Physiological Basis for the Theory of Music (English translation). Longmans, Green (1895). https://archive.org/details/onsensationston02helmgoog
17. Hughes, J.R.: Generalizing the orbifold model for voice leading. Mathematics **10**(6), 939 (2022). https://www.mdpi.com/2227-7390/10/6/939
18. Itoh, K., Sakata, H., Kwee, I., Nakada, T.: Musical pitch classes have rainbow hues in pitch class-color synesthesia. Nat. Sci. Rep. **7**, 17781 (2017) https://www.nature.com/articles/s41598-017-18150-y
19. Jedrzejewski, F.: Hétérotopies Musicales. Hermann, Paris (2019)
20. Kandinsky, W.: Complete writings on art. In: Lindsay, K.C., Vergo, P. (eds.) Da Capo Press (1994)
21. Kubota, A., et al.: A new kind of aesthetics-the mathematical structure of the aesthetic. Philosophies **2**, 1–14 (2017)

22. Lurie, J.: Higher Topos Theory. Academic Search Complete. Princeton University Press (2009)
23. Mac Lane, S.: Categories for the Working Mathematician. Springer, New York (1998). https://doi.org/10.1007/978-1-4757-4721-8
24. Mannone, M.: Introduction to gestural similarity in music. An application of category theory to the orchestra. J. Math. Music **12**, 63–87 (2018)
25. Mannone, M.: Knots, music and DNA. J. Creat. Music Syst. **2**(2), 1–22 (2018). https://www.jcms.org.uk/article/id/523/
26. Mannone, M., Distefano, V., Santini, G.: Classes of Colors and Timbres: A Clustering Approach. Under review
27. Mannone, M., Santini, G., Adedoyin, E., Cella, C.E.: Color and timbre gestures: an approach with bicategories and bigroupoids. Mathematics **10**(4), 663 (2022).https://doi.org/10.3390/math10040663
28. Mannone, M., Turchet, L.: Shall we (math and) dance? In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O.A. (eds.) MCM 2019. LNCS (LNAI), vol. 11502, pp. 84–97. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_7
29. Mannone, M., Favali, F., Di Donato, B., Turchet, L.: Quantum GestART: identifying and applying correlations between mathematics, art, and perceptual organization. J. Math. Music **15**(1), 62–94 (2021)
30. Mazzola, G., et al.: The Topos of Music. Birkhäuser (2002)
31. Mazzola, G., Andreatta, M.: Diagrams, gestures and formulae in music. J. Math. Music **1**, 23–46 (2007)
32. Newton, I.: Opticks, or, A Treatise of the Reflections, Refractions, Inflections and Colours of Light. S. Smith and B. Walford, London (1704). https://www.loc.gov/resource/rbctos.2017gen39060/?st=gallery
33. Palmer, S., Schloss, K., Xu, Z., Prado-Leon, L.: Music-color associations are mediated by emotion. Proc. Natl. Acad. Sci. **110**(22): 8836–8841 (2013). https://www.pnas.org/doi/10.1073/pnas.1212562110
34. Provenzi, E.: Geometry of color perception. Part 1: structures and metrics of a homogeneous color space. J. Math. Neurosci. **10**, 1–19 (2020). https://mathematical-neuroscience.springeropen.com/articles/10.1186/s13408-020-00084-x
35. Resnikoff, H.: On the psychophysical function. J. Math. Biol. **2**, 265–276 (1975)
36. Rosenblum, L., Dias, J., Dorsi, J.: The supramodal brain: implications for auditory perception. J. Cogn. Psychol. **1**, 65–87 (2016)
37. Sethares, W.: Tuning, Timbre, Spectrum. Springer, Heidelberg (2005)

Transformations for Pairwise Well-Formed Modes

Thomas Noll[1] and David Clampitt[2(✉)]

[1] Departament de Creació i Teoria Musical, Escola Superior de Música de Catalunya, Barcelona, Spain

[2] School of Music, Ohio State University, Columbus, USA
`clampitt.4@osu.edu`

Abstract. The paper extends the transformational approach to pairwise well-formed (PWWF) modes, represented as words over a 3-letter alphabet. In particular it addresses open problems arising from an earlier approach (Noll, T and D. Clampitt. 2018. "Kaleidoscope substitutions and pairwise well-formed modes: Major-Minor duality transformationally revisited", *Journal of Mathematics and Music.* 12(3)), wherein kaleidoscope transformations were shown to generate PWWF modes, but it was not yet shown that all PWWF modes could be so generated. This gap is filled. A further problem is that the kaleidoscope transformations are not closed under composition. The final section introduces a new construction for the generation of PWWF words/modes, transformations of words on a 4-letter alphabet. A strongly-supported conjecture is that these transformations form a monoid.

Keywords: Pairwise well-formed scales · Kaleidoscope transformations · Diatonic and syntonic modes · Sturmian morphisms · Algebraic combinatorics on words

1 Motivation

The present paper revisits some open issues arising from [8], where its authors—Noll and Clampitt—propose a quite convoluted transformational approach to the study of *pairwise well-formed modes* for a brief introduction to PWWF scales see [3]. This concept builds upon the more elementary concept of *well-formed modes*, and the transformational modeling of the pairwise well-formed modes builds upon an earlier transformational modeling of the well-formed modes in terms of *Sturmian morphisms* (see [7] or [5] for the mathematical foundations of this approach and see [4] for the music-theoretical application).

The mathematical objects under investigation are words, which encode modal interval species in terms of their step interval patterns. The underlying alphabet for these words consists of as many letters as there are different step intervals in the mode. For the encoding of the Glarean modes Ionian, Dorian, Phrygian etc., only two different letters are needed. One—say *a*—for the major second and another—say *b*—for the minor second.

The step interval patterns of the pairwise well-formed modes consist of three different step intervals, which leads to their study in terms of three-letter-words, i.e., words which are made of the three letters a, b, c. The attribute *pairwise* refers to the projections of a given three-letter-word to the two-letter words, which arise whenever two letters (a pair) out of the three letters are identified with one another. This happens, for example, when the word *acbacab* is projected to *aabaaab*, where every instance of c is replaced by an instance of a. The condition of well-formedness must be satisfied by all these projections. Recall that a two-letter word is well-formed if it is conjugate to a *Christoffel word*. Roughly speaking, this means that the modes, which exemplify these words through their scale step patterns, behave like the diatonic modes. They need to be well-formedly generated, i.e., their generating interval must be stable in that all its instances have the same span in the sense of step-counting.

A prominent example is Zarlino's major mode (*modo sintonico diatonico*) (see Fig. 1) with the step interval pattern *acbacab* whose diatonic and syntonic projections $\pi_{c \to a}(acbacab) = aabaaab$ and $\pi_{c \to b}(acbacab) = abbabab$ both satisfy the condition of well-formedness (for a detailed investigation of this example see [9]).
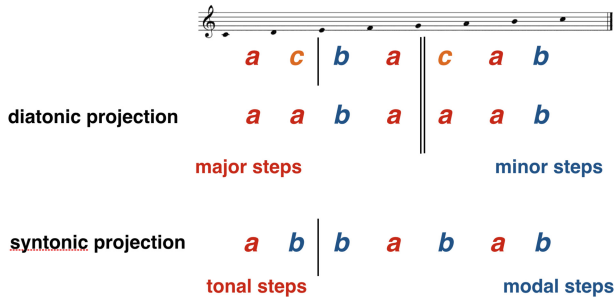

Fig. 1. Zarlino's major scale (just intonation major) as a PWWF mode

Most of the new ideas of the present article emerged from the investigation of a self-similar manifestation of Zarlino's major mode alongside with its syntonic and diatonic projections within a regular 41-tone system. In $\mathbb{Z}_{41}$ one finds specific realizations of all three modes: syntonic, diatonic and major (=syntonic diatonic) and the last is also a superposition of the the latter two:

Mode	Syntonic	Diatonic	Major
Generic step pattern	a b b a b a b	a a b a a b a	a c b a c a b
Specific steps in $\mathbb{Z}_{41}$	(7, 5, 5, 7, 5, 7, 5)	(7, 7, 3, 7, 7, 7, 3)	(7, 6, 4, 7, 6, 7, 4)

The central subject of the investigations in [8] is the concept called *kaleidoscope substitution*, which characterizes a certain subclass of substitutions on a

monoid $\mathcal{A}^*$ of words on the three-letter alphabet $\mathcal{A} = \{a, b, c\}$. And the kaleido-scope transformations are meant to generate all these step-interval patterns as images $f(abc)$ of the starting word abc. Under this perspective each pairwise well-formed mode gains a three-fold division on the form $f(a)|f(b)|f(c)$ into three interval species. The motivation for the usage of the term *kaleidoscope* is born of the idea to lift the very definition of Pairwise Well-Formedness to the transfor-mational language. The comparison of these transformations with some patterns in a system of mirrors (a kaleidoscope) is based upon the observation that these 3-fold divisions are not just analogues to the 2-fold authentic and plagal divisions of the well-formed modes, but rather compound configurations thereof. And this extends to the transformations: Kaleidoscope transformations arise somehow by gluing together two *Sturmian morphisms*, which themselves are closely related to each other. One difficulty of the paper arises from the fact that—unlike the Sturmian morphisms—the kaleidoscope transformations do not form a group or monoid. To control their extension requires a troublesome engagement with gen-erative grammars. On the one hand, it is desirable to overcome this difficulty and to reformulate the entire approach in a more elegant way. Section 5 is dedi-cated to such a project. On the other hand the paper leaves—despite its length of nearly 40 pages—a crucial argumentative gap. The authors indeed show, that for every kaleidoscope transformation κ the images $\kappa(abc)$ are pairwise well-formed modes. But what they claim and do not explicitly demonstrate is that every pairwise well-formed word w can actually be expressed as an image $w = \kappa(abc)$ of a suitable kaleidoscope transformation. Sections 2, 3, 4 and 5 of the present paper deal with the task to close this gap. Section 4 bridges the two parts and brings forward a new idea about the construction of specific PWWF scales from specific instances of their syntonic and diatonic projections.

2 Letter Frequencies and Periods in Pairwise Well-Formed Words

We consider the three-letter alphabet $\mathcal{A} = \{a, b, c\}$. Let $\mathcal{A}^*$ and $\{a, b\}^*$ denote the monoids of words in the three letters a, b, c and the two letters a, b, respec-tively. Three-letter-words can be reduced to two-letter-words by means of letter projections:

Definition 2.1. *Let x, y and z denote the three different letters of $\mathcal{A}$. Let $\pi_{x \to y}$: $\mathcal{A}^* \to \{y, z\}^*$ denote the substitution which is induced by the projection of the letter x to the letter y: $\pi_{x \to y}(x) = y, \pi_{x \to y}(y) = y, \pi_{x \to y}(z) = z$. The three letters x, y and z are called the* abandoned, receiving *and* uninvolved *letters, respectively. Let further $E_{xy} : \mathcal{A}^* \to \mathcal{A}^*$ denote the substitution which is induced by the exchange of the letters x and y.*

In particular, we consider the three projections

$$\pi_{c \to b}, \pi_{c \to a}, \pi_{c \to a} \circ E_{bc} : \{a, b, c\}^* \to \{a, b\}^*.$$

We consider the set of all non-singular pairwise well-formed (PWWF) words over the 3-letter alphabet $\{a, b, c\}$.

Definition 2.2. *A word $w \in \mathcal{A}^*$ is called pairwise well-formed (PWWF), if all three projections $\pi_{c \to b}(w)$, $\pi_{c \to a}(w)$, and $\pi_{c \to a}(E_{bc}(w))$ are well-formed, i.e., if they are standard words (images of ab under compositions of G, D, and E) or conjugates thereof. $\pi_{c \to b}(w)$ is called the* syntonic *and $\pi_{c \to a}(w)$, $\pi_{c \to a}(E_{bc}(w))$ are called the* first and second diatonic projections *of w, respectively.*

It is known from [2] that with the single exception of the word *abacaba* (or some variant with permuted letters) pairwise well-formedness implies that precisely two of the three letters have the same multiplicity. Thus—excluding the singular case—but without further loss of generality, we may restrict our investigation to the set $\Omega \subset \{a, b, c\}^*$ of all PWWF words w, satisfying $|w|_b = |w|_c$, and in the rest of this article all instances of the term *pairwise well-formed* should be understood as *non-singular pairwise well-formed*.

Consider a PWWF word $w \in \Omega$. The two letter projections $u = \pi_{c \to b}(w)$ and $v = \pi_{c \to a}(w)$ are called the *syntonic* and (first) *diatonic* projections of w, respectively. Their well-formedness implies that the multiplicities $|u|_b$ and $|v|_b$ of the letter b are both co-prime with the word length $n = |w|$ and so we may consider their inverses $p = |u|_b^{-1}$ and $q = |v|_b^{-1} \bmod n$. The two natural numbers p and q (i.e., the smallest positive representatives of the residue classes $|u|_b^{-1}$ and $|v|_b^{-1} \bmod n$) are called the *authentic* periods of the words u and v respectively. The numbers $n - p$ and $n - q$ are called their *plagal* periods, accordingly.

Proposition 2.3. *For every PWWF word $w \in \Omega$ precisely one of the following properties holds: Either the authentic period of the diatonic projection is twice the authentic period of the syntonic projection: $q = 2p$ or the plagal period of the diatonic projection is twice the plagal period of the syntonic projection: $n - q = 2(n - p)$.*

Proof. From the initial assumption $|w|_b = |w|_c$ we may conclude that $|u|_b = 2|v|_b$ and hence (taking into account that n must be odd) we see that

$$p = |u|_b^{-1} = (2|v|_b)^{-1} = 2^{-1} q \bmod n, \text{ i.e., } q = 2p \bmod n.$$

We also observe that

$$n - q = n - 2p \bmod n = 2n - 2p \bmod n = 2(n - p) \bmod n.$$

Precisely one of the two equations, either $q = 2p$ or $n - q = 2(n - p)$ holds also in natural numbers (rather than only $\bmod n$), depending whether $q < n - q$ or $n - q < q$.

The usage of the term *period* for these inverse letter multiplicities p and q or their n-complements $n - p$ and $n - q$ is motivated by the desire to represent certain syntonic and diatonic projections as images under Sturmian morphisms.

Definition 2.4. *A PWWF word shall be called* pairwise morphic, *if its syntonic and diatonic projections are morphic (i.e. images $f(ab)$ under suitable Sturmian morphisms $f \in St$). A pairwise morphic word w is said to be authentic if the authentic period of the diatonic projection is twice the authentic period of the syntonic projection, otherwise it is said to be plagal.*

This distinction between authentic and plagal pairwise morphic words turns out to be equivalent to the distinction between special and non-special Sturmian morphisms for their projections.

Proposition 2.5. *A pairwise morphic word is authentic iff its (syntonic and diatonic) projections are images $f(ab)$ under special Sturmian morphisms $f \in \langle G, D \rangle$. In other words, it is plagal iff its projections are images $f(E(ab)) = f(ba)$ under non-special Sturmian morphisms $f \circ E$ with $f \in \langle G, D \rangle$.*

Proof. In the case of a special Sturmian morphisms f with

$$det(M_f) = det \begin{pmatrix} |f(a)|_a & |f(b)|_a \\ |f(b)|_a & |f(b)|_b \end{pmatrix} = |f(a)|_a |f(b)|_b - |f(b)|_a |f(a)|_b = 1$$

the primary period $|f(a)|$ is the multiplicative inverse $mod\, n = |f(ab)|$ of the secondary letter multiplicity $|f(ab)|_b$, because

$$
\begin{aligned}
|f(a)||f(ab)|_b &= (|f(a)|_a + |f(a)|_b)(|f(a)|_b + |f(b)|_b) \\
&= |f(a)|_a |f(a)|_b + |f(a)|_a |f(b)|_b + |f(a)|_b^2 + |f(a)|_b |f(b)|_b \\
&= 1 + |f(a)|_b (|f(a)|_a + |f(b)|_a + |f(a)|_b + |f(b)|_b) \\
&= 1 + |f(a)|_b |f(ab)| = 1 \, mod\, n.
\end{aligned}
$$

Analogously, one shows that the secondary period $|f(a)|$ is the multiplicative inverse of the primary letter multiplicity $|f(ab)|_a$.

In Definition 2.6, Proposition 2.7, and Corollary 2.8 we restrict our attention to standard morphisms. The corollary restates a finding from [8], namely that the syntonic and first diatonic projections are "in sync" in the sense that one of them is a standard morphism iff the other one is. It follows further from proposition 6.16 in [8] that in this case the second diatonic projection is also morphic. All this motivates the following definition.

Definition 2.6. *For a given PWWF word w' we call the pairwise morphic conjugate w, whose syntonic projection $\pi_{c \to b}(w)$ is a standard word, the associated standard conjugate of w'.*

In the sequel we will assume that the standard conjugate w of a PWWF word ends either on ab or ba. To achieve this one may – without loss of generality – chose the word $E_{cb}(w)$ instead.

Proposition 2.7. *The standard conjugate w of a PWWF word is the unique conjugate of the form $w = vab$ or $w = vba$, for which v is a palindrome.*

Proof. Palindromicity implies a literal correspondence between the elements of v and its reverse $\tilde{v}$. In the very middle of v (which is of odd length) there must be an instance of either b (when $|w|_b$ is even) or c (when $|w|_b = |w|_c$ is odd). Due to the strict alternation of bs and cs within v, the palindromicity property of $\pi_{c \to b}(v)$ still holds in v itself, (as cs then correspond to cs and bs to bs). The uniqueness of this property among the conjugates of w follows from the fact that the central words $\pi_{c \to b}(v)$ and $\pi_{c \to a}(v)$ are the only palindromic factors of length $n - 2$ in w (seen as a cyclic word).

Corollary 2.8. *For the standard conjugate w of a PWWF word both projections $\pi_{c\to b}(w)$ (syntonic) and $\pi_{c\to a}(w)$ (first diatonic) are standard (either both special or both non-special).*

Proof. Suppose that $w = vab$ or $w = vba$ for a palindrome v. Therefore both projections $\pi_{c\to b}(v)$ and $\pi_{c\to a}(v)$ are central words and the suffix either ab or ba is shared by $\pi_{c\to b}(w)$ and $\pi_{c\to a}(w)$

It is our goal to show that the standard conjugate w of any given PWWF word takes the particular form $w = \kappa_f(cba)$ for a suitable kaleidoscope substitution κ_f. This substitution will be obtained from the particular standard morphism f, which generates the syntonic projection of w. But before having a closer look at the syntonic and diatonic morphisms in the subsequent section, we have a look at another type of projection, namely the deletion of all instances of the letter c see also [11]:

Definition 2.9. *For any word $w \in \mathcal{A}^*$, let $\tilde{w} \in \{a,b\}^*$ denote the reduction of w to the word in the letters a and b, which remains after the deletion of all instances of the letter c in w.*

Proposition 2.10. *If $w \in \Omega$ is a PWWF word (with $|w|_b = |w|_c$) ending on the letter b, we can express its (first) diatonic projection by virtue of $\pi_{c\to a}(w) = G(\tilde{w})$.*

Proof. The diatonic projection $v = \pi_{c\to a}(w)$ is of the form $v = a^{e_1}ba^{e_2}b\ldots a^{e_k}b$, where $k = |w|_b$. And the exponents of a satisfy $e_1,\ldots e_k \geq 1$, because in w every corresponding factor contains an instance of c. And for $\tilde{w}$ we have $\tilde{w} = a^{e_1-1}ba^{e_2-1}b\ldots a^{e_k-1}b$, accordingly. And obviously: $G(a^{e_1-1}ba^{e_2-1}b\ldots a^{e_k-1}b) = a^{e_1}ba^{e_2}b\ldots a^{e_k}b$.

Corollary 2.11. *If $w \in \Omega$ is a PWWF word (with $|w|_b = |w|_c$) its reduction $\tilde{w}$ is (non-degenerate) WF.*

Proof. Let w' denote a conjugate of w, which is ending on the letter b. Then $\tilde{w}' = G^{-1}(\pi_{c\to a}(w'))$ must be well-formed, because $\pi_{c\to a}(w')$ is well-formed. And hence, $\pi_{c\to a}(w)$ must also be well-formed.

3   Syntonic and Diatonic Morphisms Revisited

Definition 3.1. *A Sturmian morphism f is called proto-syntonic, iff the letter b occurs in $f(b)$ an even number of times, i.e. iff $|f(b)|_b$ is even. A Sturmian morphism f is called proto-diatonic, iff the letter a occurs in $f(a)$ an even number of times, i.e. iff $|f(a)|_a$ is even.*

Definition 3.2. *A Sturmian morphism g is called syntonic, iff it is conjugate to a standard morphism of the form $f \cdot G$, for a proto-syntonic standard morphism f. A Sturmian morphism g is called diatonic, iff it is conjugate to a standard morphism of the form $G \cdot f$ for a proto-diatonic standard morphism f.*

Proposition 3.3. *In a syntonic morphism f the multiplicity $|f(a)|_b$ of the letter b in the image $f(a)$ is odd.*

Proof. The incidence matrix M_f is of the form

$$M_f = \begin{pmatrix} r & s \\ t & 2u \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} r & r+s \\ t & t+2u \end{pmatrix}.$$

Hence M_f can only be in $GL_2(\mathbb{Z})$ if $t = |f(a)|_b$ is odd.

Proposition 3.4. *Any given PWWF word $w \in \Omega$ is conjugate to a PWWF word w', whose diatonic projection is of the form $\pi_{c \to a}(w') = f(ab)$ for a suitable diatonic standard morphism f.*

Proof. We may choose a conjugate w' of w such that $v = \pi_{c \to a}(w') = f(ab)$ for a standard morphism f. And we know (Proposition 2.3) that $|f(a)|$ must be even. And therefore the left column of the incidence matrix $M_f = \begin{pmatrix} |f(a)|_a & |f(b)|_a \\ |f(a)|_b & |f(b)|_b \end{pmatrix}$ of f must add up to an even number. And from applying Proposition 2.10 to the incidence matrices we conclude that M_f must be of the form

$$M_f = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} r' & s \\ t & u \end{pmatrix} = \begin{pmatrix} r'+t & s+u \\ t & u \end{pmatrix}.$$

And when $(r' + t) + t = r' + 2t$ is even, r' must be even itself. Hence, M_f is the incidence matrix of a diatonic standard morphism.

Definition 3.5. *(Graham construction) Consider a syntonic standard morphism $f : \{a, b\}^* \to \{a, b\}^*$. From the two-letter word $w = f(ab) = w_n w_{n-1} \ldots w_2 w_1$ of length n we construct the three-letter word $w_\prec$ of the same length by replacing every other instance of the letter b with the letter c (starting with b on the right): $w_\prec = v_n v_{n-2} \ldots v_2 v_1 \in \{a, b, c\}^*$:*

$$v_k := \begin{cases} a & \text{if } w_k = a, \\ b & \text{if } w_k = b \text{ and } |w_k \ldots w_1|_b \text{ is odd} \\ c & \text{if } w_k = b \text{ and } |w_k \ldots w_1|_b \text{ is even.} \end{cases}$$

The (r, r, s)-partitioning of $w_\prec$, with $r = |f(a)|$ and $s = n - 2r$, defines a substitution $\kappa_f : \mathcal{A}^* \to \mathcal{A}^*$, namely

$$\kappa_f(c) = v_n \ldots v_{n-r}, \quad \kappa_f(b) = v_{n-r-1} \ldots v_{n-2r}, \quad \kappa_f(a) = v_{n-2r-1} \ldots v_1.$$

The substitution κ_f shall be called the standard kaleidoscope substitution *derived from f.*

Proposition 3.6. *Consider the standard conjugate w of a PWWF word and let f denote the associated standard morphism for which $\pi_{c \to b}(w) = f(ab)$. And let κ_f denote the associated kaleidoscope substitution. Then f is a syntonic standard morphism in the sense of Definition 3.2 and $w = \kappa_f(cba)$.*

Proof. In accordance with Corollary 2.8 the first diatonic projection is the image of a standard morphism, i.e. $\pi_{c\to a}(w) = g(ab)$. In accordance with Proposition 3.4, this is a diatonic standard morphism in the sense of Definition 3.2. In accordance with propositions 7.1 and 7.2 in [8] we can conclude that the standard morphism f is also a syntonic morphism in the sense of Definition 3.2, and therefore proposition 8.2. in [8] implies that w itself is of the form $w = \kappa_f(cba)$.

To exemplify our finding it is useful to inspect two particularly small instances of PWWF words:

Remark 3.7. To make the exceptions *abc* and *bcabc* in [8] was unnecessary.

First we consider (the conjugate of *abc*) $w = bca$. $u = \pi_{c\to b}(bca) = bba$ and $v = \pi_{c\to a}(bca) = baa$. $|u|_b = 2$ with $|u|_b^{-1} = 2 \bmod 3$ and $|v|_b = 1$ with $|v|_b^{-1} = 1 \bmod 3$. And in fact $2 \cdot 2 = 1 \bmod 3$. This is a plagal division, because $2 \cdot (3-2) = (3-1)$ holds in natural numbers. The syntonic mode is $b|ba$ with the syntonic incidence matrix

$$M_u = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 2 \cdot 0 \end{pmatrix} \cdot R$$

and the diatonic mode is $ba|a$ with the diatonic incidence matrix

$$M_v = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = R \cdot \begin{pmatrix} 2 \cdot 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Both matrices have determinant -1 in accordance with the plagal divisions.

Now consider $w = bcabc$. $u = \pi_{c\to b}(bcabc) = bbabb$ and $v = \pi_{c\to a}(bcabc) = baaba$. $u_b = 4$ with $u_b^{-1} = 4 \bmod 5$ and $v_b = 2$ with $v_b^{-1} = 3 \bmod 5$. And in fact $2 \cdot 4 = 3 \bmod 5$. This is a plagal division, because $2 \cdot (5-4) = (5-3)$ holds in natural numbers. The syntonic mode is $b|babb$ with the syntonic incidence matrix

$$M_u = \begin{pmatrix} 0 & 1 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 2 \cdot 1 \end{pmatrix} \cdot R$$

and the diatonic mode is $ba|aba$ with the diatonic incidence matrix

$$M_v = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = R \cdot \begin{pmatrix} 2 \cdot 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Also here, both matrices have determinant -1 in accordance with the plagal divisions.

4   Specific PWWF Superpositions of Specific Syntonic and Diatonic Modes

A conclusion of the above findings is the possibility to construct specific PWWF words as superpositions of specific syntonic and diatonic WF words. To illustrate

this, we inspect the three types of PWWF words of length 7. Table 1 collects all three types of PWWF words of length 7 through their kaleidoscope transformations and displays their corresponding syntonic and diatonic projections along with the decomposition of their incidence matrices.

Table 1. Three types of PWWF words of length 7 with their syntonic and diatonic projections and incidence matrices.

w	$b\|c\|\|bacbc$	$ac\|ba\|\|cab$	$baa\|aca\|\|a$
$\pi_{c\to b}$	$b\|bbabbb$	$ab\|babab$	$baa\|abaa$
	$\begin{pmatrix} 0 & 1 \\ 1 & 5 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 4 \end{pmatrix} \cdot R$	$\begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \cdot R$	$\begin{pmatrix} 2 & 3 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} \cdot R$
$\pi_{c\to a}$	$ba\|baaba$	$aaba\|aab$	$aaabaa\|a$
	$\begin{pmatrix} 1 & 3 \\ 1 & 2 \end{pmatrix} = R \cdot \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}$	$\begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix} = R \cdot \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 5 & 1 \\ 1 & 0 \end{pmatrix} = R \cdot \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix}$

The left side of Fig. 2 conveys the essential combinatorics of the various types of well-formedly and ill-formedly generated scales with up to seven tones per octave. Each of the concentric circular rings represents the circular one-parameter-family of generated scales of a fixed cardinality $n = 2, 3, 4, 5, 6, 7$. Each generator $0 \le g < 1$ is represented by an angle, like the phase of a complex number. The angles are measured counterclockwise and the zero-angle corresponds to the right side of the horizontal middle line. The gray zones within these circular rings represent all g, generating specific well-formed scales of the same type. Each type (=zone) is labeled with a ratio $\frac{k}{n}$, where n denotes the number of scale steps per octave and k the shared number of steps per each instance of the generator. White zones represent generators of ill-formed n-note scales. The axial mirror symmetry along the horizontal middle line exemplifies the redundancy between the generators g and co-generators $1 - g$ Thus, the six gray zones within the outer circle ring, represent essentially three types of well-formed seven-note scales (such as listed in Table 1). The same arrangement of well-formed and ill-formed phase-zones for generated 7-note scales have been transferred, both to the horizontal and vertical coordinate axes of the square on the right side of Fig. 2. Each point of that square thus represents a pair (f, g) consisting of two well-formed modes of the same cardinality. Gray or colored rectangles represent families of such pairs, where both of these modes are well-formed. The colored rectangles represent precisely the zones of syn-diatonic pairs. Each point in a colored rectangle represents a possible superposition of a specific syntonic and a specific diatonic mode.
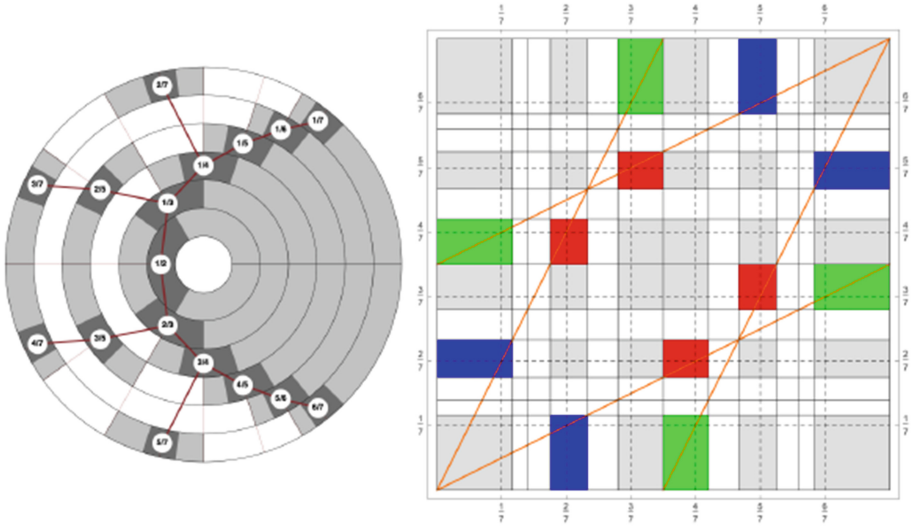
(a) Generated well-formed scales with up to 7 scale steps. Well-formed scales are located in gray zones.

(b) Pairs of specific generated 7-note scales. Pairs, yielding PWWF superpositions of a syntonic and a diatonic scale are located in colored areas. The three colors blue, green and red correspond to the three types of PWWF 7-note scales.

Fig. 2. Types of PWWF 7-note scales and their superpositions.

The specific heights of Zarlino's *modo sintonico diatonico* are $h(a) = Log_2(\frac{9}{8})$, $h(c) = Log_2(\frac{10}{9})$, $h(b) = Log_2(\frac{16}{15})$. The augmented prime $c - b$ has height $h(c - b) = Log_2(\frac{25}{24})$. It is the superposition of the syntonic and diatonic scales with the following step sizes:

$$a_{dia} = \frac{1}{7} + \frac{4}{7}Log_2(\frac{25}{24}) \sim 0.176511, \ b_{dia} = \frac{1}{2}(1 - 5a_1) \sim 0.0587233$$
$$a_{syn} = 2a - a_{dia} \qquad \sim 0.163339, \ b_{syn} = 2c - a_{dia} \ \sim 0.127496.$$

5 Supersymmetries: Exploring a Refined Graham Construction

The term *Graham Construction* (see also Definition 3.5) has recently been mentioned in [1] with reference to the early paper [6]. In Definition 3.5 it refers to a method for turning binary words into ternary words by replacing every other instance of one of the two letters by a third one. The present section relies upon a further refinement of this basic idea.

An obstacle for a full appreciation of the kaleidoscope transformations in [8] is the fact, that they are not closed under composition. This closing section presents an improved approach, where this obstacle can be overcome.

While the letters b and c form a symmetric pair in accordance with the Graham construction of the kaleidoscope transformations from syntonic morphisms, the behavior of the letter a appears to be more autonomous. A higher amount of symmetry can therefore be achieved by adding a companion letter d to the letter a. This new letter can be identified with a whenever it comes to the consideration of a $PWWF$ mode, but it can still have a different transformational behavior. The following example shall illustrate this idea. Concatenating the substitution $\kappa(b) = ba$, $\kappa(c) = ca$, $\kappa(a) = bac$ with itself yields: $\kappa^2(b) = babac$, $\kappa^2(c) = cabac$, $\kappa^2(a) = babacca$. The image $\kappa(bca) = bacabac$ is a PWWF word. But the image $\kappa^2(bca) = babaccabacbabacca$ is not. In particular, the alternating order of the occurrences of the letters b and c is violated. Also modifications of this substitution, such as $\kappa(b) = ca$, $\kappa(c) = ba$, $\kappa(a) = cab$ lead to analogous violations. In this case we have $\kappa^2(b) = bacab$, $\kappa^2(c) = cacab$, $\kappa^2(a) = bacabca$ with the non-PWWF image $\kappa^2(bca) = bacabcacabbacabca$.

There is a possibility to overcome this obstacle by considering suitable transformations on 4-letter words. To understand the underlying idea we recall the following property of the kaleidoscope transformations: Consider a PWWF word w and a kaleidoscope substitution κ such that $w = \kappa(bca)$ in the sense of [8]. Further suppose that κ is obtained from the Graham construction applied to a syntonic standard morphism f on words on the letters a and b. (i.e. the syntonic projection of w is of the form $\pi_{c \to b}(w) = f(ab)$. In this case we observe that $\kappa(c) = E_{(bc)}(\kappa(b))$, where $E_{(bc)}$ denotes the letter exchange automorphism between b and c on $\{a, b, c\}^*$. This is an immediate consequence of (1) the Graham construction, namely to replace ever other instance of b with and instance of c, (2) that $f(a)$ is a prefix of $f(b)$ and (3) that $|f(a)|_b$ is odd (Proposition 3.3). To introduce a fourth letter d as a symmetric companion to the letter a in substitutions ξ, requires two kinds of decisions:

1. Which instances of a in $\xi(b)$, $\xi(c)$ and $\xi(a)$ shall be replaced by d?
2. How should $\xi(d)$ be defined?

Before answering the first question for $\xi(b)$ and $\xi(a)$ in Definition 5.1, we may already extend the equation $\kappa(c) = E_{(bc)}(\kappa(b))$ to

$$\xi(c) := E_{(bc)(ad)}(\xi(b)), \quad \xi(d) := E_{(bc)(ad)}(\xi(as)).$$

In other words, the images $\xi(c)$ and $\xi(d)$ are essentially the same as $\xi(b)$ and $\xi(a)$, but with all instances of b being exchanged with c (and vice versa), and all instances of a being exchanged with d (and vice versa). The following definition refines definition 8.1. from [8]:

Definition 5.1 *(Two-fold Graham construction). Consider a syntonic standard morphism $f : \{a, b\}^* \to \{a, b\}^*$. From the two-letter word $w = f(ab) = w_1 w_2 \ldots w_n$ of length n we construct the four-letter word $w_\prec$ of the same length by replacing every other instance of the letter b with the letter c (starting from c)*

and all immediately following instances of the letter a with d: $w_\prec = v_1 v_2 \dots v_n \in \{a, b, c, d\}^*$:

$$v_k := \begin{cases} c \ \textit{if}\ w_k = b \ \textit{and}\ |w_1 \dots w_k|_b \ \textit{is odd.} \\ d \ \textit{if}\ w_k = a \ \textit{and}\ |w_1 \dots w_k|_b \ \textit{is odd.} \\ b \ \textit{if}\ w_k = b \ \textit{and}\ |w_1 \dots w_k|_b \ \textit{is even.} \\ a \ \textit{if}\ w_k = a \ \textit{and}\ |w_1 \dots w_k|_b \ \textit{is even.} \end{cases}$$

With the help of the (r, r, s)-partitioning of $w_\prec$, with $r = |f(a)|$ and $s = n - 2r$, we define the associated supersymmetry $\sigma_f : \{a, b, c, d\}^* \to \{a, b, c, d\}^*$ *by virtue of:*

$$\sigma_f(c) = v_1 \dots v_r, \qquad \sigma_f(b) := E_{(bc)(ad)}(\sigma_f(c)),$$
$$\sigma_f(a) = v_{2r+1} \dots v_n, \quad \sigma_f(d) := E_{(bc)(ad)}(\sigma_f(a)).$$

Definition 5.2. *For any substitution $f : \{a, b, c, d\}^* \to \{a, b, c, d\}^*$, let $w_f := \pi_{d \to a}(f(cba))$ denote its associated* three-letter-trace.

The following proposition follows from the construction of the supersymmetries:

Proposition 5.3. *Consider a syntonic standard morphism f. Then the word w_{σ_f} is PWWF.*

The major motivation for the present section is the following Result. For a proof and further exploration of this proposition see [10].

Proposition 5.4. *Consider two syntonic standard morphisms $f, g : \{a, b\}^* \to \{a, b\}^*$ together with their associated supersymmetries σ_f and σ_g. Write $f = f' \cdot E \cdot G$ with the associated proto-syntonic standard morphism f'. Then we obtain for the concatenation $\sigma_f \cdot \sigma_g = \sigma_{f'g}$. In particular, this concatenation is a supersymmetry as well.*

References

1. Bulgakova, D.V., Buzhinsky, N., Goncharov, Y.O.: On balanced and abelian properties of circular words over a ternary alphabet. ArXiv:2012:15818 (2021). WORDS 2021
2. Clampitt, D.: Pairwise well-formed scales: structural and transformational properties. Ph.D. diss, SUNY at Buffalo (1997)
3. Clampitt, D.: Mathematical and musical properties of pairwise well-formed scales. In: Klouche, T., Noll, T. (eds.) MCM 2007. CCIS, vol. 37, pp. 464–468. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04579-0_46
4. Clampitt, D., Noll, T.: Modes, the height-width duality, and Handschin's tone character. Music Theory Online **17**(1) (2011). http://www.mtosmt.org/issues/mto.11.17.1/mto.11.17.1.clampitt_and_noll.html
5. De Luca, A.: Sturmian words: structure, combinatorics, and their arithmetics. Theor. Comput. Sci. **183**, 45–83 (1997)
6. Graham, R.L.: Covering the positive integers by disjoint sets of the form $\{[n\alpha + \beta] : n = 1, 2, \dots\}$. J. Combin. Theory (A) **15**, 354–358 (1973)

7. Lothaire, M.: Algebraic Combinatorics on Words. Cambridge University Press, Cambridge (2002)
8. Noll, T., Clampitt, D.: Kaleidoscope substitutions and pairwise well-formed modes: Major-Minor duality transformationally revisited. J. Math. Music **12**(3), 171–211 (2018)
9. Noll, T., Clampitt, D.: Exploring the syntonic side of major-minor tonality. In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O.A. (eds.) MCM 2019. LNCS (LNAI), vol. 11502, pp. 125–136. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_10
10. Noll, T., Clampitt, D.: Supersymmetries for pairwise well-formed words (2022). (in preparation)
11. Paquin, G., Reutenauer, C.: On the superimposition of Christoffel words. Theor. Comput. Sci. **412**, 402–418 (2011)

Algorithms and Modeling for Music and Music-Related Phenomena

Spline Modeling of Audio Signals with Cycle Interpolation

Matt Klassen[✉]

DigiPen Institute of Technology, Redmond, WA 98052, USA
mklassen@digipen.edu
http://www.digipen.edu

Abstract. In this paper we introduce methods to model brief audio signals with cubic splines, presupposing a fundamental frequency f_0. The signal is broken up into cycles using zero-crossings, and each cycle is modeled with a C^2 cubic interpolating spline based on some target number of interpolation points. We reduce the data in the model by reducing the number of key cycles which are used to generate intermediate cycles by the method of interpolation of B-spline coefficients, or cycle interpolation.

Keywords: Spline · Audio · Signal · Interpolation · Cycle

1 Introduction

1.1 Motivation and Background

The modeling project which we summarize in this paper has grown out of collaboration on the research and software project UPISketch [2], which provides a framework for graphical manipulation of sound. The UPISketch software makes use of splines to form models of curves drawn by the user. These graphical gestures are translated through the spline model into musical gestures which modulate elements such as pitch, or fundamental frequency. The spline representation can be thought of as a discrete set of B-spline coefficients, which is an intermediary between the graphical and musical gestures. In this paper we focus on the modeling of timbre with splines, through the evolution of cycles. The notion of cycles, instead of periods, is meant to imply that we expect that realistic waveforms will be at best *almost periodic*, but not exactly. So, the B-spline coefficients will form a discrete representation of timbre.

The use of piecewise polynomials, or splines, to model audio signals has its origins in the familiar piecewise linear audio generators such as square, triangle and sawtooth waveforms. Higher degree piecewise polynomial models, in particular cubic splines, were used as the basis for waveforms in Nick Collins' work [3], in which a single cycle, or period, of the waveform can be manipulated by the

---

movement of control points. Collins developed software which provides the user with the ability to modulate the timbre of the resulting waveform by moving control points. Additionally, Collins uses the analogy of key-framing from animation to "morph" the shape of successive cycles between two key cycles. Our approach is similar but begins with the approximation of a brief audio sample with interpolating cubic splines. In this way the goal is to start with something which models a recorded sound quite closely, but might also be used for synthesis.

In addition to timbral shaping, B-splines have been used in the modeling of envelopes for f_0 synthesis, or fundamental frequency, as in [4]. The authors say in their introduction: "While B-splines have been widely used in computer graphics, very few applications can be found in the field of sound processing". We propose that it is computationally feasible to model audio signals at the cycle level, especially for short segments with a perceived f_0 in the lower end of the audible frequency range, say below 1000 Hz. Although it would also be reasonable to model the envelope with B-splines, in this paper we extract the maximum amplitude value per cycle, from the original audio sample, and apply this as a normalization factor to the B-spline model of each cycle.

As in Collins' work, we also take some motivation from animation with key framing. In this approach, the cycle plays the role of one frame in animation, so we have key cycles which are used to generate the intermediate cycles. Both the choice of interpolation points to represent a key cycle, and the choice of the key cycles, affect the amount of data used to define, and the quality of, the final model. Another analogy with animation is useful: the key frames might come from an artist's imaginative drawing, or they might come from motion capture data. In the same way, one key cycle might come from some idea or model of timbre, or it might come from one cycle which is captured from a particular audio recording. In this paper we explore mostly the latter case.

The spline models of audio data that we propose can also be considered from a few other perspectives:

1. They are *low resolution* models, since they can be used to reconstruct an audio signal from a smaller set of data.
2. They are *time domain* models, differing from typical compression models which work with frequency bands.
3. They are *locally computable* models, which means that the reconstruction of an approximation of the full resolution signal can be computed from a smaller amount of data, the B-spline coefficients, which give a local representation of the audio signal.

This reduction of data, from the original samples, happens in two basic directions: first in the use of some fraction of the number of samples per cycle as interpolation points, and second in the use of some fraction of the cycles as key cycles. Low resolution audio models may also be of interest in realtime rendering of audio where prioritization and level of detail can be important.

A word about regularity: Both for the placement of interpolation points in one cycle, and the selection of key cycles, we begin by choosing regularly spaced points and cycles. This is partly in order to make these choices work nicely

together. For example, if the interpolation points are not of the same number and regularly spaced, from cycle to cycle, then interpolation between cycles becomes much more difficult. There are many schemes, to be considered later, which attempt to take into account the amount of variation in one cycle, changes in local curvature, and other properties which suggest that a non-regular placement of interpolation points would be more optimal.

First, we construct the basic model of an audio sample, cycle by cycle. This means that there is no cycle interpolation. Then we summarize various techniques for doing the cycle interpolation, and follow this by some observations on problems that arise in this process. Finally, we point to further development and questions.

2   The Basic Model

The basic model consists of a cubic spline on each cycle which is used to recompute the same number of samples in the original audio file on each cycle. This preserves any irregularity in lengths of cycles, which is expected. We now describe more details in the context of a basic reference example. An important point is that we consider the original audio data as forming a piecewise linear function of time, so that we can compute the value at any time, not just at samples or integer points. This also allows us to work with zero crossings between samples using linear interpolation.

For a reference sample we have used a guitar pluck with f_0 approximately 440 Hz recorded at standard CD quality (sample rate 44100 Hz and bit depth 16). The file has length 60184 samples, or about 1.37 s. Our first step is to determine all zero-crossings in the audio sample, which we treat as a piecewise linear function of time measured in samples. The zero-crossings are typically not integer points. We will refer to the audio signal values $x(t)$ for time values t at or between samples. The user chooses a fundamental frequency f_0 which then determines an approximate period or cycle length. The choice of such f_0 might be based on spectral analysis or prior knowledge of the audio sample, for instance a recorded instrument sound. This f_0 "guess" is then used iteratively to compute the end of the next cycle by choosing the closest zero crossing. This way cycles are defined using zero crossings closest to the period length guess. In the reference example, we choose to use $f_0 = 220$, one octave below the supposed 440 (guitar pluck on first string fifth fret). This can be seen as a useful guess from the shape of the audio signal graph. With $f_0 = 220$ there are 1531 zero-crossings, of which 302 are used as cycle endpoints. The average number of samples per cycle is quite steady with mean value very close to 200. However, there are some problems that inevitably show up with this method. First, it may be that the evolution of cycles based on zero crossings leads to discontinuities or abrupt changes in cycle shape. We will illustrate this with cycle number 181 in the reference example. Second, it may be that the model improves by keeping the cycle lengths fixed, say at the average value, rather than mimicking the original cycle lengths. We will return to these questions in the context of cycle interpolation.

Continuing with the basic model, once the cycles are determined by the sequence of zeros z_i, $i = 0, \ldots, p$, we do a cubic B-spline fit to the audio sample data. To specify this spline function, say $f(t)$, we need to make various choices. Many of these choices are for reasons of simplicity, but may be revisited or subject to change later. The degree $d = 3$ is the natural and most common default for spline modeling in almost any context. It allows us to achieve second order derivative continuity for the smallest cost. (For details on splines, bases, and knot sequences, we refer to [1]).

When working with one particular cycle, it is useful to normalize the interval to $[0, 1]$. A cubic spline on $[0, 1]$ is a sequence of cubic polynomials $p_i(t)$ on some connected sequence of k subintervals $[u_i, u_{i+1}]$ which can be given by a sequence of *breakpoints* $0 = u_0 < u_1 < u_2 < \cdots < u_{k-1} < u_k = 1$. Rather than work with an explicit polynomial sequence, we use a B-spline basis for the vector space of all such splines $f(t)$ which have continuous second derivative on the interval $[0, 1]$. The sequence of k subintervals can be chosen to have uniform length as a reasonable starting point. The dimension of the vector space V of C^2 cubic polynomial splines on the sequence of k subintervals is then $n = k + 3$. This means that each cubic spline function $f(t)$ is a linear combination of n basis functions:

$$f(t) = c_0 \mathcal{B}_0(t) + \cdots + c_{n-1} \mathcal{B}_{n-1}(t).$$

To solve for such an f as an interpolating spline, we need n points (t, s) with inputs $t \in [0, 1]$ and outputs $x(t) \in [-1, 1]$ obtained from the audio data function. There are many possible B-spline bases for this vector space V, which can be specified by a choice of *knot sequence*, which is a non-decreasing sequence of numbers which includes the subinterval breakpoints $u_1, \ldots, u_{k-1}$ and another $d + 1 = 4$ values on either end. With these conditions we write the knot sequence as $\mathbf{t} = \{t_0, \ldots, t_N\}$ where $t_0 \leq t_1 \leq t_2 \leq t_3 \leq 0$, and $1 \leq t_{N-3} \leq t_{N-2} \leq t_{N-1} \leq t_N$.

We prefer to use the following knot sequence $\mathbf{t}$ which encodes some information at the endpoints of the cycle:

$$\mathbf{t} = \{t_0, \ldots, t_N\} = \{0, 0, 0, 0, \frac{1}{k}, \frac{2}{k}, \ldots, \frac{k-1}{k}, 1, 1, 1, 1\}.$$

Writing the B-spline basis functions associated to $\mathbf{t}$ as

$$\mathcal{B}_0(t), \mathcal{B}_1(t), \ldots, \mathcal{B}_{n-1}(t)$$

we note that $\mathcal{B}_0(0) = 1$ and $\mathcal{B}_{n-1}(1) = 1$, and all the other basis splines vanish at both 0 and 1. Since each cycle is defined based on endpoints which are zeros, we set $c_0 = c_{n-1} = 0$ and solve for the other $n - 2$ coefficients for each cycle. In order to approximate the audio data in one cycle, we find an interpolating cubic spline which matches the (piecewise linear) audio data function $x(t)$ at $n - 2 = k + 1$ specified points. Using the interval $[0, 1]$ we choose the $k - 1$ subinterval breakpoints $(u_i, i = 1, \ldots, k - 1)$ and then add two more points at the middle of the first and last subintervals. Since we do not have derivative information for the audio data, this placement will help to approximate the

derivatives at the ends. These choices are summarized below as the list of data for each spline which represents the model on an interval between two zero crossings:

- degree d = 3
- number of subintervals k
- subinterval sequence: $[u_i, u_{i+1}] = [\frac{i}{k}, \frac{i+1}{k}]$, $i = 0, \ldots, k-1$
- knot sequence: $\mathbf{t} = \{t_0, \ldots, t_N\} = \{0, 0, 0, 0, \frac{1}{k}, \frac{2}{k}, \ldots, \frac{k-1}{k}, 1, 1, 1, 1\}$
- dimension of vector space V of B-spline functions: $n = k + 3 = N - 3$
- B-spline basis functions: $\mathcal{B}_0(t), \mathcal{B}_1(t), \ldots, \mathcal{B}_{n-1}(t)$
- interpolating spline: $f(t) = c_1 \mathcal{B}_1(t) + \cdots + c_{n-2} \mathcal{B}_{n-2}(t)$ $(c_0 = c_{n-1} = 0)$
- input values: $s_0 = 0, s_1 = \frac{1}{2k}, s_i = u_{i-1}, i = 2, \ldots, n-2, s_{n-2} = 1 - \frac{1}{2k}, s_{n-1} = 1$
- target (signal) values: $x(s_0), x(s_1), \ldots, x(s_{n-1})$

Now we can define the basic model to be the sequence of zero crossings z_j, $j = 0, \ldots, p$ and the sequence of spline functions $f_j(t)$ on the interval $[z_j, z_{j+1}]$. We also refer to each of these splines and their associated data as one "cycle" C_j, so that the basic model is the sequence of cycles C_j for $j = 0, \ldots, p-1$. Further, let the B-spline coefficients for cycle C_j be c_i^j, $i = 0, \ldots, n-1$.

The B-spline functions can be defined with divided differences as:

$$\mathcal{B}_i^d(t) = (-1)^{d+1}(t_{i+d+1} - t_i)[t_i, t_{i+1}, \ldots, t_{i+d+1}](t-x)_+^d$$

or using the (de Boor-Cox) recursion formula as

$$\mathcal{B}_i^d(t) = \frac{t - t_i}{t_{i+d} - t_i} \mathcal{B}_i^{d-1}(t) + \frac{t_{i+d+1} - t}{t_{i+d+1} - t_{i+1}} \mathcal{B}_{i+1}^{d-1}(t)$$

with base case:

$$\mathcal{B}_i^0(t) = \begin{cases} 1, & t_i \leq t < t_{i+1} \\ 0, & \text{elsewhere} \end{cases}.$$

We compute values of the B-spline functions, or of the interpolating spline f, with nested linear interpolation, known as the de Boor algorithm:

- (Stage zero): Set $c_i^0 = c_i$ for $i = 0, \ldots, N - d - 1$.
- For $t \in [t_d, t_{N-d})$ set J to be the index so that $t \in [t_J, t_{J+1})$.
- (Stage p): For $p = 1, \ldots, d$, for $i = J - d + p, \ldots, J$: set

$$c_i^p = \frac{t - t_i}{t_{i+d-(p-1)} - t_i} c_i^{p-1} + \frac{t_{i+d-(p-1)} - t}{t_{i+d-(p-1)} - t_i} c_{i-1}^{p-1}$$

- $f(t) = c_J^d$.

Note: to solve for the coefficients c_i, $i = 0, \ldots, n-1$, we set up the interpolation problem as a linear system by requiring that f agrees with the audio output data at each of the n input values. Since the input values are evenly distributed, we are

guaranteed a unique solution, according to the Schoenberg-Whitney Theorem on B-spline interpolation (see [5]).

If we compute the basic model for the reference example with $f_0 = 220$ and $k = 47$ ($d = 3$, and dimension $n = 50$) the model takes about 83 s to compute on a mac laptop in our implementation with JUCE. This model uses about 1/4 of the original samples in the audio file and matches the signal very closely. In terms of audio quality, this model is close enough to the original to say that there are no obviously noticeable artifacts, or distortion.

It is not our intention to present this as a model of data compression, but rather as a starting point to understand how cycles evolve in time during the course of a brief audio sample, and to investigate whether it is possible to simulate this evolution through artificial means which use a smaller set of data. The fact that the model is an adequate representation of an original audio sample can be thought of in the way that the position of a human figure can be approximately captured by sensors which measure the positions of a number of points on the surface of the figure in 3D space.

3   Cycle Interpolation

First, we should specify what it means to interpolate between two cycles, each represented by an interpolating cubic spline, as in the basic model. Keeping the assumption that all cycles still have a common basis of n B-splines, the first obvious approach is to linearly interpolate between pairs of B-spline coefficients with the same index. But it is also reasonable to consider using quadratic or cubic splines. In all of these cases we will refer to this set of splines as *meta-splines*. In order to distinguish these from the splines used to generate each cycle in the basic model, we refer to the latter as *cycle-splines*. The cycle-splines then are modeling the timbre of the audio signal, whereas the meta-splines are controlling the evolution of the timbre, or the way in which one cycle morphs between key cycles.

Second, it is natural to use an approximation of the envelope of the audio signal which is being modeled. Since we are already partitioning into cycles, it is straightforward to extract the maximum absolute value in each cycle, then when computing a cycle-spline to normalize the B-spline coefficients so that they produce the same maximum.

There should be one meta-spline per B-spline coefficient of the cycle-splines. For example, if we are interpolating between key cycles C_{j_1} and C_{j_2}, then to compute the first B-spline coefficient c_0^j of cycle C_j, with $j_1 < j < j_2$, we use meta-spline $g_0(t)$. This meta-spline will have inputs and targets set proportionally to match the sequence of key cycles.

This means that if $g_0(t_{j_1}) = c_0^{j_1}$ and $g_0(t_{j_2}) = c_0^{j_2}$ then we compute for some t_j between t_{j_1} and t_{j_2}:

$$c_0^j = g_0(t_j).$$

Note that in forming the meta-splines it may be that we do not place the inputs for interpolation in a uniform sequence. This arises naturally when we choose to

place key cycles more densely near the beginning than at the end, for instance when modeling an audio sample which has a more varied attack phase and a much less varied sustain and release phase, or tail.

We can now define the Cycle Interpolation Model to be given by the data:

- key cycles $C_{j_0}, \dots, C_{j_{q-1}}$
- B-spline coefficients of key cycle C_{j_r}: $c_i^{j_r}$, $i = 0, \dots, n-1$
- meta-splines $g_i(t)$ with target values c_i^j, $j = j_0, \dots, j_{q-1}$
- max values for envelope α_i, $i = 0, \dots, p-1$
- cycle zeros z_i, $i = 0, \dots, p$

Two important things are still to be determined: 1) how to determine the form of the meta-splines, and 2) how to choose the key cycles.

Regarding the choice of meta-splines: If we use degree 1, or simply linear interpolation between corresponding B-spline coefficients, then the effect is similar to a cross-fade between two key cycles. As Collins points out, this is not just a cross-fade between B-spline coefficients (or control points) but with no other special conditions it is also a cross-fade between audio sample values. But since we are allowing for varying cycle lengths $z_{i+1} - z_i$ and we are also normalizing by the envelope values, the result is not strictly a cross-fade. For simplicity and computational speed we have mostly used the linear case, but it is interesting to also note some properties of the cubic case.

It is tempting to use both cubic meta-splines and also cubic cycle-splines, if for no other reason than elegance and consistency. But there are a few pitfalls. There are limits to the practical computation of linear systems used in solving for each interpolating spline. In the case of cycle-splines this dimension is $n = k+3$, the dimension of the vector space of splines. In the case of the meta-splines the dimension is equal to q, the number of key cycles. If we have say 300 cycles, as with the reference example, and we choose to use half of these as key cycles, then we have dimension $q = 150$ which can be prohibitively large for running experiments with modeling. Fortunately, we are interested in models with fewer key cycles. Another pitfall is that there can be too much variation in the cubic spline model which fits the key cycle data, especially if there are long gaps between key cycles, as may be the case in the tail of a signal. This suggests that there may be reason to explore other types of spline models for the meta-splines.

We also note a type of duality that arises when using cubic splines as both meta-splines and cycle-splines. In particular, the number of meta-splines is n, which is also the dimension for cycle-splines, and the number of cycle-splines is q which is also the dimension of the meta-splines. When generating the output audio data from a model, the spline function $f_j(t)$ is evaluated N_j times according to the number of samples per cycle N_j. Similarly, when producing intermediate B-spline coefficients for non-key cycles, the spline function $g_i(t)$ is evaluated Q times according to the total number of cycles Q to be generated by the model. Since we are interested in models which depend on a small amount of data, this can lead to some special cases. Suppose, although N_j typically varies, that it is constant: $N_j = N$ for all j. Additionally, suppose that $q = n$,

and that $Q = N$, and further that $f_i = g_i$ for all i. This means our model is generated by one set of cubic splines which play the dual roles of cycle-splines and meta-splines. Since $Q = N$, the number cycles coincides with the number of samples per cycle, and we only need to evaluate each spline at the appropriate points once. This assumes that the intermediate cycles are represented by uniform subdivision of the interval $[0, 1]$, just as the samples in one cycle are. It also implies that each meta-spline has the value zero at the ends, since this is required for the cycle-splines. If a cycle interpolation model satisfies all of the above requirements, we call it a *reflexive* model.

Next, we consider different choices of key cycles.

Regular Cycle Interpolation

The first approach is to use regularly spaced key cycles, which we call *regular cycle interpolation*. If the entire set of cycles is C_i, $i = 0, \ldots, p-1$, then we specify some positive integer m and choose cycles C_{jm} as key cycles, $j = 0, \ldots, r$, with $rm \leq M - 1$ and r maximal. There are two variations on this regular cycle interpolation, the first of which is to insist on the last cycle also being included. This allows for the obvious cycle interpolation between C_{rm} and C_{p-1}. The second variation is to not include the last cycle as key, and to simply use C_{rm} repeatedly for the final cycles. In this case, it useful to still follow the envelope of the original audio signal, as indicated above.

Exponential Cycle Interpolation

In order to focus more closely on the attack phase of a short audio sample, such as the reference example, it makes sense to choose more key cycles near the beginning and fewer towards the end, or tail. One useful sequence is to have key cycle indices 0 and then 2^i: $0, 1, 2, 4, 8, \ldots$ which we refer to as *exponential cycle interpolation*.

Fibonacci Cycle Interpolation

Similar to exponential, but a slightly denser pattern is $0, 1, 2, 3, 5, 8, \ldots$, which we call *Fibonacci cycle interpolation* using the recursion $k_{i+1} = k_i + k_{i-1}$.

These are some of the first types of sequences of key cycles that we found natural or useful. In the next section we discuss some of the problems that arise with cycle interpolation.

4 Problems with Cycle Interpolation

There are various problems which can cause spectral deficiencies or audio artifacts with cycle interpolation. We discuss two of these here: missing sub-harmonics, and cycle shape discontinuities.

Missing sub-harmonics:

This is a natural defect of cycle interpolation, since it is not a priori designed to follow the sub-harmonic oscillatory pattern of an original audio waveform, but
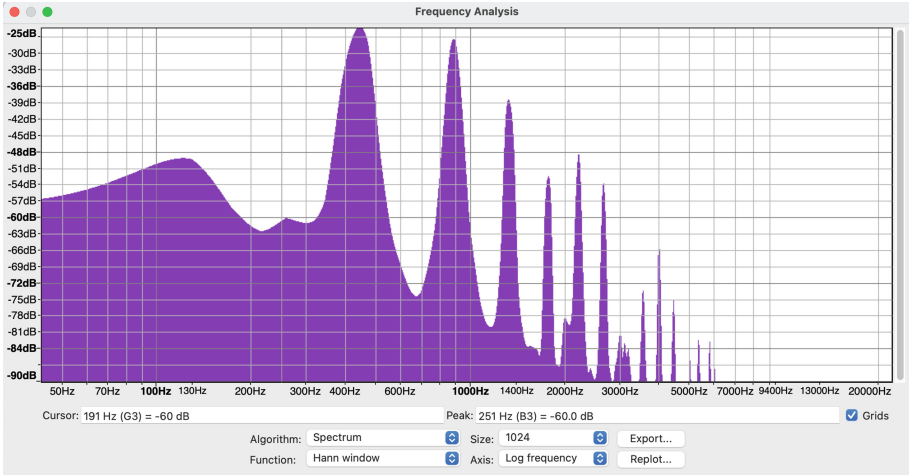
Fig. 1. Spectrum for reference example

rather to fill in the non-key cycles artificially with B-spline interpolation. This can be observed with the reference example. In Fig. 1 is a spectral analysis of a portion of the reference example using Audacity software.

Next, in Fig. 2, is a similar plot for the model using the Fibonacci cycle interpolation with 15 key cycles. Although the frequency guess 220 Hz and higher harmonics are well represented, the sub-harmonic 110 Hz is markedly absent. There are various ways to attempt to address this issue. Rather than delving into DSP post-processing, we prefer in this paper to first attempt to work with the spline model. This is a good opportunity to make a simple adjustment to the model which is surprisingly cheap and effective. We can represent the sub-harmonic with a quadratic spline, represented in the same basis as the cubic cycle-splines. This requires simply solving for a spline on one cycle of the form $A(1-t)t$ where $A > 0$ represents the appropriate amplitude borrowed from the spectrum of the original waveform. We then mimic a sinusoid on successive cycles by alternating between A and $-A$. The B-spline coefficients of this "parabolic sinusoid" can be added directly to the B-spline coefficients of each cycle-spline. Note that this method preserves the allowance of varying cycle lengths as well. The resulting spectrum is shown in Fig. 3.

This example brings up a point about audio rendering and mixing with spline models. Given several spline models, which are to be mixed, one can do the mixing prior to the final rendering, as long as the B-spline bases are consistent. This can lead to a significant reduction in computation. By consistent we mean that the models have the same degree d and the same number of subintervals per cycle k. These could be constant, or could vary but still agree per cycle. The two models do not need to have the same key cycles, but if they do there will be a greater reduction in computation.
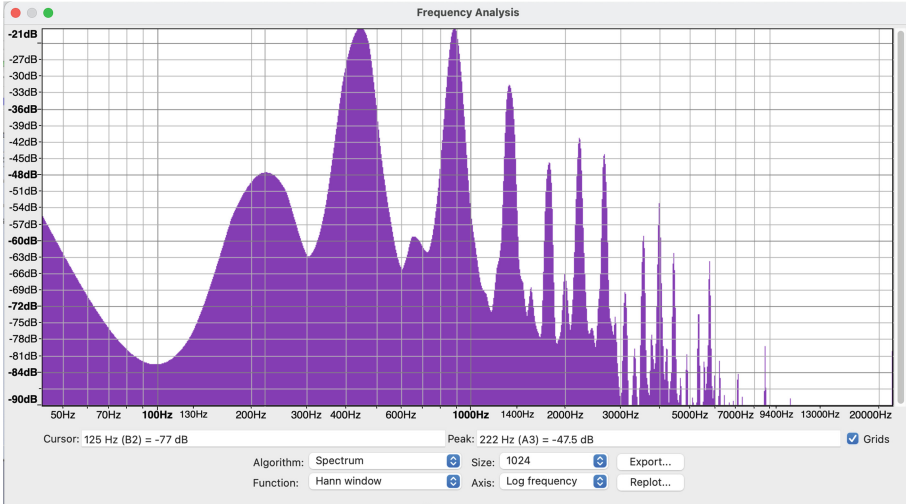
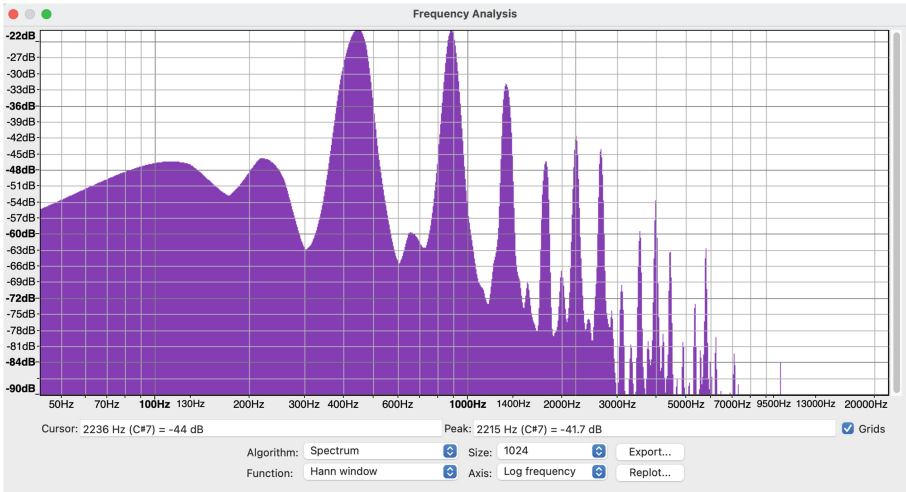**Fig. 2.** Spectrum of model without sub-harmonic

Fig. 3. Spectrum of model with sub-harmonic

Cycle shape discontinuities:

The next type of problem is due to the inherent inaccuracies which occur when attempting to break up an audio sample into cycles. This is illustrated in the reference example, with $f_0 = 220$ Hz (guess at fundamental frequency). With the help of our visualization software for cycle interpolation (written with JUCE) we note that a problem occurs in the region around cycle 180. There are 8 prominent stationary points in each cycle starting at cycle 70. The last of these is a relative minimum occurring just before the right endpoint of each cycle,

which we will call feature A. We can see that feature A begins to drift upwards in cycles 100 through 170, then finally transitions above the time axis by cycle 182. This is illustrated in Fig. 4 around cycle 180. This behavior of the graph is part of the natural evolution of the shape of the curve but it is a problem for cycle interpolation. We will call this problem a *shape discontinuity*, meaning an abrupt change in the shape of consecutive cycles. This is not a property of the original waveform, just of the cycle interpolation model. The original waveform has a smooth progression of cycles, but these cycles are not always best defined by zero crossings.

It is worth noting that this poses no problem for the basic model (without cycle interpolation) since the cycles are each modeled independently. In fact, in the basic model, the cycles could be assigned almost any sequence of lengths and cause no problems. It is also worth noting that the reference example has only one instance of an obvious shape discontinuity of this type.

To see the effect on cycle interpolation, we illustrate in Fig. 5 a graph which also includes the model (in blue) with regular cycle interpolation value $m = 5$. The shape discontinuity causes the B-spline coefficients of the model to change drastically between key cycles. It is no surprise that this causes audible artifacts in the resulting audio.

There are a few ways to mitigate this problem by adjusting the model. The most obvious is to add key cycles at and around the transition point, for instance by adding key cycles 178 through 182. This method does cause reduction in the audible artifacts, but does not really solve the problem which stems from the shape discontinuity. Another approach is to release the waveform at an earlier point from its obligation to match any more key cycles from the original waveform. In other words, this approach removes rather than adds key cycles. One obvious result is that the tail produced by the model will differ significantly from the original, but this may have a small effect on the outcome. For instance, if we use the Fibonacci cycle interpolation model for the reference example, the initial sequence is: $0, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 301$, with 14 key cycles. If we remove the last two then the tail produced by the model is based on cycle 144, and has only 12 key cycles. If these key cycles use about 1/3 of the samples per cycle (say $k = 63$) then we have a good model which removes most of the audible artifacts. The data in this model can be further reduced by setting the cycle length to a constant, say 200 samples. If we use the simplest type of meta-splines (linear interpolation) then the data can be measured by the number of floats needed to reconstruct the model as a fraction of the original audio data. We include in the model the normalization value of one float per cycle. In this case the percent of the original data stored in the model is about 0.183%.

We have found that cycle interpolation has the potential to form good models of recorded sounds with a small fraction of the original sampled data. There are interesting problems which occur in representing sub-harmonics and avoiding errors inherent in the process of breaking up audio signals into cycles. We are encouraged by the results so far and look forward to seeing these techniques used in the design of new sounds in UPISketch.

Fig. 4. Shape discontinuity near cycle 180

Fig. 5. Cycle interpolation near cycle 180

5   Future Work

There are many aspects to explore in more detail. A few of these are:

1) From the sound design perspective, use cycle interpolation together with randomization, or cellular automata, to explore new timbres.

2) Investigate how shape discontinuities arise when cycle interpolation is applied to various musical instrument samples.
3) Use cycle interpolation to model phones for concatenative speech synthesis.
4) Reduce the data in cycle interpolation further, taking into account the prioritization of interpolation points based on discrete curvature.
5) Address the shape discontinuity problem by allowing cycles to be defined at points which are not zero-crossings.

References

1. Klassen, M.: Lecture notes on Bezier curves and polynomial splines. http://azrael.digipen.edu/notes/
2. Bourotte, R., Kanach, S.: UPISketch: the UPIC idea and its current applications for initiating new audiences to music. Organ. Sound **24**(3), 252–260 (2019)
3. Collins, N.: SplineSynth: an interface to low-level digital audio. In: Proceedings of the Diderot Forum on Mathematics and Music, Vienna, pp. 49–61 (1999). ISBN 3-85403-133-5
4. Ardaillon, L., Degottex, G., Roebel, A.: A multi-layer F0 model for singing voice synthesis using a B-spline representation with intuitive controls. In: Interspeech 2015, Dresden, Germany. hal-01251898v2, September 2015
5. de Boor, C.: A Practical Guide to Splines, revised edition. Springer, New York (1980)

Transposition and Time-Scaling Invariant Algorithm for Detecting Repeated Patterns in Polyphonic Music

Antti Laaksonen[(✉)], Kjell Lemström, and Otso Björklund

Department of Computer Science, University of Helsinki, Helsinki, Finland
`ahslaaks@cs.helsinki.fi`

Abstract. This paper presents an algorithm for the time-scaled repeated pattern discovery problem in symbolic music. Given a set of n notes represented as geometric points, the algorithm reports all time-scaled repetitions in the point set. The idea of the algorithm is to use an onset-time-pair representation of music, which reduces the musical problem of finding repeated patterns to the geometric problem of detecting maximal point sets where all points are located on one line. The algorithm works in $O(n^4 \log n)$ time, which is almost optimal because the size of the output can be $\Theta(n^4)$. We also experiment with the algorithm using real musical data, which shows that when suitable heuristics are used to restrict the search, the algorithm works efficiently in practice and is able to find small sets of potentially interesting repeated patterns.

Keywords: Music pattern discovery · Transposition and time-scaling invariance · Geometric algorithms

1 Introduction

In this paper we consider the problem of finding repetitions in Western, equal tempered, polyphonic music. Various kinds of repetitions are frequent both in pop and in classical music. For example, already the structure of a pop song is often based on repetitions such as the usual ABABCBB structure. However, here we concentrate on repetitions taking place at the note level. In classical music one can find various forms of such repetitions, e.g., themes, motifs, imitations, drones, pedals and Alberti basses. Heinrich Schenker [14] stated already in 1954, that repetitions form the basis for music as an art. In general, repetitions make it easier for listeners to detect and remember musical ideas [10].

As the musical repetitions tend to appear in different pitches (different perceived height), it is important to apply transposition invariance in the searching process. If the repetitions are searched for in monophonic music, i.e., music with just one voice, or within a single voice, string-based algorithms can be used efficiently for the task (see e.g. [1,3,4]). Combining transposition invariance with polyphonic music – where repetitions may be scattered around distinct instrument or voices – makes a complex problem setting for which solutions based on

Table 1. A taxonomy of musical pattern matching and repeated pattern discovery problems that shows the gap filled by this paper. The table shows associated literature and the best known solutions for the problems.

Problem	Time complexity
Exact pattern matching	
Plain	$O(nm)$ [15]
Time-scaled	$O(n^2m)$ [5]
Time-warped	$O(n(m + \log n))$ [5]
Partial pattern matching	
Plain	$O(nm \log m)$ [15]
Time-scaled	$O(n^2m^2 \log n)$ [7]
Time-warped	$O(n^2m^2 \log n)$ [8]
Repeated pattern discovery	
Plain	$O(n^2 \log n)$ [11]
Time-scaled	$O(n^4 \log n)$ (new)
Time-warped	$O(n^2 \log n)$ [6]

linear string representations are not sufficient but a geometric point set (pitch-against-time) representation can be used effectively [11]. Furthermore, the data may be acquired by converting audio data into symbolic form where any voice information is lost during the process. In the sequel, we shall concentrate on the more general and complex problem and, therefore, use the point set representation. An example of the point set representation with real musical repetitions that are transposed and time-scaled is given in Fig. 5.

Let us denote by S a two-dimensional point set that represents a musical work (or several musical works concatenated one after another). The number of points in S is denoted by n. If a pattern in S is moved vertically or horizontally, it is *transposed* or *time-shifted*, respectively. A *translated* pattern may be both transposed and time-shifted; we call this the *plain* case. Moreover, a translated pattern may also be *time-warped*, in which case the translated points are time-shifted by some order-preserving, individual amount. *Time-scaling* is a special case of time-warping where time-shifting is applied using a universal multiplication factor to the onset times of the points.

The problem of finding repeating musical patterns is closely related to the musical pattern matching problem where the occurrences of a pattern P of m points are searched for in S. The matching may be *exact* or *partial*, meaning that all or only some of the points in P have to appear in a match, correspondingly. There are algorithms for the pattern matching problem for the plain translated case [13,15], time-scaled case [5,7,13] and time-warped case [5,8]. For the repeated pattern finding problem, there are algorithms for the plain [11] and time-warped case [6], but to our best knowledge there is no published algorithm

for the time-scaled case which is discussed in this paper. Table 1 summarizes the literature and the best known time complexities for all these problems.

Rather interestingly, time-scaled problems seem to be harder than time-warped problems, although the invariance needed for the latter problems is stronger [9]. The reason for this is that we can use dynamic programming to efficiently solve time-warped problems that consists of independent subproblems. However, it seems difficult to use a dynamic programming based approach in time-scaled problems because they have a global scaling factor that affects all subproblems.

As stated above, time-scaling is actually a special case of time-warping. However, the stronger the invariance, the more false positives the searching algorithms produce, i.e., using a time-warping algorithm for our problem would require a separate post-processing phase to filter out the vast majority of time-warped but non-time-scaled occurrences. We expect this to be much more tedious and time consuming than what we present in this paper.

The paper is organized as follows: In Sect. 2, we define the time-scaled repeated pattern discovery problem, show a lower bound for the output size of the algorithm, and present a simple algorithm for the problem. In Sect. 3, we describe our $O(n^4 \log n)$ algorithm for the problem and discuss heuristics that can be used with the algorithm. In Sect. 4, we study the efficiency of the algorithm and the effect of the heuristics. Finally, in Sect. 5, we present our conclusions.

2 Problem Definition

The input for the problem is a two-dimensional point set S that consists of n real-valued points. Each point $p \in S$ corresponds to a musical note: the coordinates $p.x$ and $p.y$ denote the onset time and pitch of the note, respectively.

Given a real number α (time-scaling factor) and a real-valued vector v (translation vector), let

$$\mathrm{MTTP}(\alpha, v) = \{p \mid p \in S, (\alpha p.x + v.x, p.y + v.y) \in S\}$$

denote a *maximal time-scaled translatable pattern* which corresponds to a repeated time-scaled pattern in music. For example, if $\alpha = 2$, the duration of each note is doubled in the repetition.

The problem discussed in this paper is to create an algorithm that discovers and reports all MTTPs in a point set. However, to make the problem more meaningful, we have two restrictions in the search. First, we only report patterns where $\alpha \geq 1$ because if $\alpha \neq 1$, any pattern can be represented in two ways using scaling factors α and $1/\alpha$. In addition, we only consider patterns that have two points with different x values. If this is not the case, the problem would not be well-defined because there would be an infinite number of possible (α, v) pairs.

Note that if $\alpha = 1$, $\mathrm{MTTP}(\alpha, v)$ corresponds to a two-dimensional maximal translatable pattern $\mathrm{MTP}(v)$ that can be found using the SIA algorithm [11].
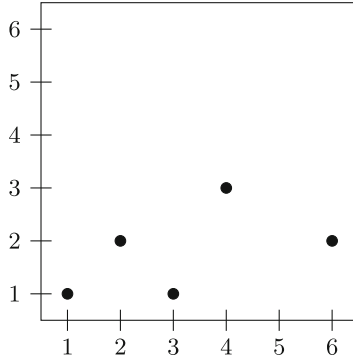
Fig. 1. Example point set

2.1   Example

As an example, consider a point set

$$S = \{(1,1), (2,2), (3,1), (4,3), (6,2)\},$$

shown in Fig. 1. In this case the patterns are as follows:

- MTTP$(3/2, (0, -1)) = \{(2,2), (4,3)\}$
- MTTP$(5/3, (-4, 0)) = \{(3,1), (6,2)\}$
- MTTP$(2, (0,1)) = \{(1,1), (2,2), (3,1)\}$
- MTTP$(5/2, (-4, -1)) = \{(2,2), (4,3)\}$
- MTTP$(3, (-8, 0)) = \{(3,1), (4,3)\}$
- MTTP$(3, (0,0)) = \{(1,1), (2,2)\}$
- MTTP$(5, (-4, 0)) = \{(1,1), (2,2)\}$

Note that there are two ways to produce the patterns $\{(2,2), (4,3)\}$ and $\{(1,1), (2,2)\}$ using two distinct (α, v) combinations.

2.2   Lower Bound

Next we show that the size of the output of the algorithm can be $\Theta(n^4)$, which means that any algorithm for the problem requires $\Omega(n^4)$ time in the worst case.

Consider a point set

$$S = \{(1,1), (2,1), (3,1), \ldots, (n,1)\}$$

where each note has the same pitch. There are $\Theta(n^4)$ ways to select four distinct points $p_1 < p_2 < q_1 < q_2$ such that $p2 - p1 \le q2 - q1$. In each such case we have found two points p_1 and p_2 with distinct x values that have a repetition (points q_1 and q_2) with scaling $a \ge 1$, which means that the algorithm reports them. In addition, for fixed p_1 and p_2, each repetition has a distinct pair (α, v), so p_1 and p_2 are reported separately for each case. Thus, we have found a construction where the size of the output is $\Theta(n^4)$.

2.3   Simple Algorithm

Before presenting our actual algorithm, an interesting question is what would be a simple brute force algorithm for the problem.

Even creating such an algorithm is not trivial, but one possible idea is to go through all subsets of four points $p_1, p_2, q_1, q_2 \in S$ where $p_1.x < p_2.x$ and $q_1.x < q_2.x$. If $p_2.y - p_1.y = q_2.y - q_1.y$, we have found a potential repeating pattern and can define

$$\alpha = (q_2.x - q_1.x)/(p_2.x - p_1.x)$$

and

$$v = (q_1.x - ap_1.x, q_1.y - p_1.y).$$

This corresponds to a nonempty pattern $\mathrm{MTTP}(\alpha, v)$, which will be reported if $\alpha \geq 1$ and the pattern contains two points with different x values. Note that the algorithm can generate a pair (α, v) several times and should only process the first occurrence.

This algorithm works in $O(n^5 \log n)$ time, because there are $O(n^4)$ subsets of four points, and for each subset it takes $O(n \log n)$ time to go through the points in S and find the points that belong to the corresponding pattern. The algorithm can be used to process small data sets, but there is no obvious way to improve it.

3   Algorithm Description

In this section, we describe an $O(n^4 \log n)$ time algorithm for the time-scaled pattern discovery problem. The algorithm uses the onset-time-pair representation presented in [6], and it reduces the problem of finding time-scaled repetitions into the problem of finding all maximal point sets where the points are located on one line.

The algorithm forms for each possible transposition a set C_i ("canvas") which consists of point pairs in S whose pitch interval is i. Since each point pair in a canvas has a constant pitch interval, it is enough to encode the onset times of the pair: $(x_1, x_2) \in C_i$ means that there are two points in S with onset times x_1 and x_2 and interval i. Now, a maximal set of points on the same line corresponds to a a a maximal time-scaled translatable pattern (MTTP) whose time-scaling factor is the slope of the line (see Fig. 2).

More formally, given a set S of n points, the algorithm creates a collection of sets where each set is of the form

$$P_i = \{(a, b) \mid a \in S, b \in S, b.y - a.y = i\},$$

i.e., it contains all note pairs (a, b) whose interval $b.y - a.y$ is a constant i. Then, the algorithm generates for each set P_i an onset-time-pair representation

$$C_i = \{(a.x, b.x) \mid (a, b) \in P_i\}.$$

whose each point consists of x coordinates of a note pair in P_i. This representation is useful when finding time-scaled repetitions, because each repetition corresponds to a set of points that are on the same line. Thus, the remaining problem is to detect all such maximal point sets.

Let us assume that a set C_i consists of k notes. We can find all maximal point sets in $O(k^2 \log k)$ time as follows. We go through all point pairs $p_1, p_2 \in C_i$ where $p_1.x < p_2.x$ and $p_1.y < p_2.y$, and calculate for each such pair two values: a *slope*

$$s = \frac{p_2.y - p_1.y}{p_2.x - p_1.x}$$

of the line defined by the points, and an *offset*

$$z = p_1.y - p_1.x \cdot s,$$

which corresponds to the y coordinate where the line would intersect with the y axis. Then, the triples (s, z, p_1) and (s, z, p_2) are added to a list. After processing all $O(k^2)$ point pairs, we sort the list in $O(k^2 \log k)$ time, and after that, all points that are on the same line are next to each other in the list and we can detect them in $O(k^2)$ time.

Note that there is a direct correspondence between the parameters of a maximal translatable pattern and the parameters of a line in the onset-time-pair representation. Each pattern with parameters (α, v) corresponds to a line in $C_{v.y}$ so that the slope of the line is $s = \alpha$ and offset of the line is $z = v.x$.

Since the total number of points in C_i sets is $O(n^2)$, the algorithm works in $O(n^4 \log n)$ time.

3.1   Example

Consider again the point set

$$S = \{(1, 1), (2, 2), (3, 1), (4, 3), (6, 2)\},$$

shown in Fig. 1. Let us focus on repetitions whose interval is 1 which can be found by creating the sets

$$P_1 = \{((1, 1), (2, 2)), ((1, 1), (6, 2)), ((2, 2), (4, 3)),$$
$$((3, 1), (2, 2)), ((3, 1), (6, 2))\}$$

and

$$C_1 = \{(1, 2), (1, 6), (2, 4), (3, 2), (3, 6)\}.$$

In this case, the points $(1, 2)$, $(2, 4)$ and $(3, 6)$ are on the same line (Fig. 2) with slope 2 and offset 0. This point set corresponds to

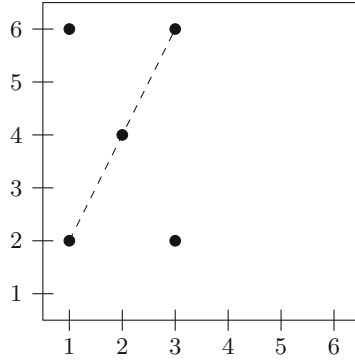$$\mathrm{MTTP}(2, (0, 1)) = \{(1, 1), (2, 2), (3, 1)\}.$$

Fig. 2. Points $(1, 2)$, $(2, 4)$ and $(3, 6)$ are on the same line in the onset-time-pair representation C_1. This corresponds to $\text{MTTP}(2, (0, 1))$.

3.2   Filtering Repetitions

Since the algorithm typically produces a large number of repetitions, we can add heuristics (based on musical knowledge) to improve the results of the algorithm (see e.g. [1,2,12]). In this paper, we consider the following heuristics:

Inter-onset-Intervals.   The inter-onset-intervals between two consecutive notes in a musical pattern cannot be large. Thus, when processing a set C_i in the algorithm, we can choose a constant max_d and only consider pairs $p_1, p_2 \in C_i$ where

$$p_2.x - p_1.x < max_d$$

and

$$p_2.y - p_1.y < max_d.$$

This heuristic can improve the running time of the algorithm, because it can be applied when generating point pairs for the onset-time-pair representation.

Pattern Properties.   Most of the patterns found by the algorithm are usually short, while musically interesting repetitions are likely longer. For this reason, we can choose a constant min_n and only report patterns that have at least min_n notes.

In addition, notes in musically interesting patterns typically have several different pitches, so we can choose a constant min_p and only report patterns that have at least min_p different pitches.

Scaling Factors.   Since we are interested in time-scaled repetitions, we can focus on repetitions where $\alpha \neq 1$. When combined with other heuristics, this can greatly reduce the number of patterns reported by the algorithm.
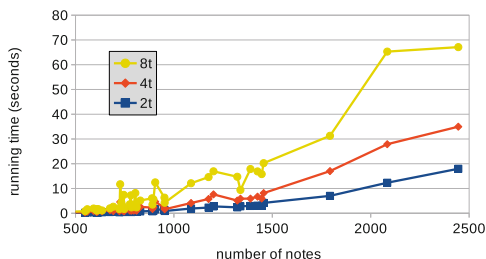
Fig. 3. Efficiency of the algorithm for max_d values $2t$, $4t$ and $8t$ where t denotes the number of ticks in a beat.

4   Experiments

In this section, we study the efficiency and usefulness of our algorithm on real musical data. We have implemented the algorithm in C++, and verified that it produces correct results. The implementation is available in our GitHub repository (https://github.com/c-brahms/time-scaled-repeated).

The data set used in the experiments consists of 48 MIDI files: 24 preludes and fugues from the first book of Bach's *Das wohltemperierte Klavier*. We converted each file into a point set where onset times are MIDI time values (ticks) and pitches are MIDI note numbers. The number of note events in a file ranges from about 400 to 2500.

We conducted the experiments using a 1.8 GHz Intel Core i7 computer in a Linux environment. In all experiments we searched for patterns where $\alpha \neq 1$, i.e., time-scaling is used in the repetition.

4.1   Efficiency

In the first experiment, we measured the running time of the algorithm for each file. It turned out that the general algorithm without the max_d parameter would be too slow for processing the files, so we only consider tests where the max_d parameter is used. The other filtering parameters only control the reporting after the search, so they do not affect the efficiency of the algorithm.

Figure 3 shows the results of the experiment. Three max_d values were used: $2t$, $4t$ and $8t$ where t denotes the number of ticks in a beat. As expected, the greater the max_d value, the slower the algorithm. In most cases, the processing time was less than two seconds, and the maximum processing time was $67\,\mathrm{s}$ for the largest input when the value $max_d = 8t$ was used.

The experiment shows that the algorithm can process real music files efficiently. While the time complexity $O(n^4 \log n)$ of the algorithm could indicate that it is of limited practical use, the max_d parameter considerably improves the practical efficiency of the algorithm. On the other hand, if the max_d parameter is not used, the algorithm can only be used for small inputs.
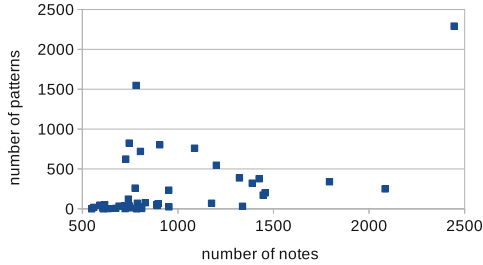
Fig. 4. The number of notes and discovered patterns in each file ($max_d = 4$, $min_n = 8$, $min_p = 5$).

4.2   Pattern Discovery

In the second experiment, we examined the patterns discovered by the algorithm. In this experiment, we used parameters $max_d = 4$, $min_n = 8$, and $min_p = 5$, i.e., the maximum inter-onset-interval is 4 ticks and the pattern must have at least 8 notes and 5 distinct pitches. We chose the parameters so that they filter musically interesting patterns and produce a sufficient number of results.

Figure 4 shows the results of the experiment. The x axis shows the number of notes in each file, and the y axis shows the number of discovered patterns. In most cases, the number of discovered patterns is small and it would be possible to check them all manually.

In almost all discovered patterns, the scaling factor α is one of 4/3, 3/2, 2, 3, and 4. This is not surprising because there are no tempo changes in our data set, and such scaling factors are also expected in real musical repetitions. This suggests that in some cases we could also only focus on finding repetitions whose scaling factors belong to a constant set and achieve an $O(n^2 \log n)$ time algorithm by using a standard pattern discovery algorithm several times. However, such a search would not find repetitions with unexpected scaling factors.

While the used heuristics reduce the number of results, it seems that most of the discovered patterns are still not musically interesting. A possible additional heuristic would be to somehow restrict the intervals between consecutive pattern notes. However, it seems to be difficult to add such a heuristic to the algorithm because the intervals are ignored in the onset-time-pair representation which is the main building block of the algorithm.
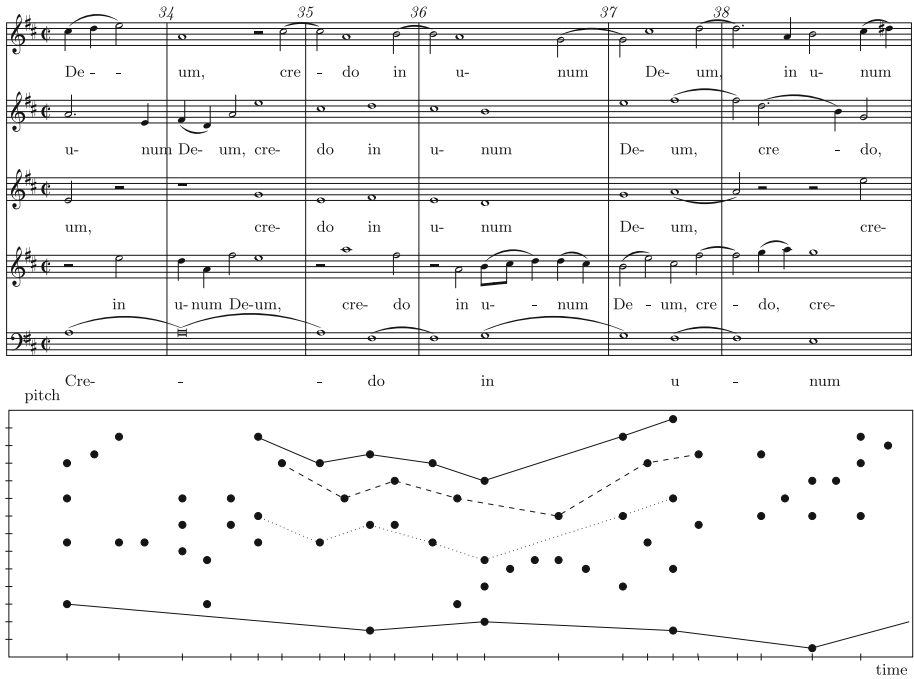
Fig. 5. An extract of the choir parts of J.S. Bach's *Mass in B minor* (*Credo in unum Deum*) and the corresponding, synchronized point set representation where each note-on is represented by a point. The theme is introduced in the beginning of the movement. In the depicted measures the fireworks start: the theme has four partly overlapping occurrences in different keys. Our algorithm detects also the repetition in the bass voice (continued beyond the illustrated area) although it has twice the duration of the other occurrences.

5   Conclusions

In this paper, we have presented an $O(n^4 \log n)$ time algorithm for solving the time-scaled repeated pattern discovery problem in symbolic music. The presented algorithm is more efficient than an $O(n^5 \log n)$ time brute force approach, and it is almost an optimal algorithm because any algorithm for the problem requires $\Theta(n^4)$ time.

Our algorithm can be seen as a missing piece in the taxonomy of pattern matching and discovery algorithms in symbolic music. Exact and time-warped algorithms have been proposed for both pattern matching and discovery, but time-scaled algorithms have only been used in pattern matching. Like in pattern matching, the time-scaled problem is the most difficult also in pattern discovery.

Based on our experiments, our algorithm can be quite efficient in practice when some heuristics are used to filter interesting musical patterns. While there are theoretical constructions where the output size is $\Theta(n^4)$, the number of interesting patterns in actual musical data is much smaller and we can find them

efficiently by implementing the algorithm so that it avoids creating patterns that are not musically meaningful.

It is an interesting question whether the $O(n^4 \log n)$ time complexity of the algorithm could be improved. Since we use onset-time-pair representations and reduce the problem to a geometric problem of detecting all maximal sets of points that are on the same line, one way to that end would be to solve the geometric problem more efficiently. The problem at hand is somewhat easier than the general geometric problem: while the total number of points in onset-time-pair representations is $O(n^2)$, the number of *distinct* x and y values is only $O(n)$. In the future, we will study if we can use this observation to improve the algorithm and process the points in groups that have the same x or y coordinate.

References

1. Cambouropoulos, E.: Musical parallelism and melodic segmentation: a computational approach. Music Percept. **3**(23), 249–268 (2006)
2. Collins, T., Arzt, A., Flossmann, S., Widmer, G.: SIARCT-CFP: improving precision and the discovery of inexact musical patterns in point-set representations. In: Proceedings of the 14th International Conference on Music Information Retrieval (ISMIR 2013), pp. 549–554 (2013)
3. Hsu, J.L., Liu, C.C., Chen, A.: Discovering nontrivial repeating patterns in music data. IEEE Trans. Multimed. **3**(3), 311–325 (2001)
4. Knopke, I., Jürgensen, F.: A system for identifying common melodic phrases in the masses of Palestrina. J. New Music Res. **38**, 171–181 (2009)
5. Laaksonen, A.: Efficient and simple algorithms for time-scaled and time-warped music search. In: Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research, pp. 621–630 (2013)
6. Laaksonen, A., Lemström, K.: Discovering distorted repeating patterns in polyphonic music through longest increasing subsequences. J. Math. Music **15**(2), 99–111 (2021)
7. Lemström, K.: Towards more robust geometric content-based music retrieval. In: Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010), pp. 577–582 (2010)
8. Lemström, K., Laitinen, M.: Transposition and time-warp invariant geometric music retrieval algorithms. In: Proceedings of the 2011 International Conference on Multimedia and Expo (ICME 2011), pp. 1–6 (2011)
9. Lemström, K., Wiggins, G.: Formalizing invariances for content-based music retrieval. In: Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009), pp. 591–596 (2009)
10. Lerdahl, F., Jackendoff, R.: A Generative Theory of Tonal Music. MIT Press, Cambridge (1983)
11. Meredith, D., Lemström, K., Wiggins, G.: Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. J. New Music Res. **31**(4), 321–345 (2002)
12. Nieto, O., Farbood, M.M.: Perceptual evaluation of automatically extracted musical motives. In: Proceedings of the 13th International Conference on Music Information Retrieval (ISMIR 2012), pp. 723–727 (2012)

13. Romming, C.A., Selfridge-field, E.: Algorithms for polyphonic music retrieval: the Hausdorff metric and geometric hashing. In: Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007), pp. 457–462 (2007)
14. Schenker, H.: Harmony. University of Chicago Press, London (1954)
15. Ukkonen, E., Lemström, K., Mäkinen, V.: Geometric algorithms for transposition invariant content-based music retrieval. In: Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003), pp. 193–199 (2003)

On the Memory Usage of the SIA Algorithm Family for Symbolic Music Pattern Discovery

Antti Laaksonen[✉] and Kjell Lemström

Department of Computer Science, University of Helsinki, Helsinki, Finland
`ahslaaks@cs.helsinki.fi`

Abstract. SIA is a fundamental algorithm in symbolic musical pattern discovery, which reports all maximal translatable patterns in a point set. The original SIA algorithm requires $O(kn^2 \log n)$ time and $O(kn^2)$ space, where n is the number of points in the data set, and k is the number of coordinates in each point. In this paper, we present a sweepline algorithm that shares the running time of SIA but requires only $O(kn)$ space, enabling to process of larger data sets without running out of memory. Since SIA is the first step in many pattern discovery tasks, our new algorithm can have a broad impact. For example, we discuss the problem of finding all occurrences of maximal translatable patterns with specific properties. We also compare the algorithms in practice and show that reduced memory usage can benefit real data sets.

Keywords: Pattern discovery · SIA algorithms · Memory usage

1   Introduction

This paper revisits the nearly 20-year-old musical pattern discovery algorithm SIA, [14] which detects all maximal translatable patterns (MTPs) in a geometric point set S corresponding to music data (see Fig. 1). The versatile algorithm has thence seen various modifications for many focal music retrieval applications (see, e.g., [1–5, 7–13, 15]). When the algorithm was published, its $O(kn^2 \log n)$ time complexity was a more restrictive factor than its $O(kn^2)$ space complexity, where n and k are the number of points and the number of coordinates in each point. However, the speed of processors has increased over time, emphasizing the importance of the memory consumption of the algorithm.

We aim to redesign the SIA algorithm to require remarkably less space than the original algorithm. Our new, seepline-based algorithm only needs $O(kn)$ space, which significantly improves the original $O(kn^2)$ space complexity. The original algorithm is based on sorting a vector table that consists of $O(n^2)$ vectors, but it turns out that we can avoid this by using a binary heap structure that only contains a small number of vectors at a time.

An extension of SIA is the SIATEC algorithm [14] which finds all occurrences of MTPs and classifies them in translationally equivalence classes (TECs) in
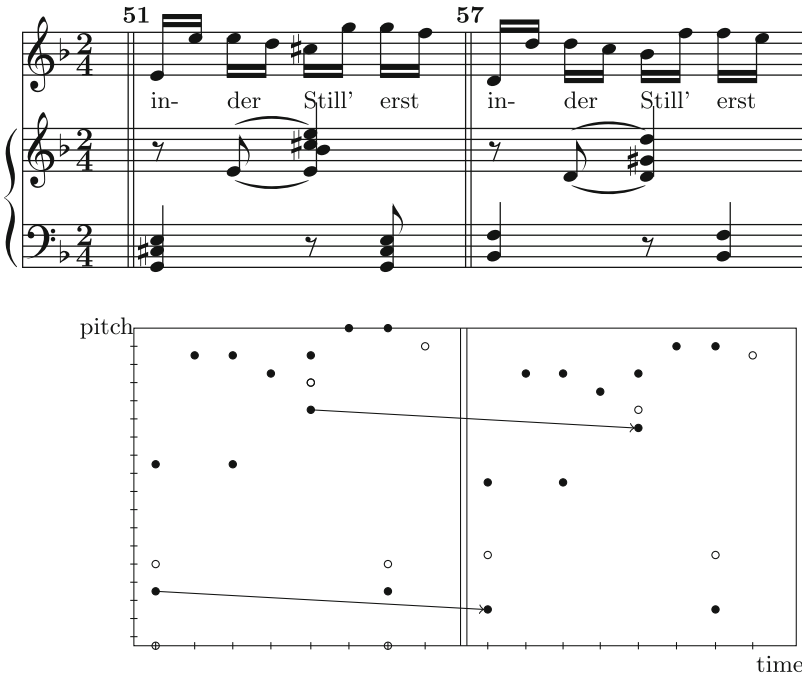
Fig. 1. Two excerpts from Schubert's song cycle *Winterreise* (song *Rast*, measures 51 and 57), in common music notation (above) and the corresponding geometric point set representation (below) where each note is represented as a point whose x coordinate is the onset time and y coordinate is the pitch. The two representations are synchronized, i.e., the points below are matched with the corresponding notes above. The maximal translatable pattern (MTP) is illustrated as filled circles: all these points in the first excerpt can be translated to a point in the second extract by using the same translation vector (two instances of the translation vector are shown as examples).

$O(kn^3)$ time. Also, in this case, the original space complexity of the algorithm is $O(kn^2)$, but we can improve it using our new method. For example, if we want to report all distinct patterns that have specific properties, we can create an algorithm that works in $O(k(n + t))$ space where t denotes the total number of points in reported patterns. Such an algorithm is useful when t is small compared to n^2, saving a significant amount of memory.

Many algorithms in the literature are based on SIA and SIATEC and can benefit from our sweepline method. For example, the COSIATEC and SIATEC-Compress algorithms [11,13] greedily compute a compressed encoding of a point set that, in some cases, may be improved by using set-covering techniques [5,12]. Although general-purpose compression algorithms may also be adapted to this end, COSIATEC has been shown to outperform them [10]. Moreover, many studies have extended the SIA framework, such as similar techniques in musical pattern matching [8,9,15], finding distinctive musical patterns [3], improving the precision of the discovery [1,2], finding repeated structures in audio music [4], and including the time-warping invariance for the pattern discovery problem [7].
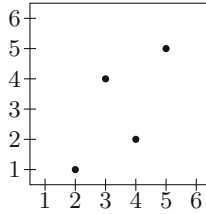
Fig. 2. An example data set that consists of $n = 4$ points. One of the MTPs in this set is $((2,1),(4,2))$ that can be translated using the vector $(1,3)$.

The structure of the rest of the paper is as follows: In Sect. 2, we define the problem that the SIA algorithm solves and describe the original algorithm. Section 3 presents our sweepline algorithm that produces the same results as SIA but only requires a small amount of memory. In Sect. 4, we show how we can create a SIATEC style algorithm that benefits from the new way to calculate MTPs. Finally, in Sect. 5, we present our conclusions.

2   Background

We represent a musical work as a point set D of n points. Each point consists of k coordinates. In a typical setting, $k = 2$ and the coordinates denote a musical note's onset time and pitch. We assume that the points have some order and can be accessed using the notation $D[1], D[2], \ldots, D[n]$.

Given two points $a, b \in D$, the translation vector $v = b - a$ translates the point a to the point b. A *maximal translatable pattern* (MTP) is defined as

$$\mathrm{MTP}(v, D) = \{d \mid d \in D, d + v \in D\}$$

which means that each MTP is a maximal set of points in D that can be translated using a vector v so that they still belong to D. In music, each MTP corresponds to a potential repetition of a melody or other musical pattern.

A typical task in musical pattern discovery is to find all nonempty MTPs in a point set. To avoid reporting each pattern twice, we here only consider MTPs such that the translation vector v satisfies $v > -v$. For example, the vectors $(1,-2)$ and $(-1,2)$ correspond to the same MTP and only the first vector is reported because $(1,-2) > (-1,2)$. Note that we use the lexicographic order to compare points and vectors throughout the paper.

As an example, let us consider the following data set of $n = 4$ points

$$D = \{(2,1), (3,4), (4,2), (5,5)\}$$

(shown in Fig. 2). Our goal is to find the following MTPs:

– $\mathrm{MTP}((1,-2), D) = \{(3,4)\}$
– $\mathrm{MTP}((1,3), D) = \{(2,1), (4,2)\}$

- MTP$((2,1), D) = \{(2,1), (3,4)\}$
- MTP$((3,4), D) = \{(2,1)\}$

For example, the vector $(1,3)$ translates the points $(2,1)$ and $(4,2)$ because $(2,1) + (1,3) = (3,4)$ and $(4,2) + (1,3) = (5,5)$. In this case, the total number of points in MTPs is 6 and in general the number of points can be calculated using the formula $n(n-1)/2$.

The SIA algorithm [14] can construct all MTPs for a data set. The usual way to implement the algorithm is as follows. The algorithm first sorts D and then creates a *vector table*

$$V = \{(D[j] - D[i], i) \mid 1 \le i < j \le n\}$$

whose each element consists of a translation vector and an index to the point that can be translated using that vector. Then, the algorithm sorts V, after which all points that belong to the same MTP are next to each other. Finally, the algorithm scans V from left to right and reports the discovered MTPs.

In our example point set, the sorted vector table is

$$V = \{((1,-2), 2), ((1,3), 1), ((1,3), 3),$$
$$((2,1), 1), ((2,1), 2), ((3,4), 1)\}$$

revealing the MTPs corresponding to the vectors $(1,-2)$, $(1,3)$, $(2,1)$, and $(3,4)$.

The number of elements in the vector table is $O(n^2)$, and each element requires $O(k)$ space. Thus, the algorithm works in $O(kn^2 \log n)$ time[1] and $O(kn^2)$ space.

3   Sweepline Algorithm

Our goal is to reimplement SIA to use a smaller amount of memory. So, we cannot create a vector table or use another data structure having a quadratic number of elements. We will adopt the sweepline based idea of the $P2$ algorithm [15] to go through efficiently the contents of a lexicographically sorted vector table without actually creating the table. $P2$ was designed for the problem of finding all partial matches of a musical pattern in a data set.

The idea is to use a binary heap data structure supporting two operations:

- `add(x)`: add an element x to the heap
- `fetch()`: return and remove the smallest element

Both the operations work in logarithmic time.

We create a heap whose each element is an index pair (a, b) where $a < b$. Each pair (a, b) corresponds to the translation vector $D[b] - D[a]$. The elements in the heap are ordered so that $(a, b) < (c, d)$ exactly when $(D[b] - D[a], a, b) < (D[d] - D[c], c, d)$ lexicographically, i.e., translation vectors determine the order of the elements but they are not stored in the heap.

---

[1] The logarithmic factor comes from the need of sorting the $O(n^2)$ vectors.

We initialize an empty heap by adding, for $i = 1, 2, \ldots, n-1$, an element

$$(i, i+1)$$

to the heap. Each such element corresponds to a sweepline. After that, on each step we fetch the smallest element (a, b) from the heap and process it. If $b < n$, we then add a new element

$$(a, b+1)$$

to the heap, which means that the corresponding sweepline proceeds, and otherwise, we do not add any new elements. Finally, we stop when the heap is empty. Using this algorithm, we can go through all point pairs in the order of their translation vectors using only a small amount of memory, as the total number of elements in the heap is $O(n)$.

3.1   Correctness and Analysis of the Algorithm

Let us denote by H the set of elements in the heap, and by $\overline{H}$ the set of (unprocessed) elements still to be added to the heap. Moreover, $min\{S\}$ denotes the minimum element in set S. An index pair $(a, b-1)$ is the *predecessor* of (a, b) and any pair $(a, b-i)$, where $i \geq 1$, is an *ancestor* of (a, b).

As the heap always returns the index pair whose translation vector is minimal, the points translatable by the same translation vector are returned one after another[2] from the heap. Formally, the algorithm works correctly if the following holds:

1. Each index pair (a, b) is added to the heap at some point,
2. No index pair (a, b) is added twice to the heap,
3. $min\{H\} \leq min\{\overline{H}\}$ holds at all time.

Let us next consider these one by one.

1. Initially, the heap consists of index pairs $(i, i+1)$, for $i = 1, 2, \ldots, n-1$. Let us assume that at some point the heap contains a pair $(j, j+k)$, where $1 \leq j \leq n-1$, and $1 \leq k \leq n-j+1$, and argue that the pair $(j, j+k+1)$, where $j + k < n$, will be added to the heap at some point. We fetch one element from the heap on each step, and at some point $(j, j+k)$ becomes the returned one. At this point, the algorithm adds $(j, j+k+1)$ to the heap. So, all pairs are added to the heap at some point.
2. Assume that (a, b) is the smallest index pair that is added twice to the heap. Initially each index pair in the heap is distinct. Then, since (a, b) is added after removing $(a, b-1)$, this would mean that $(a, b-1)$ is also added twice, which is a contradiction. Thus, the algorithm adds no pair twice to the heap.

---

[2] Thus, counting multiplicity of a translation vector can be done by observing how many times each translation vector is returned in a row.

3. Obviously $min\{H\} < min\{\overline{H}\}$ holds after the initialization. It remains to be shown that no updating iteration (fetching the smallest element and adding another one) of the heap can break its soundness. Let us make a counter-assumption that after an iteration there is an element $(a, b) \in \{\overline{H}\} < min\{H\}$. Due to the construction of the algorithm, we know that its predecessor was not the one that was fetched from the heap, that is, either this very predecessor or some other ancestor of the element resides in the heap. Thus, $(a, b) > min\{H\}$, which contradicts the counter-assumption.

Theorem 1. *Our sweepline algorithm works correctly.*

Proof. A direct consequence of using the heap and that 1, 2 and 3 above hold. □

Let us next compare SIA and our sweepline algorithm. In SIA, the vector table V consists of elements of the form $(D[b] - D[a], a)$. In our algorithm, each such element corresponds to an element (a, b) that is first added to the heap and later fetched from the heap. Since we always fetch the smallest element, the processing order of the elements is the same in both algorithms. Thus, our new algorithm correctly finds all MTPs like the SIA algorithm.

Our algorithm works in $O(kn^2 \log n)$ time because it processes $O(n^2)$ elements and adds each element to the heap and later removes it. Since the algorithm first adds $n - 1$ elements to the heap and then always fetches one element and adds at most one element, the total number of elements in the heap is always at most $n - 1$ and the algorithm needs $O(kn)$ space. Thus, our sweepline algorithm has the same time complexity as SIA, but uses much less space.

3.2   Example

As an example, consider again the data set

$$D = \{(2, 1), (3, 4), (4, 2), (5, 5)\},$$

shown in Fig. 2. In this case, the heap initially consists of the elements

$$\{(2, 3), (1, 2), (3, 4)\}.$$

The smallest element is $(2, 3)$, which is removed and a new element $(2, 4)$ is added to the heap. After that, the heap has the elements

$$\{(1, 2), (3, 4), (2, 4)\},$$

the smallest element $(1, 2)$ is removed, and a new element $(1, 3)$ is added to the heap. This process continues until all elements have been generated.

4   Classifying Patterns

Two patterns A and B are called *translationally equivalent* ($A \equiv B$) if there is a vector v such that

$$\{x + v \mid x \in A\} = B.$$

For example,

$$\{(2, 1), (4, 2)\} \equiv \{(3, 4), (5, 5)\},$$

because we can choose $v = (1, 3)$. Given a pattern $P \in D$, its *translational equivalence class* (TEC) consists of all translationally equivalent patterns in D:

$$\mathrm{TEC}(P, D) = \{X \mid P \equiv X, X \in D\}$$

4.1   Finding All TECs

Next we consider the problem of finding the TEC of each MTP in a data set. For example, in our data set

$$D = \{(2, 1), (3, 4), (4, 2), (5, 5)\}$$

the TEC of $\mathrm{MTP}((1, -2), D)$ is

$$\{\{(2, 1)\}, \{(3, 4)\}, \{(4, 2)\}, \{(5, 5)\}\}$$

and the TEC of $\mathrm{MTP}((1, 3), D)$ is

$$\{\{(2, 1), (4, 2)\}, \{(3, 4), (5, 5)\}\}.$$

A classical algorithm for solving the problem is SIATEC [14] which works in $O(kn^3)$ time and uses $O(kn^2)$ space. Next we show how our new algorithm can be used to discover TECs using less memory.

Now that we can generate all MTPs, the remaining task is to determine for each pattern P the class $TEC(P, D)$. This can be done using a pattern matching algorithm (e.g. $P1$ [15]) that finds all occurrences of P in D in $O(knm)$ time and $O(k(n + m))$ space, where m is the number of points in P. Since the total number of points in MTPs is $O(n^2)$ and the number of points in a pattern is $O(n)$, the TECs of all MTPs are generated in $O(kn^3)$ time and $O(kn)$ space. Thus, this new algorithm works as fast as SIATEC but requires only a small amount of memory.

4.2   Finding Distinct TECs

We may also require that each distinct TEC must be generated only once. For example, in our data set $\{(2, 1)\}$ and $\{(3, 4)\}$ are distinct MTPs, but they belong to the same TEC, so we may not want to generate a TEC for both of them. In particular, the original SIATEC algorithm guarantees that each generated TEC is distinct. While the number of distinct TECs is usually smaller than the number

of MTPs, it turns out that it is more difficult only to generate the distinct TECs than to generate the TEC of each MTP.

We can uniquely represent a TEC by choosing one of its patterns P and creating a *normalized* pattern

$$P' = \{d - x \mid d \in P\}$$

where x is the smallest point in P. For example, patterns $\{(2,1),(3,4)\}$ and $\{(4,2),(5,5)\}$ belong to the same TEC and their normalized pattern is $\{(0,0),(1,3)\}$. Note that the normalized pattern always contains the point whose all coordinates are zero, which can be removed to create a more compact representation.

To only generate distinct TECs, we can use a trie data structure that keeps track of the TECs found so far. We need the following operations:

– add(P'): add P' to the data structure
– check(P'): check if P' is in the data structure

A trie supports both the operations in $O(m \log n)$ time where m is the number of points in P'. Since the total number of points in MTPs is $O(n^2)$, the trie requires $O(kn^2)$ space. Thus, using our new algorithm and a trie, we can find all distinct TECs in $O(kn^3)$ time and $O(kn^2)$ space, like in the SIATEC algorithm.

In this setting, it may seem that our new algorithm does not help much. While we can generate all MTPs and find all distinct TECs using a small amount of memory, we still need a quadratic amount of memory for storing the representatives of the TECs to avoid reporting any TEC more than once. However, in practice, we often do not want to report all TECs but only some of them: a selection of musically interesting patterns. For this reason, while generating the MTPs, we can *filter* patterns that look interesting and find TECs only for them.

Consider a situation where filtering is used and the number of points in interesting patterns is t. Here we need only $O(kt)$ space for storing the representatives of distinct TECs, and the total memory usage of the algorithm is $O(k(n + t))$. Thus, the benefit of using our new algorithm is that memory is only needed for the patterns that are actually reported after filtering, not for all possible MTPs. In practice t can be much smaller than the number of points in all MTPs.

5   Experiments

We have implemented both the original SIA algorithm and our sweepline algorithm using C++ and verified that they produce the same results. Our implementations are available in our GitHub repository https://github.com/c-brahms/sia-memory.

Next, we describe the results of two experiments where we study the efficiency of the algorithms and the effect of filtering when searching for patterns. In both the experiments, we used a data set based on the preludes and fugues from the first book of Bach's *Das wohltemperierte Klavier*, concatenated one after another. The total number of notes in the data set is 40,000.
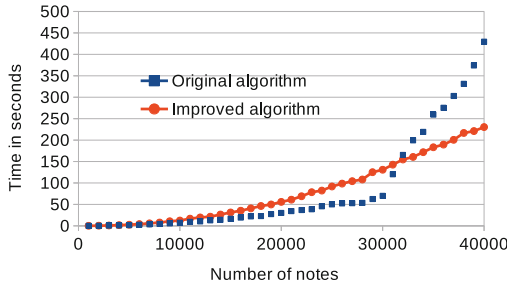
Fig. 3. The efficiency of the algorithms. The original algorithm is faster for small data sets, but it becomes slow for larger data sets because of its high memory usage.
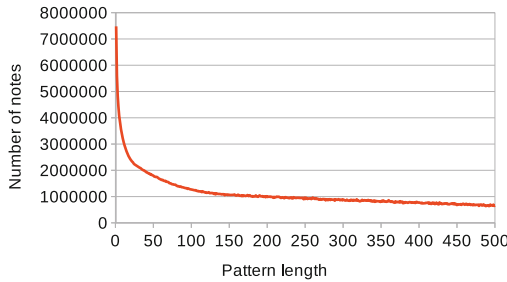

Fig. 4. The number of notes in patterns, length ranging in $[1, 500]$. Most of the patterns are short and the number of patterns decreases as a function of the pattern length.

5.1   Efficiency

In the first experiment, we compared the running times of the algorithms. We conducted the experiment in a Linux environment on a laptop computer with a 1.8 GHz Intel Core i7 processor and 4 GB RAM.

We created test inputs where n is between 1,000 and 40,000 and each input contains n first notes of the data set. The algorithm was given the test input, and it counted and printed the number of MTPs. In each test, we ran both algorithms five times and calculated the averages of their running times.

Figure 3 shows the results of the experiment. When n is small, the original SIA algorithm is about twice as fast as our algorithm. However, when n gets larger, the original algorithm becomes slow and our algorithm is more efficient. In our environment, the new algorithm runs faster when $n \geq 32,000$.

The probable reason for the results of the experiment is that SIA runs out of memory and begins to use swap memory which is slow. The number of pairs in the vector table is $n(n-1)/2$ and each element is a pair of two values $(2 \cdot 4$ bytes in our implementation). Thus, the memory needed for the vector table when $n = 32,000$ is about 4 GB which equals the amount of RAM in our environment.

For the largest input $n = 40{,}000$ the running times of SIA and the sweepline algorithm are 430 and 230 s, respectively. The running time of the sweepline algorithm seems to be stable at about n^2c seconds where $c = 1/(7 \cdot 10^6)$.

5.2 Filtering Patterns

In the second experiment, we studied the effect of filtering while generating the patterns. When generating distinct TECs, filtering can reduce the memory usage because the number of patterns stored in the memory becomes smaller. We focus on filtering based on the pattern length.

We used the entire data set of $n = 40{,}000$ notes in the experiment and calculated for each $k = 1, 2, \ldots, n - 1$ the number of notes in MTPs whose length is k. Figure 4 shows the distribution of notes for k between 1 and 500. For example, there are 818,553 MTPs of length 5, so the number of notes for $k = 5$ is $5 \cdot 818{,}553 = 4{,}092{,}765$.

The total number of notes in all MTPs is $n(n-1)/2$. In this data set, 5.5%, 25.8% and 70.1% of the notes belong to MTPs where $k \leq 10$, $k \leq 100$ and $k \leq 500$, respectively. As Fig. 4 shows, there is a peak in the distribution for small pattern lengths, and after that the number of notes slowly decreases when the pattern length increases.

If we use our algorithm and only report distinct TECs of patterns with specific lengths, this can significantly reduce the memory usage of the search. For example, if we only report MTPs whose length is between 10 and 100, we only need memory for about 20% of all notes in MTPs.

Also other filtering methods could be used. For instance, a technique that only focuses on patterns with musically interesting properties can reduce memory usage even further. A topic for future research would be to evaluate the results with different filtering methods to find ways to reduce the memory usage of the algorithm without ruling out potentially interesting patterns.

6   Conclusions

In this paper, we redesigned the SIA algorithm so that it uses much less space. Our new sweepline algorithm reduces the space complexity from $O(kn^2)$ to $O(kn)$ allowing the procession of large data without running out of memory. Our new algorithm can be used to create a SIATEC style algorithm that requires $O(k(n + t))$ space where t is the number of points in reported patterns.

The reduced memory usage of our algorithm is important both in theory and practice. Our experiment shows that the original SIA algorithm performs poorly on large data sets when it runs out of memory and needs to use swap memory to store the vector table. However, our new algorithm only requires a small amount of memory and has a stable running time also when working with large data sets. Since the SIA algorithm family is used in various music information retrieval applications, they can also benefit from our sweepline algorithm.

In our experiment, we had 4 GB of RAM and SIA becomes slow when $n \approx$ 32,000. If we had more RAM, the algorithm would probably perform better but still become slow when the size of the vector table exceeds the amount of RAM. For example, for 8 GB and 16 GB of RAM, the input size where the algorithm runs out of memory would be about 45,000 and 64,000, respectively.

Could it be possible to improve the SIA algorithm further? At least it seems difficult to improve the time complexity $O(kn^2 \log n)$ or space complexity $O(kn)$ of the algorithm. Any algorithm that reports translation vectors in sorted order (like both algorithms discussed in this paper) also solves the Sorting X + Y problem [6] (with $k = 2$), and it is conjectured that this problem cannot be solved faster than in $O(n^2 \log n)$ time. In theory, however, there could be a possibility to dispose the logarithmic factor of the time complexity if the MTPs could somehow be discovered without sorting the translation vectors. Moreover, it does not seem easy to reduce the space complexity without affecting the time complexity.

However, a way to improve the practical performance of the algorithm could be to use parallel features of modern processors, such as vector instructions and multiple threads. This could be especially useful in applications where only a small number of patterns are reported and the output size is not large.

References

1. Collins, T.: Improved methods for pattern discovery in music, with applications in automated stylistic composition. Ph.D. thesis, Faculty of Mathematics, Computing and Technology, The Open University, Milton Keynes (2011)
2. Collins, T., Arzt, A., Flossmann, S., Widmer, G.: SIARCT-CFP: improving precision and the discovery of inexact musical patterns in point-set representations. In: Proceedings of the 4th International Society for Music Information Retrieval (ISMIR 2013), pp. 549–554 (2013)
3. Collins, T., Arzt, A., Frostel, H., Widmer, G.: Using geometric symbolic fingerprinting to discover distinctive patterns in polyphonic music corpora. In: Meredith, D. (ed.) Computational Music Analysis, pp. 445–474. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-25931-4_17
4. Collins, T., Böck, S., Krebs, F., Widmer, G.: Bridging the audio-symbolic gap: the discovery of repeated note content directly from polyphonic music audio. In: Audio Engineering Society Conference: 53rd International Conference: Semantic Audio (2014)
5. Forth, J.: Cognitively-motivated geometric methods of pattern discovery and models of similarity in music. Ph.D. thesis, Department of Computing, Goldsmiths, University of London (2012)
6. Harper, L., Payne, T., Savage, J., Straus, E.: Sorting $X+Y$. Commun. ACM **18**(6), 347–349 (1975)
7. Laaksonen, A., Lemström, K.: Discovering distorted repeating patterns in polyphonic music through longest increasing subsequences. J. Math. Music **15**(2), 99–111 (2021)
8. Lemström, K.: Towards more robust geometric content-based music retrieval. In: Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010), pp. 577–582 (2010)

9. Lemström, K., Laitinen, M.: Transposition and time-warp invariant geometric music retrieval algorithms. In: Proceedings of the 2011 International Conference on Multimedia and Expo (ICME 2011), pp. 1–6 (2011)
10. Louboutin, C., Meredith, D.: Using general-purpose compression algorithms for music analysis. J. New Music Res. **45**(1), 1–16 (2016)
11. Meredith, D.: COSIATEC and SIATECCompress: pattern discovery by geometric compression. In: Music Information Retrieval Evaluation Exchange (MIREX Competition on "Discovery of Repeated Themes & Sections") (2013)
12. Meredith, D.: Music analysis and point-set compression. J. New Music Res. **44**(3), 245–270 (2015)
13. Meredith, D.: RecurSIA-RRT: recursive translatable point-set pattern discovery with removal of redundant translators. In: Cellier, P., Driessens, K. (eds.) ECML PKDD 2019. CCIS, vol. 1168, pp. 485–493. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-43887-6_42
14. Meredith, D., Lemström, K., Wiggins, G.: Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. J. New Music Res. **31**(4), 321–345 (2002)
15. Ukkonen, E., Lemström, K., Mäkinen, V.: Geometric algorithms for transposition invariant content-based music retrieval. In: Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003), pp. 193–199 (2003)

A Proposal to Compare the Similarity Between Musical Products. One More Step for Automated Plagiarism Detection?

Aarón López-García[1]([✉]) , Brian Martínez-Rodríguez[2]([✉]) ,
and Vicente Liern[3]([✉])

[1] Department of Computer Science, Universitat de València, Valencia, Spain
logara8@alumni.uv.es
[2] Universidad Internacional de la Rioja (UNIR), Logroño, La Rioja, Spain
brian.martinez@unir.net
[3] Dep. Matemáticas para la Economía y la Empresa, Universitat de València,
Valencia, Spain
vicente.liern@uv.es

Abstract. In previous works, the authors presented a measure of similarity between melodies by identifying them with sequences of ordered vectors and using a clustering process based on fuzzy logics. Along the same line, we propose a measure of musical similarity between fragments of digital audio. We present the SpectroMap algorithm that allows us to detect the local maxima of the audio spectrogram representation (also known as constellation map) and we compared the similarity between different maps belonging to different audio excerpts. As a result, it is obtained a value that represents the resemblance between two musical products. This procedure could be used as a non-subjective tool in automatic plagiarism detection. To illustrate this method, three experiments have been carried out comparing different versions famous pop songs. The results point to the usefulness of the method, although this should be contrasted with an analysis of the human perception of this similarity.

Keywords: Fuzzy clustering · Similarity · Plagiarism

1 Introduction

In past editions of Mathematics and Computation in Music (MCM) we have presented a method to estimate the similarity between different characteristics of symbolic music (melody, rhythm, harmony, tunning) [9,10]. In 2019 we presented *Mercury*, a computer framework in which techniques from fuzzy clustering were implemented to Computer-Assisted Musical Composition. This saved, to a certain extent, the uncertainty/inaccuracy inherent in any kind of music [7]. Despite the use of software, the approach has always been from the point of view

of symbolic music, never from the pure treatment of sound. In this paper we propose to extend the applicability of the techniques showed in [9] to comparison of digital audio, based on some attributes of the spectrogram representation.

To achieve our goal, it is necessary to previously process the audio. For this, we have designed an algorithm, which we have called *SpectroMap*, for filtering local maxima (peaks) of the spectrograms. Once this filtering process has been carried out, we obtain the constellation map or fingerprint [12] of the audio fragment. Constellation map can be easily incorporated into the similarity calculations implemented in Mercury, thus obtaining a non-subjective numerical value of similarity between digital audio fragments.

The assessment of the similarity in the conditions described above can be considered as an important element to take into account for the detection of plagiarism. We do not mean to say that the subjective and perceptual part is not important, but if the calculation of similarity between two musical productions is automated, a high value of similarity between them should alert us. In this case, we could also conduct the traditional and pertinent tests [4] to evaluate the existence or not of plagiarism.

We present some examples of similarity estimation between different versions of the same song, using both the spectrogram filtering methods and the similarity calculation methods implemented in Mercury over three different corpora, one for each reference song. The results obtained, beyond giving a ranking of which version is most similar to the original, also provide information about possible compositional interrelationships between the different versions.

2   Theoretical Background

2.1   Clustering Methods

Clustering methods are aimed to create groups of elements within an initial data set, so that the elements included in each group can be considered similar to each other. Unlike the classification methods, in which the elements are assigned with a pre-existing class, in the clustering methods the different classes or subgroups in which the data set is going to be divided have to be defined beforehand the execution of the analysis phase. The clustering procedure will consist of finding a partition of a data set $\mathbf{X}$ that satisfies certain grouping criteria. Following the criteria exposed in [6], given a data set, we will call *element* to each minimum unit of information belonging to it. Each element will have associated a total of q scalar magnitudes called *characteristics* or *attributes*. The term *cluster* (also *group* or *class*) will be used to designate each of the c groupings made from the data set. In *hard* clustering, it is understood that the elements that belong to a certain cluster share properties or characteristics with each other and are differentiable from the elements belonging to another cluster. In *fuzzy* clustering this distinction is no longer so clear. The term *centroid* denotes the central point of each of the clusters. The set of n data is

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\} \subset \mathbb{R}^q. \tag{1}$$

where each $\mathbf{x}_i \in \mathbb{R}^q$, will be a point of q characteristics belonging to a metric space q-dimensional $\mathbb{R}^q$. The index i will designate the i-th element $\mathbf{x}_i$; the number x_{ik} will designate the value of the k-th characteristics of $\mathbf{x}_i$. The total amount of characteristics q is known as the dimensionality of the data set X, and it will have to be a finite and integer number greater than zero.

2.2   Hard and Soft Partitions

In [1] and [2] we find the necessary theoretical fundaments to define the different types of partitioning of a data set. Suppose that $\mathbf{X}$ is a finite set of n elements such that $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$ and we want to distribute the elements of the set $\mathbf{X}$ in a number c of subsets $\mathbf{C} = \{\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_c\}$ with $2 \leq c \leq n$. This family of subsets $\{\mathbf{C}_j : 1 \leq j \leq c\} \subset \mathbf{X}$ will be a partition of type *hard* if:

$$\bigcup_{j=1}^{c} \mathbf{C}_j = \mathbf{X}, \quad \mathbf{C}_j \cap \mathbf{C}_k = \varnothing, \quad 1 \leq j \neq k \leq c. \tag{2}$$

The matrix $\mathbf{U} = [u_{ij}]$ will represent the membership coefficients of each element $\mathbf{x}_i$ to each subset $\mathbf{C}_j$.

2.3   K-Means Clustering

The *k-means* algorithm, first described by [8], is one of the most widely used clustering methods. It can be classified as a non-hierarchical partitioning method of clustering, in which the data set is divided into a number k of groups, each with a *centroid* called *mean*. This algorithm requires setting the number of clusters k in advance, as well as perform a previous initialization of the groups. The grouping results obtained will depend deterministically both on the number of clusters and on the initialization performed, so to trust the results it will be convenient to repeat the procedure with different initializations.

As we have seen, the operation of the algorithm has two main phases: the initialization phase and the iteration phase. In the first phase, each of the n elements will be randomly assigned to one of the k clusters. Is it possible to formulate the k-means algorithm as an optimization problem of an objective function that will be minimized under given convergence conditions [3].

Definition 1. *Let* $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \subset \mathbb{R}^q$ *be a data set of n elements. The* k-means *objective function* $J_w \colon \boldsymbol{M}_c \times \mathbb{R}^{cq} \to \mathbb{R}^+$ *is defined as*

$$J_W(U, \mathbf{v}) = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}(d_{ij})^2. \tag{3}$$

where $d_{ij} = d(\mathbf{x}_i, \mathbf{v}_j)$ is a distance function calculated between the element i and the centroid j; $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c) \in \mathbb{R}^{cq}, \mathbf{v}_j \in \mathbb{R}^q \forall j$ is the set of centroids from

the clusters; $\mathbf{v}_j$ is the centroid of the cluster $u_j \in U, 1 \leq j \leq c$; and the matrix $\mathbf{U} = [u_{ij}] \in \mathbf{M}_{cp}$ is the belonging matrix to a *hard* partition, accomplishing

$$u_{ij} \in [0,1], \quad \sum_{j=1}^{c} u_{ij} = 1, \quad 1 \leq i \leq n, \quad 1 \leq j \leq c. \tag{4}$$

2.4 Fuzzy C-Means Clustering (FCM)

The *fuzzy c-means* algorithm supposes a generalization of the functions described in 3, transforming them into an infinite family of functions. The first of these generalizations was made in [5], later formulated by [1] as an extension of the well known k-means algorithm.

Definition 2. *Let* $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\} \subset \mathbb{R}^q$ *be a data set of* n *items. The fuzzy objective function* c-means $J_w \colon \mathbf{M}_{fc} \times \mathbb{R}^{cq} \to \mathbb{R}^+$ *is defined as*

$$J_\lambda(U, \mathbf{V}) = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^\lambda (d_{ij})^2. \tag{5}$$

where $U \in \mathbf{M}_{fc}$ is a fuzzy partition of $\mathbf{X}$, and $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c) \in \mathbb{R}^{cq}$, $\mathbf{v}_j \in \mathbb{R}^q$ is the set of centroids associated to the clusters $u_j, 1 \leq j \leq c$; and $d_{ij} = d(\mathbf{x}_i, \mathbf{v}_j)$ is any distance function in $\mathbb{R}^q$; u_{ij} is the membership coefficient of the element $\mathbf{x}_i$ to the cluster j; and finally $\lambda \in [1, \infty)$ is the weight exponent, or *fuzziness degree* of the process.

The function originally proposed by [5] is obtained by setting $\lambda = 2$ and selecting the Euclidean distance $d(ij) = d_{euc}(ij)$. It was later generalized by [1] into the following family of functions $\{J_\lambda | 1 \leq \lambda < \infty\}$. We can now see that the objective functions have the distance weighted by the membership coefficients u_{ij}. Since $\mathbf{M}_{fc}$ is a fuzzy partition, the coefficients $u_{ij} \in [0, 1]$.

The fuzzy clustering process will be achieved through an iterative optimization of the objective function J_λ, updating in each iteration both the membership coefficients u_{ij} and the centroids $\mathbf{v}_j$ by following the expressions (see [1]):

$$u_{ij} = \left(\sum_{k=1}^{c} \left[\frac{d(\mathbf{x}_i, \mathbf{v}_j)}{d(\mathbf{x}_i, \mathbf{v}_k)} \right]^{\frac{2}{\lambda-1}} \right)^{-1}, \qquad \mathbf{v}_j = \left(\sum_{i=1}^{n} u_{ij}^\lambda \mathbf{x}_i \right) / \sum_{i=1}^{n} u_{ij}^\lambda. \tag{6}$$

The matrix $U = (u_{ij}), 1 \leq i \leq n, 1 \leq j \leq c$ is now a fuzzy partition of $\mathbf{X}$, built by the membership coefficients u_{ij}. The fuzzy partition verifies

$$\sum_{j=1}^{c} u_{ij} = 1, \quad 1 \leq i \leq n. \tag{7}$$

In what follows we will show the implementation of the fuzzy c-means clustering algorithm proposed by Bezdek in [1]:

STEP 1. Fix a number of clusters m, $2 \leq m < n$. Choose any inner product norm metric for $\mathbb{R}^q$; fix λ, $1 \leq \lambda < \infty$. Initialize $U^{(0)}$.

STEP 2. Calculate the fuzzy centroids $\{v_j^{(k)}\}$ with $U^{(k)}$ and expression (6).

STEP 3. Update $U^{(k)}$ using expression (6) and $\{v_j^{(k)}\}$.

STEP 4. Compare $U^{(k)}$ to $U^{(k+1)}$ using a convenient matrix norm, being $\epsilon \in (0,1)$ and arbitrary termination criterion. If $\| U^{(k+1)} - U^{(k)} \| \leq \epsilon$ then stop, otherwise set $k = k + 1$ and return to STEP 2.

A Dissimilarity Based on FCM Algorithm

Definition 3. *Let $\mathscr{T}^A = \{\mathbf{x_1}, \ldots, \mathbf{x_n}\} \subset \mathbb{R}^q$ and $\mathscr{T}^B = \{\mathbf{y_1}, \ldots, \mathbf{y_m}\} \subset \mathbb{R}^q$ be two data sets, where $n > m$. Let $d : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}$ be a distance function. Let u_{ij} be the final membership coefficients calculated with FCM algorithm, using data set $\mathscr{T}^A$ as data to be partitioned and $\mathscr{T}^B$ as initial set of centroids. The average dissimilarity $\mathscr{D}$ from the data set $\mathscr{T}^A$ to the data set $\mathscr{T}^B$ is defined by*

$$\mathscr{D}(\mathscr{T}^A, \mathscr{T}^B) = \frac{1}{n \cdot m} \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ij} d(\mathbf{x_i}, \mathbf{y_j}) . \tag{8}$$

It is noteworthy that dissimilarity $\mathscr{D}$ does not consider the possible natural order that could exist in both data sets, achieving a partition of $\mathscr{T}^A$ without any special weight to the elements whose degree of neighbourhood is stronger.

2.5   Fuzzy Ordered C-Means Clustering (FOCM)

In [9] and [11] we presented FOCM, an improvement of the FCM algorithm in which the order of both data set and centroids sequences were taken into account during the partition process. Instead of partitioning a data set $\mathbf{X}$ with a given set of c centroids belonging to C categories, let us consider the possibility to implement the partition process introducing the order of the elements in the fuzzy partition process. For that purpose, let us consider two sequences $\mathscr{S}^A$ and $\mathscr{S}^B$ with different number of elements. Sequence $\mathscr{S}^A$ will be the ordered data set to be partitioned, and sequence $\mathscr{S}^B$ will represent the initial set of centroids.

In FOCM, the Neighbourhood Functions will provide higher weights of comparison to the pair of elements of the sequences that share closer positions in the order of each sequence. At the same time, they will decrease the contribution to the global dissimilarity to those pair of elements that are ordinally distant.

The purpose of FOCM is to modify the algorithm FCM so the natural order of both data set sequence $\mathscr{S}^A = \{\mathbf{x_1}, \ldots, \mathbf{x_n}\}$ and centroids sequence $\mathscr{S}^B = \{\mathbf{y_1}, \ldots, \mathbf{y_m}\}$, with, $n < m$, is considered during the partition process.

FOCM algorithm works as follows: for every step in which the fuzzy partition U has been calculated, the coefficients u_{ij} will be multiplied by a weight by

means of a specific neighbourhood function $f(i, j)$. For accomplishing with de convergence criterion, the matrix should be normalized as follows into $\widetilde{U}$

$$\tilde{u}_{ij} = \frac{u_{ij} f(i, j)}{\sum\limits_{k=1}^{m} u_{ik} f(i, k)}. \tag{9}$$

STEP 1. Set $\{v_j^{(0)}\} = \{y_j\}$. Let m, n be the number of notes of $\mathscr{S}^B$ and $\mathscr{S}^A$, respectively. Choose any convenient neighbourhood function.

STEP 2. Choose any inner product norm metric for $\mathbb{R}^q$, and fix $\lambda \geq 1$. Calculate the initial $\widetilde{U}^{(0)}$ using (6), (9) and $\{v_j^{(0)}\}$.

STEP 3. Calculate the fuzzy cluster centers $\{v_j^{(k)}\}$ with $\widetilde{U}^{(k)}$ and the equation (6).

STEP 4. Update $\widetilde{U}^{(k)}$ using the equations (6), (9) and $\{v_j^{(k)}\}$.

STEP 5. Compare $\widetilde{U}^{(k)}$ to $\widetilde{U}^{(k+1)}$ using a convenient matrix norm; being $\epsilon \in (0, 1)$ and arbitrary termination criterion. If $\| \widetilde{U}^{(k+1)} - \widetilde{U}^{(k)} \| \leq \epsilon$ then stop; otherwise set $k = k + 1$ and return to STEP 3.

There is a big number of possible neighbourhood functions (Gaussian, Triangular, Exponential, Sigmoidal, etc.) [9]. In this paper we have chosen the Gaussian neighbourhood function, i.e.

$$f_G(i, j) = Ae^{-\frac{1}{2\sigma^2}\left[i+1-\frac{(n-1)\cdot(j-1)}{(m-1)}\right]^2}. \tag{10}$$

2.6   Definition of a Dissimilarity Based on FOCM Clustering

Using the FOCM algorithm, in [9] was defined a dissimilarity between any pair of sequences with different number of elements.

Definition 4. *Let $\mathscr{S}^A = \{\mathbf{x_1}, \ldots, \mathbf{x_n}\} \subset \mathbb{R}^q$ and $\mathscr{S}^B = \{\mathbf{y_1}, \ldots, \mathbf{y_m}\} \subset \mathbb{R}^q$ be two sequences, where $n > m$. Let $d : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}$ be a distance function. Let u_{ij} be the final membership coefficients calculated with FOCM algorithm, using sequence $\mathscr{S}^A$ as data to be partitioned and sequence $\mathscr{S}^B$ as initial set of centroids. The average dissimilarity $\widetilde{\mathscr{D}}$ from the sequence $\mathscr{S}^A$ to the sequence $\mathscr{S}^B$ is defined by*

$$\widetilde{\mathscr{D}}(\mathscr{S}^A, \mathscr{S}^B) = \frac{1}{n \cdot m} \sum_{i=1}^{n} \sum_{j=1}^{m} \tilde{u}_{ij} d(\mathbf{x_i}, \mathbf{y_j}). \tag{11}$$

In what follows we show the utility of expression (11) for evaluating de dissimilarity between songs or music compositions.

3   A Comparison of Musical Products Based on FOCM

Establishing an objective measurement for calculating the dissimilarity between musical products like pop, rock songs or classical music compositions, can be very useful as a tool for automatic plagiarism detection. Our approach for comparing two musical products will consist of: selecting the digital audio excerpts to be compared; extracting the constellation map (proposed in [12]) from the spectrogram of each excerpt; calculating the FOCM dissimilarity between constellation maps of both excerpts, with Eq. 11, taking into account that a constellation map is a sequence of points formed by time and frequency.

3.1   FFT Process

With the aim of implementing a fingerprint extraction for a given musical signal X_t, we have designed an algorithm that computes a global peak detection over the spectrogram associated to give us its constellation map. Let N_{FFT} and N_O be the length of the Fast Fourier Transform (FFT) window and the number of elements to overlap between segments respectively, we first compute the spectrogram of the signal (S_{tfa}), by using the Hamming window, in order to get the (time, frequency, amplitude) vectors by considering these two parameters. Such representation contains the amplitude spatial information to analyze. Our engine search determines whether a time-frequency point can be considered locally relevant according to its neighbourhood. Then, the detection is processed regarding a required band. Let $\{T_i\}_{i=1}^n$ and $\{F_j\}_{j=1}^m$ be the time and frequency bands of the spectrogram with the amplitude of the event, we can reformulate the spectrogram $S_{tfa} = (T_i)_{i=1}^n = (F_j)_{j=1}^m$ as its rows and columns representations (Figs. 1 and 2).
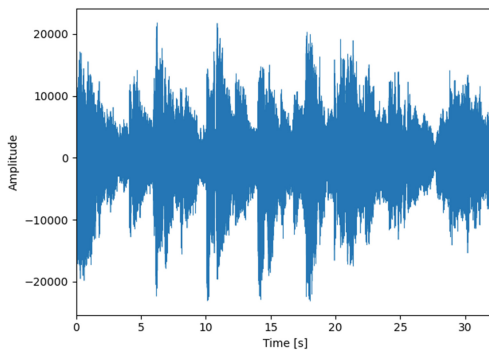
Fig. 1. Example of waveform from an excerpt of a pop song.
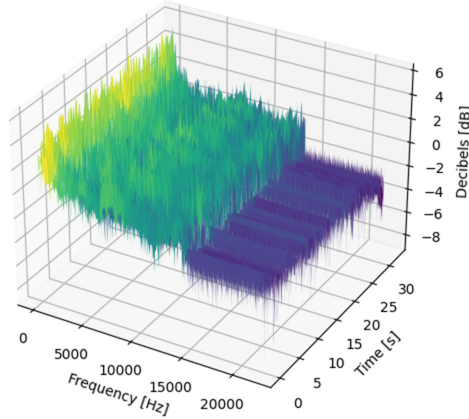
Fig. 2. Spectrogram visualization of the previous excerpt.

3.2   Peak Detection Algorithm

As part of the engine search, we define two windows $\phi_T^{d_T}$ and $\phi_F^{d_F}$ to process the local pairwise comparisons with a respective length of d_T and d_F, whose functionality is to extract a number of elements of the band and return the local maximum. We can describe the time-band window mechanism with length of $0 < d_T \leq n$ and structure $T_i = (T_i^1, ..., T_i^n)$ as

$$\phi_T^{d_T}(T_i) = \left(\max \{T_i^k, ..., T_i^{k+d_T}\} \right)_{1 \leq k \leq n - d_T - 1}, 1 \leq i \leq n. \tag{12}$$

When we group all the values we drop those elements that have equal index to avoid duplicates. We can group the window of each band to create the set:

$$\Phi_T^{d_T} = \{\phi_T^{d_T}(T_i)\}_{i=1}^n. \tag{13}$$

This way, we get the topologically prominent elements per each feature vector. With 12, it is easy to note that even though there are $n - d_T - 1$ matches, the window $\phi_T^{d_T}(T_i)$ may contain a smaller number of elements whenever $d_T > 2$. Depending of how restrictive we need to be, we can proceed with just one of the bands or combine them to create a more stringent search since it is returned only if the peaks that are prominent in both directions. Finally, the algorithm merges all the band-dependent peaks 13, to give us the total number of spatial points that determines the audio fingerprint. Our engine search, *SpectroMap*, processes audio signals in order to return an output file with the (time, frequency, amplitude) peaks detected. The algorithm works in these steps:

STEP 1. Decide the window to use and set the parameters N_{FFT} and N_O.
STEP 2. Read the audio file to get its amplitude vector and its sample rate.
STEP 3. Compute the spectrogram through the associated Fourier transformations.

STEP 4. Set a fixed window length (d_T, d_F or both) for the pairwise comparisons.

STEP 5. Choose the settings to proceed with the peak detection over a selected band or a combination of both.

STEP 6. Create an identification matrix which consists in a binary matrix with the same shape as the spectrogram with the position of the highlighted prominences.

STEP 7. Extract such elements and create a file with the (time, frequency, amplitude) vectors.

Regarding the step 5, authors highly recommend to select both bands to perform the peak detection since the output is more filtered and spatially consistent. For the remainder steps, its choice is a personal decision that depends on the scope of the research (Fig. 3).
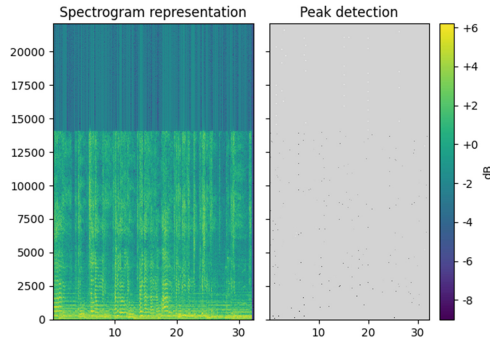
Fig. 3. Spectrogram and result of peak detection algorithm.

3.3 Constellation Map

As it was previously explained, the constellation map is obtained by means of the filtering of local maximum (peak detection) using the algorithm SpectroMap. The sequence of peaks is created by sorting each peak by its appearance time.

Definition 5. *A Constellation Map is the sequence $\mathscr{M}^A = \{\mathbf{x_1}, \ldots, \mathbf{x_n}\} \subset \mathbb{R}^2$ where each $\boldsymbol{x}_i \in \mathbb{R}^2$ is an observable defined by 2 features: time and frequency. Each element has been sorted by its appearance time (Fig. 4).*
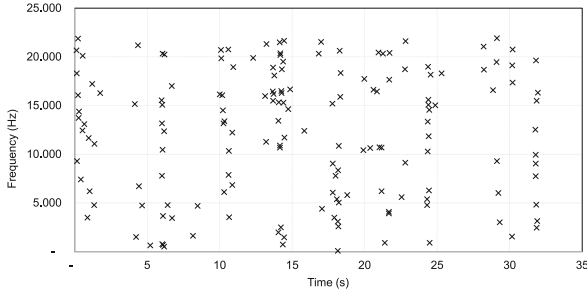
Fig. 4. Example of constellation map generated from an audio excerpt.

3.4   Calculation of the Dissimilarity Based on FOCM

Using the FOCM algorithm, we can define a dissimilarity between any pair of constellation maps.

Definition 6. *Let $\mathcal{M}^A = \{\mathbf{x_1}, \ldots, \mathbf{x_n}\} \subset \mathbb{R}^2$ and $\mathcal{M}^B = \{\mathbf{y_1}, \ldots, \mathbf{y_m}\} \subset \mathbb{R}^2$ be two constellation maps, where $n > m$. Let $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ be a distance function. Let u_{ij} be the final membership coefficients calculated with FOCM algorithm, using constellation map $\mathcal{M}^A$ as data to be partitioned and constellation map $\mathcal{M}^B$ as initial set of centroids. The average dissimilarity $\widetilde{\mathcal{D}}$ between this two constellation is defined by*

$$\widetilde{\mathcal{D}}(\mathcal{M}^A, \mathcal{M}^B) = \frac{1}{n \cdot m} \sum_{i=1}^{n} \sum_{j=1}^{m} \tilde{u}_{ij} \cdot d(\mathbf{x_i}, \mathbf{y_j}). \tag{14}$$

The expression (14) allows us to evaluate the musical plagiarism between any two given excerpts of digital audio.

4   Experiments

To illustrate the applicability of this method, we have designed three experiments to estimate the similarity between different versions of three different pop songs. We have chosen the songs: *Someone Like You*, by Adele; *When I was your man*, by Bruno Mars; *All of me*, by John Legend. This selection is convenient for creating the three different corpora, since there are numerous and different covers available on YouTube. The videos have been downloaded, and the digital audio has been extracted in wav format at 44.100 Hz and 16 bits, selecting the same fragment of the song. With this excerpts we have created three experimental corpora. For each corpora we will calculate the similarities using the method explained in the previous section: applying the SpectroMap algorithm and Eq. 14. In Tables 1, 3 and 5 are shown the audio sources used in each experiment. For each corpora, we will compare the different versions with each other and with

the original, in order to sort them from greater to lesser similarity. The results obtained are shown in Tables 2, 4 and 6.

For experiment No. 1, according to the data shown in the Table 2, the closest resemblance to the official one is the version by the artist jordan (8.21). However, the *leo*, *imy2* and *masha* versions are more similar to each other than to the official version. This fact could indicate a notable influence between these three artists. The version farthest from the official one is that of *leo* (13.62). Once this result is obtained, we listen to the version and verify that the artist has made a version in *rock* style of Adele's theme, effectively far removed in perceptual terms from the original version. The furthest versions are those of *nursera* and *leo* (22.85). Again, if we listen to both versions, the auditory difference is evident, since both versions represent antagonistic musical styles.

The results for experiment No. 2 are shown in the Table 4. The closest resemblance to the official one is the live version by the same artist John Legend (1.311), due to the big similarity of tempi between two versions. The versions of artist *smith* and *imy2* are more similar to the live version. The artist closest to the official version is *scaccia* (1.326). The version farthest from the official one is of *stewart* (13.62).

In experiment No. 3 (Table 6). The closest resemblance to the official song is the cover by artist *scaccia* (2.099), that is also the closest to the live version (1.935). The farthest version from the official one is of *leroy* (2.927).

Table 1. Audio sources used in the first experiment for song *Someone like you*

Artist	YouTube url
Adele Oficial Video (oficial)	https://www.youtube.com/watch?v=hLQl3WQQoQ0
Adele Live Performance (britawards)	https://www.youtube.com/watch?v=qemWRToNYJY
Angelina Jordan (jordan)	https://www.youtube.com/watch?v=nU9TA70fXro
Nursera Yener (nursera)	https://www.youtube.com/watch?v=Z9iylN-IiUA
Masha (masha)	https://www.youtube.com/watch?v=0EwSEsSvxGY
imy2 (imy2)	https://www.youtube.com/watch?v=qIuPgPyTNKE
Leo Moracchioli (leo)	https://www.youtube.com/watch?v=pkbbd3fhcMw

Table 2. Dissimilarities calculated between different covers from experiment No. 1

Cover	Cover	Dissim.	Cover	Cover	Dissim.	Cover	Cover	Dissim.	Cover	Cover	Dissim.
leo	imy2	5.405	jordan	brit	9.693	oficial	brit	10.270	leo	oficial	13.620
leo	masha	5.416	oficial	masha	9.698	imy2	brit	10.455	leo	jordan	14.392
imy2	masha	5.603	brit	jordan	9.778	oficial	imy2	12.759	nursera	brit	14.440
masha	imy2	5.658	brit	masha	9.788	jordan	imy2	13.568	nursera	masha	17.487
oficial	jordan	8.212	jordan	nursera	9.831	oficial	leo	13.586	nursera	imy2	21.603
oficial	nursera	9.680	leo	brit	10.998	imy2	jordan	13.608	nursera	leo	22.847

Table 3. Audio sources used in the second experiment for song *When I was your man*

Artist	YouTube url
Bruno Mars Oficial Version (bmo)	https://www.youtube.com/watch?v=ekzHIouo8Q4
Bruno Mars Live performing (bml)	https://www.youtube.com/watch?v=gY4GZgZK9H0
Alexander Stewart (stewart)	https://www.youtube.com/watch?v=j_d3gq5JCAc
Sam Smith (smith)	https://www.youtube.com/watch?v=_ZaLIiV7c7Y
imy2 (imy2)	https://www.youtube.com/watch?v=uBh_7PBy8cg
Stephen Scaccia (scaccia)	https://www.youtube.com/watch?v=Nhm0MHQKYDY

Table 4. Dissimilarities calculated between different covers from experiment No. 2

Cover	Cover	Dissim.	Cover	Cover	Dissim.	Cover	Cover	Dissim.	Cover	Cover	Dissim.
smith	bml	1.018	bmo	scaccia	1.326	smith	imy2	1.427	smith	scaccia	1.571
stewart	smith	1.262	scaccia	bml	1.335	bml	imy2	1.430	imy2	stewart	1.615
imy2	bml	1.264	imy2	bmo	1.340	bml	bmo	1.466	bmo	smith	1.651
bmo	bml	1.311	imy2	smith	1.398	bml	scaccia	1.495	scaccia	stewart	1.766
imy2	scaccia	1.324	stewart	bml	1.425	scaccia	smith	1.554	stewart	bmo	2.042

Table 5. Audio sources used in the third experiment for song *All of me*

Artist	YouTube url
John Legend Oficial Video (jlo)	https://www.youtube.com/watch?v=450p7goxZqg
John Legend Live Performance (jll)	https://www.youtube.com/watch?v=s18cJqrBIOk
Leroy Sanchez (leroy)	https://www.youtube.com/watch?v=Im6_k-UMJeo
Luciana Zogbi (zogbi)	https://www.youtube.com/watch?v=39_OmBO9jVg
Stephen Scaccia (scaccia)	https://www.youtube.com/watch?v=07McLNDuffo
Scott Hoying (hoying)	https://www.youtube.com/watch?v=d0GR60bul4M

Table 6. Dissimilarities calculated between different covers from experiment No. 3

Cover	Cover	Dissim.	Cover	Cover	Dissim.	Cover	Cover	Dissim.	Cover	Cover	Dissim.
scaccia	jll	1.935	jlo	jll	2.191	jll	hoying	2.309	hoying	jlo	2.832
jll	scaccia	1.989	leroy	hoying	2.210	jll	leroy	2.316	leroy	jlo	2.917
scaccia	jlo	2.099	scaccia	hoying	2.211	scaccia	zogbi	2.662	zogbi	hoying	2.969
jlo	scaccia	2.137	hoying	scaccia	2.270	zogbi	scaccia	2.686	zogbi	leroy	3.196
scaccia	leroy	2.142	leroy	jll	2.273	jlo	zogbi	2.722	zogbi	leroy	3.196

5   Conclusions

The fuzzy logic-based procedures that were implemented for computer-assisted music composition in Mercury software can be used for automatic assessment of music plagiarism from digital audio files.

The Internet and social networks offer an excellent platform for the dissemination of musical content. However, plagiarism detection requires automatic tools that allow quickly and effectively discriminate those versions that may be suspicious in terms of their resemblance to others. Beyond the legal and ethical

aspects, this resemblance can be useful for the performers or authors themselves, who can discover their own and other influences in other artists.

References

1. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algoritms. Plenum Press, New York (1981)
2. Bezdek, J.C., Keller, J., Krisnapuram, R., Pal, N.: Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. Kluwer Academic Publishers, Boston, London, Dordrecht (1999)
3. Gan, G., Ma, C., Wu, J.: Data Clustering: Theory, Algorithms, and Applications. SIAM, Philadelphia (2007)
4. De Prisco, R., et al.: Music plagiarism at a glance: metrics of similarity and visualizations. In: 21st International Conference Information Visualisation (IV), pp. 410–415. IEEE, London (2017). https://doi.org/10.1109/iV.2017.49
5. Dunn, J.C.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J. Cybern. **3**, 32–57 (1973)
6. Jain, A.K., Dubes, R.C.: Algorithms for Clustering Data. Prentice-Hall, Inc., Hoboken (1988)
7. Liern, V.: Fuzzy tuning systems: the mathematics of musicians. Fuzzy Sets Syst. **150**(1), 35–52 (2005)
8. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 3, pp. 281–297. Oakland, USA (1967)
9. Martínez, B., Liern, V.: A fuzzy-clustering based approach for measuring similarity between melodies. In: Agustín-Aquino, O.A., Lluis-Puebla, E., Montiel, M. (eds.) MCM 2017. LNCS (LNAI), vol. 10527, pp. 279–290. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-71827-9_21
10. Martínez–Rodríguez, B., Liern, V.: Mercury: a software based on fuzzy clustering for computer-assisted composition. In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O.A. (eds.) MCM 2019. LNCS (LNAI), vol. 11502, pp. 236–247. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_19
11. Martınez-Rodrıguez B.: El fuzzy clustering y la similitud musical: aplicación a la composición asistida por ordenador. Ph.D. thesis. Universidad Politécnica de Valencia, Valencia, Spain (2019). https://doi.org/10.4995/Thesis/10251/134056
12. Wang, A.: An industrial strength audio search algorithm. In: Ismir, vol. 2003, pp. 7–13 (2003)

A New Fitness Function for Evolutionary Music Composition

Brian Martínez-Rodríguez[✉] [ID]

International University of La Rioja (UNIR), Logroño, La Rioja, Spain
brian.martinez@unir.net
http://www.brianmartinez.es

Abstract. In this paper we propose a new fitness function for Evolutionary Computation purposes, based on a weighted by neighborhood average distance between two sequences of points within any metric space. We will apply this fitness function to the field of Computer-Assisted Composition focusing on the problem of thematic bridging, consisting in the evolutionary creation of a soft set of transitions between two given different melodies, the initial and the final one. Several self-adaptive strategies will be used to perform the search. A symbolic melody will be genotypically mapped into a sequence of genes, each of then containing the information of duration, frequency and time distance to following note. We will test the implementation of the fitness function by means of two experiments, showing some of the intermediate melodies generated in a successful run, and benchmarking every experiment with performance indicators for any of the three distinct evolutionary strategies implemented. The results prove this novel fitness function to be a quick and suitable way for individual evaluation in genetic algorithms.

Keywords: Evolutionary computation · Genetic algorithm · Computer-assisted composition · Fitness · Neighborhood · Thematic bridging

1 Introduction

The use of Evolutionary Computation in the field of computer-assisted composition has been widely addressed through a large variety of evolutionary techniques [12,17]. In this paper we present a novel fitness function that can be employed on evolutionary algorithms that need to evaluate the dissimilarity between the genotype of individuals with different number of genes. We will focus on the specific problem of *Thematic Bridging* [8], conceived as obtaining automatically a set of smooth transitions from any given initial melody to any given goal melody. To achieve this, several evolution strategies [6] will be implemented and afterwards tested in two experiments, measuring the performance indicators for several settings. The evolutionary search, thanks to mutation and crossover operators, will progressively minimize the dissimilarity of every offspring, until an individual

reaches the objective genotype of the goal melody. The best individuals of every offspring will be stored and will create the desired bridging, that could be written into a *musicXML* interchange format file.

2   Related Work

The use of evolutionary algorithms in the field of Computer-Assisted Composition is said to have started at the beginning of the 90's. In the year 1991, Horner and Goldberg [8] implemented one of the first applications of the evolutionary algorithms to computer composition, describing an evolutive technique called *Thematic Bridging* that is able to produce melodic material as a result of iterative transitions between two small melodies. The composition is created by means of human selection and organization of the algorithmically created melodic material, in a form of an imitative five-voices canon.

Three years later, Biles [3] presented *GenJam*, a noteworthy application of genetic computation that generates improvisations in jazz style, keeping the hierarchical relations between different melodic ideas suggested by the harmonic chords progression that is playing. At the same time, the system retrieves feedback information in real-time from the human player. Other interesting evolutive designs were proposed by Hartman [7] and Mcintyre [13].

De la Puente et al. [4] introduced GEMUSIC, a tool that creates algorithmically melodic lines similar to human compositions, thanks to the implementation of Evolutionary Grammars. Weinberg et al. [21] described an interactive evolutive robotic system that collaborates with human players and improvises while playing on a xylophone. The system detects the musical material played by the human and evolves it using several fitness functions. Tzimeas et al. [19] developed the software *Jazz Sebastian Bach*, a system that evolves melodies originally composed by J. S. Bach and turns them into a jazzy style. The authors propose a fitness function called *Critical Damped Oscillator* that overcomes several algorithmic difficulties related to *Automatic Fitness Assessment* (AFA) or *Interactive Genetic Algorithm* (IGA).

De León et al. [16] proposed the characterization of a melody as the result of a set of rules coming from a fuzzy genetic algorithm, aimed to distinguish if a given MIDI file contains a melody or not. The figure of the human-expert knowledge is replaced by a fuzzy genetic system. Sánchez et al. developed MELOMICS project [15], a sophisticated evolution system able to compose and orchestrate whole musical pieces.

Scirea et al. presented in [18] the framework MetaCompose, for music composition that includes a chord sequence and accompaniment generator, and a melody generator that uses a novel evolutionary technique combining FI-2POP and multi-objective optimization. In 2019, Nam YW. and Kim YH. [14] automatized the production of good-quality jazzy melodies by means genetic algorithm, using a variable-length chromosome and geometric crossover.

Trump proposed in [20] a evolutionary framework for improvisation in which the improvisation is created by successive sound cells containing a musical con-

tent transformed by a creative selection. Other interesting approaches were proposed by Almada [1], Arutyunov [2] and Donelly [5].

3 Theoretical Background

3.1 Mathematical Definition of a Melody

As we presented in [10], a melody can be understood as a sequence of points, $\mathcal{M} = \{\mathbf{x}_i\}_{i=1}^{n}$, where each point $\mathbf{x}_i \in \mathbb{R}^q$ is a musical note. The most simple way to represent a note is using three musical characteristics: the duration, the frequency and the time distance that could exist until the next note (representing in this way the possible existence of a silence between this note and the next one). Thus, a musical note will be expressed by a point within a three-dimensional metric space $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$, where the feature x_1 expresses the time duration of the note, x_2 expresses the frequency and x_3 indicates time duration of an optional silence until the next note. In order to calculate the time features x_1 and x_3 of each note, we will use the *relative-duration coefficient* δ proposed in [10]. For the symbolic representation of the frequency in the feature x_2 we will use the *MIDI pitch number* associated to any musical note.

3.2 Neighbourhood Functions

Neighbourhood functions introduced in [10,11] are the key point of the fitness function that will be introduced in the following section. When making a comparison between two sequences, with the first one having a number of n elements and the second one having a number of m elements, the aim of the neighbourhood function will be to calculate the degree of similarity between any element i from the first sequence and any element j of the second one.

   In this way, when comparing two sequences A and B with very different number of elements, if a correct function is defined, the first elements of sequence A will be strongly correlated with the first elements of the sequence B, but very weakly correlated with the final elements of B. In addition, the ending elements of sequence A will be weakly correlated with the first elements of the sequence B, but strongly correlated with the final elements of B. Equation 2 shows the expression of Gaussian Neighbourhood Functions used in this paper.

$$f(i,j) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left[\frac{1}{2\sigma^2}\left(i - \frac{(n-1)}{(m-1)}j\right)^2\right]}. \tag{1}$$

3.3 Fitness Function

We propose a new fitness function for evolutionary music composition based on the definition of *Melodic Dissimilarity* proposed in [10]. Let $\mathcal{M}^A = \{\mathbf{x_1}, \ldots, \mathbf{x_n}\} \in \mathbb{R}^q$ and $\mathcal{M}^B = \{\mathbf{y_1}, \ldots, \mathbf{y_m}\} \in \mathbb{R}^q$ be two different melodies constructed by a different number of notes. Let $d : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}$ be any distance function on the metric space. Let $f(i,j)$ be any neighborhood function.

The Neighborhood Average Dissimilarity $\mathscr{D}$ from melody $\mathscr{M}^A$ to melody $\mathscr{M}^B$ is defined as

$$\mathscr{D}(\mathscr{M}^A, \mathscr{M}^B) = \frac{1}{n \cdot m} \sum_{i=1}^{n} \sum_{j=1}^{m} f(i,j) \cdot d(\mathbf{x_i}, \mathbf{y_j}). \tag{2}$$

The proposed fitness function of each individual will be constructed by de absolute difference between the dissimilarity of each individual A with the goal melody B minus the dissimilarity of the goal melody with itself. Consequently, the expression for the fitness function is

$$F_{fitness}(\mathscr{M}^A) = |\mathscr{D}(\mathscr{M}^A, \mathscr{M}^B) - \mathscr{D}(\mathscr{M}^B, \mathscr{M}^B)| \tag{3}$$

Observe how expression (2) does not accomplish any of the requirements of a distance function. Consequently $\mathscr{D}(\mathscr{M}^B, \mathscr{M}^B) \neq 0$, for the most of the cases.

4 Genetic Algorithm

4.1 Genotype Representation

A melody will be represented into an individual. In the genotype, the sequence of all notes is codified into the sequence of genes. Each gene contains the minimum information of a note. For the following experiments, three different representations of an individual genotypes will be used, each of one containing the required information by any of the evolutionary strategies [3] tested: *simple mutation, uncorrelated mutation with one step size*, and *Uncorrelated mutation with n Step Sizes.*

Simple Mutation. In this representation, the genotype of every individual is a sorted array of float numbers in which the three features x_1, x_2 and x_3 of every note (*gene*) will be stored by order. The length of the array will be $3 \times n$, where n is the number of notes of the melody that is coded for each individual. The representation is as follows

$$(x_1^1, x_2^1, x_3^1, x_1^2, x_2^2, x_3^2, \ldots, x_1^n, x_2^n, x_3^n) \tag{4}$$

Uncorrelated Mutation with One Step Size. In this representation, the genotype of each individual is again a sorted array in which the three features of every note have been stored, besides three values σ belonging to features of time duration, pitch and time distance. The length of the genotype will be $3 \times n + 3$, being n the number of notes of the melody, and its structure will be the following:

$$(x_1^1, x_2^1, x_3^1, x_1^2, x_2^2, x_3^2, \ldots, x_1^n, x_2^n, x_3^n; \sigma_1, \sigma_2, \sigma_3) \tag{5}$$

Uncorrelated Mutation with n Step Sizes. In addition to the previous information of time duration, pitch and time distance belonging to each one of the n notes, this representation will include the sigma value σ corresponding to each one of this features. In this case, the length of the genotype will be $6 \times n$ and its structure:

$$(x_1^1, x_2^1, x_3^1, \ldots, x_1^n, x_2^n, x_3^n; \sigma_1^1, \sigma_2^1, \sigma_3^1, \ldots, \sigma_1^n, \sigma_2^n, \sigma_3^n) \tag{6}$$

4.2   Restrictions on the Evolution Strategy

Some constrains will be implemented into the evolutionary strategy aimed to reduce the space of research. The first restriction is introduced in the mutation of the MIDI pitch value. The mutated pitch value will be forced to be a integer number, due to the MIDI mapping of the musical notes is enclosed from 0 to 127, so the algorithm does not consider the possible existence of intervals smaller than a semitone.

The second constrain is implemented into the possible variation of the relative-duration Coefficient δ. This feature will be mutated by means of adding or subtracting a multiple of a minimum-duration figure. The arbitrarily chosen minimum-duration is a demisemiquaver (thirty-second note), with coefficient $\delta_{min} = 0.03125$.

4.3   Mutation

The mutation in a specific gene will be done using *random resetting*, establishing an uniform mutation probability for every genes. When a gene is randomly chosen for being mutated, the feature represented by this gene into a float number will be modified adding or subtracting a certain amount, calculated according to the case that we consider.

Simple Mutation. In the case of simple mutation, the features of duration x_1, pitch x_2 and time distance x_3 of a selected note i will be modified using these expressions based on the equations exposed in [6]:

$$x'^i_1 = x^i_1 + \delta_{min} \cdot \lceil \sigma_1 \cdot N(0,1) \rceil$$
$$x'^i_2 = x^i_2 + \lceil \sigma_2 \cdot N(0,1) \rceil \tag{7}$$
$$x'^i_3 = x^i_3 + \delta_{min} \cdot \lceil \sigma_3 \cdot N(0,1) \rceil$$

where $N(0,1)$ is a generator of Gaussian distributed random numbers, centered in the zero value (mean equal to zero), and with standard deviation equal to 1. In this case, the values of σ stay constant.

Uncorrelated Mutation with One Step Size. In this kind of mutation, the values σ_1, σ_2 and σ_3 used to calculate the changes in the features of duration, pitch, and time distance, are assumed to change randomly for each individual. Therefore, the mutations on the standard deviations and features x_i will be done by means of the following expressions:

$$
\begin{aligned}
\sigma'_1 &= \sigma_1 \cdot e^{\tau \cdot N(0,1)} & x'^i_1 &= x^i_1 + \delta_{min} \cdot \lceil \sigma_1 \cdot N(0,1) \rceil \\
\sigma'_2 &= \sigma_2 \cdot e^{\tau \cdot N(0,1)} & x'^i_2 &= x^i_2 + \lceil \sigma_2 \cdot N(0,1) \rceil \\
\sigma'_3 &= \sigma_3 \cdot e^{\tau \cdot N(0,1)} & x'^i_3 &= x^i_3 + \delta_{min} \cdot \lceil \sigma_3 \cdot N(0,1) \rceil
\end{aligned}
\tag{8}
$$

For every mutation of the deviation σ_j, we should check if the new value is not too small. To achieve this, we establish a threshold value ϵ_j below which σ can not still decrease, so:

$$
\begin{aligned}
\sigma'_1 &< \epsilon_1 \Rightarrow \sigma'_1 = \epsilon_1 \\
\sigma'_2 &< \epsilon_2 \Rightarrow \sigma'_2 = \epsilon_2 \\
\sigma'_3 &< \epsilon_3 \Rightarrow \sigma'_3 = \epsilon_3
\end{aligned}
\tag{9}
$$

Uncorrelated Mutation with n Step Sizes. In this case, each one of the x^i_j features of an individual will mutate with a specific deviation σ^i_j. The mutations of the deviations and features of a chosen note i will be carried out with these expressions:

$$
\begin{aligned}
\sigma'^i_1 &= \sigma^i_1 \cdot e^{\tau \cdot N(0,1)} & x'^i_1 &= x^i_1 + \delta_{min} \cdot \lceil \sigma^i_1 \cdot N(0,1) \rceil \\
\sigma'^i_2 &= \sigma^i_2 \cdot e^{\tau \cdot N(0,1)} & x'^i_2 &= x^i_2 + \lceil \sigma^i_2 \cdot N(0,1) \rceil \\
\sigma'^i_3 &= \sigma^i_3 \cdot e^{\tau \cdot N(0,1)} & x'^i_3 &= x^i_3 + \delta_{min} \cdot \lceil \sigma^i_3 \cdot N(0,1) \rceil
\end{aligned}
\tag{10}
$$

Once again, it is necessary to check if the mutated value of every deviation σ^i_j is not smaller than a given threshold ϵ_0.

Mutation Operation Concerning the Number of Notes of a Melody. Besides mutating the features of duration, frequency and time distance of a random note from the melody coded on the genotype of each individual, it is needed to establish a mutation operation to change the number of notes of the melody, since we want to achieve an evolutionary transition from one initial melody to a second one, both having conceivably a different number of notes.

Two arbitrary probabilities for insertion and suppression of a note will be implemented in order to insert a new note in a random position p inside the genotype, or remove the note located on the position p, respectively.

When inserting a brand new note in a random position of the genotype, there exist three different possibilities:

– *Inserting the note at the beginning of the melody ($p = 0$)*: In this case, three new positions will be inserted at the very beginning of the genotypical array.

The values of these three new positions will duplicate exactly the three features of the prior first note, so the new inserted note will duplicate exactly the ancient first one.

– *Inserting the note at the end of the melody* ($p = 3 \cdot n$): In this case, three new positions will be added at the end of the genotypical array. The values of this positions will duplicate the previous last note of the melody.

– *Inserting the note in an intermediate position within the melody* (p = k, $0 < k < 3 \cdot n$): In this case, a new note will be inserted between the notes placed in the positions $k - 1$ and k. Each one of the three features of the new note will be calculated as an average value of the corresponding feature from the two adjacent notes, taking into account the previously specified constraints of mutation changes in duration and pitch.

In the cases of the genotype related to the representation of Uncorrelated mutation with one Step Size and Uncorrelated mutation with n Step Sizes, it will also be necessary to include in the genotypical array one extra position in case of one Step size, or three extra positions in case os n Step Sizes, in order to include the new σ values corresponding to the new note.

4.4 Initialization of the Population, Parents Selection and Crossover

The population will be initialized creating a number μ of different individuals, whose genotypes have been initially cloned from the starting melody one, and afterwards subjected to a random mutation process.

A number of λ couples of parents will be randomly chosen to generate a new child from every couple. The recombination operation for the new genotype will be the uniform crossover, so each gene will be randomly inherited from any of the parents.

The selection process of survivors for next generation will be guided by method $\mu + \lambda$, which involves mixing together the population of parents and offspring [3], sorting by each individual's fitness and choosing the best μ individuals.

4.5 Performance Indicators

For each execution there will be a maximum of 200 generations. Each experiment will be executed 1.000 times for any one of the pre-established setups. The algorithm will store the following performance indicators [3]:

– **SR** (Success Rate): Percentage of executions that finish successfully over the total number of executions.
– **MB** (Mean Best Fitness): Average of the best fitness value of the population when execution finishes, successfully or not.
– **MBFS** (Mean Best Fitness Success): Average of the better fitness value of the population taking into account only the successful executions.
– **AES** (Average number of Evaluations to a Solution): Average number of generations needed to reach a successful execution.
– **MST** (Mean Success Time): Mean time needed to find a successful execution.

5   Experiments

5.1   First Experiment

Given initial melody A and goal melody B, use the evolutionary strategies with fitness function (3) to generate a melodic transition from A to B (Fig. 1).

Fig. 1. Initial and final melodies of the first experiment.

Settings of the algorithm: $\mu = 20$, $\lambda = 200$, mutation prob. $= 0.15$ and note insertion prob. $= 0$. Run 1.000 times. Results shown in Table 1 and Fig. 2:

Fig. 2. Some intermediate melodies generated at one successful execution.

Table 1. Benchmarking indicators for experiment one.

	Simple	One step	n steps
SR	100.00%	100.00%	100.00%
MBFS	0.01904	0.01904	0.01903
MB	0.01904	0.01904	0.01903
AES	43.80	99.60	97.30
MST (ms)	167.86	373.03	378.77

5.2 Second Experiment

Given initial melody A and goal melody B, use the evolutionary strategies with fitness function (3) to generate a melodic transition from A to B (Fig. 3).

Fig. 3. Initial and final melodies of the second experiment.

Settings of the experiment: $\mu = 20$, $\lambda = 500$, mutation prob. $= 0.15$ and note insertion prob. $= 0.05$. Run 1.000 times. Results shown in Table 2 and Fig. 4.

Fig. 4. Some intermediate melodies generated at one successful execution.

Table 2. Benchmarking indicators for experiment two.

	Simple	One step	n steps
SR	21.30%	34.60%	40.70%
MBFS	0.38764	0.38764	0.38764
MB	0.41251	0.40294	0.40050
AES	55.60	56.50	49.90
MST (ms)	1490.69	1490.53	1331.01

5.3   Results

In Fig. 5 it is possible to compare the performance curves for experiment one and two, for simple mutation, Uncorrelated mutation with one Step Size and Uncorrelated mutation with n Steps Size. The final benchmarks for the performance indicators of SR, AES and MST are summarized in Table 3.
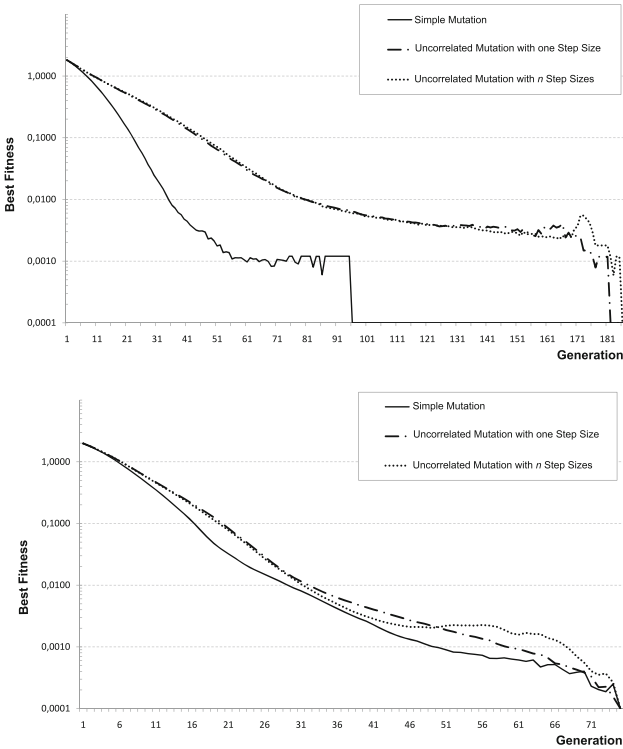
Fig. 5. Performance charts for the first and second experiment.

Table 3. Final performance indicators for the experiments.

	SR			AES			MST		
	Simp.	1 step	n steps	Simp.	1 step	n steps	Simp.	1 step	n steps
Exp. #1	100,00%	100,00%	100,00%	43.80	99.60	97.30	167.86	373.03	378.77
Exp. #2	21.30%	34.60%	40,70%	55.60	56.50	49.90	1490.69	1490.53	1331.01

6   Discussion

We have run 1000 executions for each set of experiments one and two, with a maximum number of 200 offsprings for each execution. The success rate of experiment one was 100%, due to the melodies not being very distant in terms of evolutionary search. The success rate of experiment two was 40,7% as the initial and final melodies where very distant. For the experiment one, the most efficient mutation was the simple mutation. Nevertheless, Uncorrelated mutation with n Steps Sizes has been the most efficient in experiment two. All the representations achieved low values of mean success time (MST) and low average number of evaluations to a solution (AES).

7   Conclusions

The evolutionary algorithm implemented and tested in the two experiments has proved to find solutions to the problem of thematic bridging between two melodies, thanks to the minimization of the novel fitness function proposed in this paper (3) and based on the Neighborhood Average Dissimilarity (2).

The evolutionary algorithm implemented in the experiments exercise has been shown to be capable of making transitions between two melodies by minimizing the fitness function proposed in (3), in a quick and useful way for simple evolutionary composition poruses. Future work will involve the implementation of more sophisticated evolutionary techniques, widening the rhythmical restrictions of the evolutionary algorithm and incorporating the generation of harmonic chord sequences.

References

1. Lemos Almada, C.: Genetic algorithms based on the principles of *Grundgestalt* and developing variation. In: Collins, T., Meredith, D., Volk, A. (eds.) Mathematics and Computation in Music, MCM 2015. LNCS (LNAI), vol. 9110, pp. 42–51. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20603-5_4
2. Arutyunov, V., Averkinb, A.: Genetic algorithms for music variation on genom platform. In: 9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW 2017, pp. 317–324 (2017)
3. Biles, J.A.: GenJam: a genetic algorithm for generating jazz solos. In: Proceedings of the 1994 International Computer Music Conference, pp. 131–137. International Computer Music Association, Copenhagen (1994)

4. de la Puente, A., Alfonso, R., Moreno, M.: Automatic composition of music by means of grammatical evolution. In: Proceedings of the 2002 Conference on APL: Array Processing Languages, pp. 148–155. Association for Computer Machinery, Madrid (2002)

5. Donnelly, P., Sheppard, J.: Evolving four-part harmony using genetic algorithms. In: Di Chio, C., et al. (eds.) Applications of Evolutionary Computation, EvoApplications 2011. LNCS, vol. 6625, pp. 273–282. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-20520-0_28

6. Eiben, A., Smith, J.: Introduction to Evolutionary Computing. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-44874-8

7. Hartmann, P.: Selection of musical identities. In: Proceedings of the 1990 International Computer Music Conference, pp. 234–236. International Computer Music Association, Glasgow (1990)

8. Horner, A., Goldberg, D.: Genetic algorithms and computer-assisted music composition. In: Proceedings of the 1991 International Computer Music Conference, Montreal, Canada, vol. 51, pp. 479–482 (1991)

9. Horner, A., Ayers, L.: Harmonization of musical progressions with genetic algorithms. In: Proceedings of the 1995 International Computer Music Conference, pp. 483–484. International Computer Music Association, Banff (1995)

10. Martínez, B., Liern, V.: A fuzzy-clustering based approach for measuring similarity between melodies. In: Agustín-Aquino, O.A., Lluis-Puebla, E., Montiel, M. (eds.) Mathematics and Computation in Music, MCM 2017. LNCS (LNAI), vol. 10527, pp. 279–290. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-71827-9_21

11. Martínez-Rodríguez, B., Liern, V.: Mercury: a software based on fuzzy clustering for computer-assisted composition. In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O. (eds.) Mathematics and Computation in Music, MCM 2019. Lecture Notes in Computer Science, vol. 11502, pp. 236–247. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_19

12. Miranda, E., Al Biles, J.: Evolutionary Computer Music. Springer, London (2007). https://doi.org/10.1007/978-1-84628-600-1

13. McIntyre, R.: Bach in a box: the evolution of four part baroque harmony using the genetic algorithm. In: Proceedings of the First IEEE Conference on Evolutionary Computation, pp. 852–857. IEEE, Orlando (1994)

14. Nam, Y.-W., Kim, Y.-H.: Automatic jazz melody composition through a learning-based genetic algorithm. In: Ekárt, A., Liapis, A., Castro Pena, M.L. (eds.) Computational Intelligence in Music, Sound, Art and Design, EvoMUSART 2019. LNCS, vol. 11453, pp. 217–233. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-16667-0_15

15. Sánchez, C., Moreno, F., Albarracín, D., Fernández, J., Vico, F.: Melomics: a case-study of AI in Spain. AI Mag. **34**(3), 99–103 (2013)

16. de León, P., Rizo, D., Ramirez, R and Iñesta, J.: Melody characterization by a genetic fuzzy system. In: Proceedings of the 5th Sound and Music Computing Conference. Sound and Music Computing Association, Berlin (2008)

17. Romero, J.J.: A Handbook on Evolutionary Art and Music. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-72877-1

18. Scirea, M., Togelius, J., Eklund, P., Risi, S.: MetaCompose: a compositional evolutionary music composer. In: Johnson, C., Ciesielski, V., Correia, J., Machado, P. (eds.) Evolutionary and Biologically Inspired Music, Sound, Art and Design, EvoMUSART 2016. Lecture Notes in Computer Science, vol. 9596, pp. 202–217. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-31008-4_14

19. Tzimeas, D., Mangina, E.: A GA tool for computer assisted music composition. In: Proceedings of 2007 International Computer Music Conference. International Computer Music Association, Copenhagen (2007)

20. Trump, S.: Sound cells in genetic improvisation: an evolutionary model for improvised music. In: Romero, J., Ekárt, A., Martins, T., Correia, J. (eds.) Artificial Intelligence in Music, Sound, Art and Design, EvoMUSART 2020. Lecture Notes in Computer Science, vol. 12103, pp. 179–193. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-43859-3_13

21. Weinberg, G., Godfrey, M., Rae, A., Rhoads, J.: A real-time genetic algorithm in human-robot musical improvisation. In: Kronland-Martinet, R., Ystad, S., Jensen, K. (eds.) Computer Music Modeling and Retrieval. Sense of Sounds, CMMR 2007. Lecture Notes in Computer Science, vol. 4969, pp. 351–359 . Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-85035-9_24

A Mathematical Model of Tonal Function (I): Voice Leadings

Isaac del Pozo and Francisco Gómez-Martín[(✉)]

Universidad Politécnica de Madrid, Madrid, Spain
`francisco.gomez@upm.es`

Abstract. This paper is the first in a series of two papers on mathematical models of tonal function. In this first paper, we present a mathematical model of tonal function whose scope is limited to classic music from the common practice period. After a formalization of some harmonic elements (pitch classes, chords, arrangements, voicings, voice leadings), the model of tonal function is described. Our model is based on voice leadings and the tonal function is defined in terms of them. A combinatorial optimization algorithm is used to determine the tonal function. In this work, only chords with the same number of voices are considered. The general case was left for the second paper of the series; in the second paper the model of tonal function is generalized.

Keywords: Tonal function · Voice-leadings · Chord progressions · Matrix algebra · Distance functions · Nabla distance · Optimal voice-leadings · Chord classification · Characteristic polynomials

1 What Is Tonal Function?

This paper is the first in a series of two papers about mathematical models of tonal function presented at the *Mathematics and Computation in Music 2022* conference [23]. In this paper we present a mathematical model of tonal function (also called harmonic function) in the context of classical and jazz music. In this paper basic concepts and formalization are introduced and a mathematical model of tonal function for chords with the same number of voices is presented. In the second paper [20], the case where chord do not have the same number of voices is examined. Furthermore, our model of tonal function is applied to the study of modulation.

Several models—both mathematical and musical—have been presented by a number of authors in the past few years. The concept of tonal function was first introduced by Hugo Riemannn in his 1893 work *Vereinfachte Harmonielehre* [22] (*Harmony simplified* in its English version). According to Hyer [14], Riemannn borrowed the term from mathematics. Riemannn coined the names for the three main tonal functions, tonic, dominant, and subdominant functions. His theory

soon gained popularity and was much used in composition, pedagogy of harmony, and music analysis. About a half century later, the Viennese school of harmony, led by figures such as Simon Sechter and Heinrich Schenker and later Arnold Schoenberg, created the scalar degree theory, which put more emphasis on the use of Roman numeral, the figure bass, and the relation of chords to the tonal center within a harmonic progression; see Watson [24].

Riemann's theory of harmony has undergone many extensions and revisions since its initial formulation, which have guaranteed its use in today's musical practice. For example, authors such as Lewin [17], Hyer [13], and Cohn [6,7] have formulated a neo-Riemannian theory where the operations between chords originally defined by Riemann are extended and refined to a larger set of operations resulting in a definition of tonal function capable of dealing with music from the extended common practice. See [11] for a survey on those extensions of Riemann's theory.

Strange as it may seem, there are not many formal and precise definitions of tonal function in the literature and yet some of them seem not to fully capture the essence of this musical phenomenon; in particular, there are very few mathematical definitions of tonal function. Riemann's definition is laid out in terms of a theory of consonance and dissonance and also in terms of operations on chords [22]. The resulting theory of harmony as we know it today is a mixture of Riemann's theory and the theories by the Viennese school. However, the definition of tonal function seems to be vague in many cases. For instance, in the classical textbook *Harmony* by Walter Piston [18]—a textbook used by many generations of harmony students —, we find the following definition of tonal function.

> Tonality, then, is not merely a matter of using just the tones of a particular scale. It is more a process of setting forth the organized relationship of these tones to one among them which is to be the tonal center. Each scale degree has its part in the scheme of tonality, its tonal function.

This definition is somewhat vague and similar ones can be found in other textbooks on harmony. About the adequacy of the very term "function", Kopp [15] for example contends that, "the meaning of the word has proved adaptable to support a wide variety of statements concerning harmony." and discuss the semantics of the word "function." In this paper, we will present a definition of harmonic function based on voice-leadings.

Table 1. Scalar degree characteristics of tonal function [21].

Function	Triggers	Associates	Dissonances
T	1 and 3	5 and 6	5 (if 6 present) and 7
SD	4 and 6	1 and 2	1 (if 2 present) and 3
T	5 and 7	2	4 and 6

A more complete definition is given in the project *Open Music Theory* [3], which in turns follows Quinn's definition of tonal function [21]. The definition given therein classifies chords on scalar degrees I, III, and VI as tonic chords; II and IV as subdominant chords; and V and VII as dominant chords. The

characteristics of the scale degrees are then added to his definition. Tonic's characteristic scale degrees are 1, 3, 5, 6, and 7, (which result from take the scale degrees of the tonic's chords), subdominant's characteristic scale degrees are 1, 2, 3, 4, and 6, and, finally, dominant ones are 2, 4, 5, 6, and 7. Some authors do not agree with this chord classification and expand it by assigning VI to subdominant and III to dominant.

Quinn further refines this chord classification and introduces the categories of functional triggers, functional associates, and functional dissonances; see Table 1. These categories were created based on the degree of stability of the chord notes with respect to the tonal center; other authors such as Harrison [12] or Cohn [8] have proposed similar classifications. Cohn [8] can be considered as a precedent of the work presented here (his sum-class transformations, for example).

Tonal function has been studied from many standpoints. Carpenter [5] examined how a musical idea can be shaped through tonal function; Caplin [4] in turn looked into the historical perspectives on the relation between tonal function and metrical accent; Dudeque [9] investigated Schoenberg's concept of tonal function; Agustín-Aquino and Mazzola [1] presented a theory of modulation that includes a definition of harmonic function.

The paper is organized as follows. In Sect. 2, a mathematical formalization of the basic concepts related to chords and voice leadings through matrix algebra is presented. This section is a review of the formalization presented by the authors in [19]. For the present work such review will be needed to establish terminology and results. In Sect. 3 an algorithm to find optimal voice leadings (the Hungarian algorithm) is described and a procedure to classify chords according to its tonal function is presented. This procedure is based on optimal voice leadings. The paper comes to an end with a conclusion section.

2   Basic Concepts

2.1   Pitch Classes and Chords

In this section we present a formalization of the most fundamental musical concepts. We start off by considering the range of audible frequencies, the interval $(20, 2 \cdot 10^4)$ (measured in Hz). For the sake of completeness, this interval will be extended to the real line. This has the advantage of being closed under the sum and product of frequencies. The space of frequencies will be denoted by Φ. The principle of octave equivalence will be applied here and therefore an equivalence relation $\sim$ between frequencies is defined. We write $x \sim y$ if and only if $x = 2^k \cdot y$, for some integer k. This relation identifies all the pitches that are apart any number of octaves as just one pitch. We further assume that the octave is evenly divided into 12 parts, that is, we assume equal temperament.

Given a fixed pitch class $[k]$, we define the circle of fifths $LC_k/\sim$ as the set of classes

$$LC_k/\sim = \left\{ [k], \left[k \cdot 2^{\frac{7}{12}} \right], \left[k \cdot 2^{\frac{14}{12}} \right], \left[k \cdot 2^{\frac{21}{12}} \right], \ldots, \left[k \cdot 2^{\frac{77}{12}} \right] \right\}$$

This definition is illustrated in Fig. 1 where the pitch k was set to A. Then, the circle of fifths was built up from it by multiplying the previous pitch by $2^{\frac{7}{12}}$, the distance of a fifth in terms of frequency.

A **chord** X is an unordered collection of classes of pitches taken from $LC_k/\sim$. In Western tonal music, in addition to the pitches, chords are described by specifying two more features, namely, the **root** and the **quality**. The root is the lowest pitch in any voicing of the chord (see the definition of arrangement below); the quality refers to labels given to chords to describe different variants of a the same chord. For example, a dominant seventh chord on F is the chord composed by F-A-C-E♭, in that order. The root of this chord is F and the quality dominant seventh. This label tells us that the first three notes form a major triad and that E must flat so that there is minor seventh between F and E. Usually, quality of chords is shown by using several symbols such as m or lowercase for minor chords, $+$ for augmented chords, etc. Notice that when the root and the quality are introduced, the pitches are then ordered.

A **chord progression** is a sequence of chords. As such, chords in a progression are presented in the order in which they appear in the piece of music. A natural way to deal with chord progression is to represent them as matrices whose entries are classes of pitches. If $P \in \mathcal{M}_{m \times n}(LC_k/\sim)$ is a chord progression of length n, then each chord is a vector of m notes and there are n chords in the progression. We can arrange the notes of the chord progression in a matrix as follows.
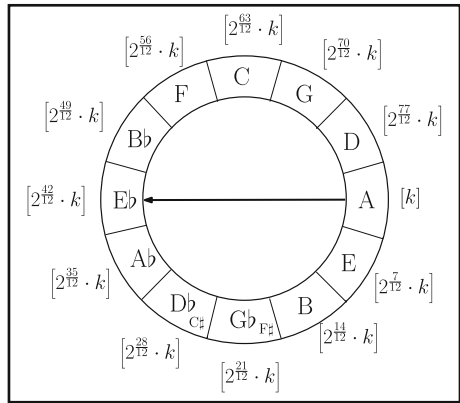
Fig. 1. The circle of fifths

$$P = \begin{pmatrix} [\theta_{11}] & \cdots & [\theta_{1n}] \\ \vdots & \ddots & \vdots \\ [\theta_{m1}] & \cdots & [\theta_{mn}] \end{pmatrix} \tag{1}$$

To fix ideas, let consider the 2-note chord progression {F, A} to {E, G}, which from now on will be notated as {F, A}$\Longrightarrow${E, G}. Its matrix representation is $P = \begin{pmatrix} [F] & [E] \\ [A] & [G] \end{pmatrix}$.

Let Φ^+ be the set of positive frequencies. A **voicing** of a chord is a mapping V_X from $\mathcal{M}_{m \times 1}(LC_k/\sim)$ to $\mathcal{M}_{m \times 1}(\Phi^+)$. This mapping takes a given pitch class to a note. Indeed,

$$V_X \left(\begin{pmatrix} [\theta_{1j}] \\ \vdots \\ [\theta_{mj}] \end{pmatrix} \right) = \begin{pmatrix} \phi_{1j} \\ \vdots \\ \phi_{mj} \end{pmatrix} \tag{2}$$

where $\phi_{ij} \in [\theta_{ij}]$, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. Following with the previous example, a voice leading for the chord progression could be (among other possibilities)

$$V_X \left(\begin{pmatrix} [F] \\ [A] \end{pmatrix} \right) = \begin{pmatrix} F4 \\ A4 \end{pmatrix}$$

For ease of reading, we will notate the frequencies by their standard names instead of their numerical values. Therefore, we will write A4 instead 440 Hz.

An **arrangement** of a chord progression is a mapping defining which notes of the chords are chosen for each voicing. Formally, it is a mapping $\psi : \mathcal{M}_{m \times n}(LC_k/ \sim) \to \mathcal{M}_{m \times n}(\Phi^+)$ written as

$$\psi \left(\begin{pmatrix} [\theta_{11}] & \cdots & [\theta_{1n}] \\ \vdots & \ddots & \vdots \\ [\theta_{m1}] & \cdots & [\theta_{mn}] \end{pmatrix} \right) = \begin{pmatrix} \phi_{11} & \cdots & \phi_{1n} \\ \vdots & \ddots & \vdots \\ \phi_{m1} & \cdots & \phi_{mn} \end{pmatrix} \tag{3}$$

where $\phi_{ij} \in [\theta_{ij}]$, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. From now on, arrangements will be notated as $(\phi_1, \ldots, \phi_n) \longrightarrow (\phi'_1, \ldots, \phi'_n)$, that is, as bijections between sequences of notes; compare this notation to that of chord progressions above.

For the chord progression {F, A}$\Longrightarrow${E, G}, A_C could take on, among others, the form of

$$A_C \left(\begin{pmatrix} [F] & [E] \\ [A] & [G] \end{pmatrix} \right) = \begin{pmatrix} F4 & E4 \\ A4 & G4 \end{pmatrix}$$

2.2   The Nabla Distance

The space of frequencies can be enriched by defining a metric on it. A metric space requires the concept of distance, in our case the distance between two notes. It is a well-known fact that our ears measure the distance in terms of ratios of frequencies. In the following, we will formalize the concept of distance between notes. Given two frequencies $\alpha, \beta \in \Phi^+$, a metric $\Delta : (\Phi^+)^2 \to \mathbb{R}$ is defined as the following application:

$$\Delta(\alpha, \beta) = \left| \int_\alpha^\beta \frac{\Omega}{\phi} d\phi \right| \tag{4}$$

where Ω is a constant such that $\left| \int_1^2 \frac{\Omega}{\phi} d\phi \right| = 12$; see [2] for a relationship between this constant and the definition of cents.

By working out the integral above, this distance can be expressed as $\Delta(\alpha, \beta) = \left| \Omega \ln \left(\frac{\alpha}{\beta} \right) \right|$. The actual value of the constant is $\Omega = 12 \cdot |\log_2(e)|$. This value was found under the assumption of equal temperament (equal division of the octave into 12 equal parts). The Δ function does hold the three properties of a metric, namely: positivity, $\Delta(\alpha, \beta) \geq 0$; symmetry, $\Delta(\alpha, \beta) = \Delta(\beta, \alpha)$; and the triangle inequality $\Delta(\alpha, \beta) \leq \Delta(\alpha, \gamma) + \Delta(\gamma, \beta)$.

The pair (Φ^+, Δ) is called the **musical metric space**. This metric can be extended to the space of pitch classes by just taking the minimum of the elements

in each pitch class. For two classes $[\theta], [\tau]$ in $LC_k/\sim$, the distance between them, notated by $\tilde{\Delta}$, is defined as

$$\tilde{\Delta}([\theta], [\tau]) = \min\left\{\Delta(\alpha, \beta) \mid \alpha \in [\theta], \beta \in [\tau]\right\} \tag{5}$$

See the work [10] of Forte for more information on distance functions. For example, $\Delta(C5, E4) = 8$ and $\Delta(C4, E5) = 16$, but

$$\tilde{\Delta}([C], [E]) = \min\left\{\Delta(\alpha, \beta) \mid \alpha \in [C], \beta \in [E]\right\} = 4$$

Notice that the maximum value the distance $\tilde{\Delta}$ can take is 6.

Let $P \in \mathcal{M}_{m \times n}(LC_k/\sim)$ be a chord progression such that $P = ([p_{ij}])$, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. Consider σ, an element in the symmetric group $\mathcal{S}_m$ defined over the set of indices $\{1, 2, \ldots, m\}$. Then, we define $E(P)$, the **extension** of P, as those matrices $B = (b_{ij})$ in $\mathcal{M}_{m \times n}(LC_k/\sim)$ such the following two conditions hold:

1. For some values of j, $[p_{ij}] = [b_{ij}]$, for all $i = 1, \ldots, m$;
2. For the rest of values of j, $[p_{ij}] = [b_{\sigma_k(i)j}]$, for all $i = 1, \ldots, m$, where σ_k is a permutation in $\mathcal{S}_m$.

These conditions state that a column in B is either the same column in P or a permutation of some column of P. $E(P)$ is the set of such matrices. Consider again the matrix associated to the chord progression {F, A}$\Longrightarrow${E, G}. Then, the extension of P is

$$E(P) = \left\{\begin{pmatrix} [A] & [G] \\ [F] & [E] \end{pmatrix}, \begin{pmatrix} [A] & [E] \\ [F] & [G] \end{pmatrix}, \begin{pmatrix} [F] & [G] \\ [A] & [E] \end{pmatrix}, \begin{pmatrix} [F] & [E] \\ [A] & [G] \end{pmatrix}\right\}$$

$E(P)$ has cardinal $(m!)^n$.

Next, we need to define the distance that a voice travels through a given chord progression. We will use the symbol $\tilde{\nabla}$ to define the **distance** of a **chord progression** P. Then, $\tilde{\nabla}(P)$ is defined as follows:

$$\tilde{\nabla}(P) = \sum_{i=1}^{m} \sum_{j=1}^{n-1} \tilde{\Delta}([\theta_{ij}], [\theta_{i(j+1)}]) \tag{6}$$

The value of $\tilde{\nabla}(P)$ is the sum of all the distances between consecutive notes of a voice over all voices in the chord progression.

The operator nabla can also be defined for the set $E(P)$ as follows: $\tilde{\nabla}(E(P)) = \left\{\tilde{\nabla}(B) \mid B \in E(P)\right\}$. Notice that $\tilde{\nabla}(P)$ is a real value and $\tilde{\nabla}(E(P))$ a set of values. Let us compute $\tilde{\nabla}(P)$ for the chord progression {E, C}$\Longrightarrow${F, A}. Indeed, $\tilde{\nabla}(P) = \sum_{i=1}^{m} \sum_{j=1}^{n-1} \tilde{\Delta}([\theta_{ij}], [\theta_{i(j+1)}]) = \tilde{\Delta}([E], [F]) + \tilde{\Delta}([C], [A]) = 1 + 3 = 4$ Actually, we do not need to consider all the matrices in $E(P)$ to compute $\tilde{\nabla}(E(P))$. It is enough to choose those where the first column is not rearranged. The nabla distances of the matrices in $E(P)$ are

$$\tilde{\nabla}\left(\left(\begin{bmatrix}[C]&[A]\\[E]&[F]\end{bmatrix}\right)\right) = 1 + 3 = 4, \quad \tilde{\nabla} = \left(\left(\begin{bmatrix}[C]&[F]\\[E]&[A]\end{bmatrix}\right)\right) = 5 + 5 = 10,$$

The nabla of the extension of P is $\tilde{\nabla}(E(P)) = \left\{\tilde{\nabla}(B) \mid B \in E(P)\right\} = \{4, 10\}$.

A **chord progression** is said to be **optimal** if $\tilde{\nabla}(P) = \min\left\{\tilde{\nabla}(E(P))\right\}$. In our example, the chord progression {E, C}$\Longrightarrow${F, A} was optimal as the nabla distance attained the minimum at that progression.

Analogously, the nabla distance can be defined for arrangements; it will be notated by ∇ (without tilde). If $A = (\phi_{ij}) \in \mathcal{M}_{m \times n}(\Phi^+)$ is an arrangement, then the formal definition of ∇ is $\nabla(A) = \sum_{i=1}^{m} \sum_{j=1}^{n-1} \Delta(\phi_{ij}, \phi_{i(j+1)})$.

An arrangement A is said to be **optimal** if $\nabla(A) = \tilde{\nabla}(E(P_A))$, where P_A is the chord progression associated to A. Let us consider two arrangements associated to the chord progression {E, C}$\Longrightarrow${F, E}, say, A_1:(E4, C4)$\longrightarrow$(F4, E4) and A_2:(E4, C4)$\longrightarrow$(F5, E5). Let us find which one is optimal by computing their nabla distances. We have $\nabla(A_1) = \Delta(E4, F4) + \Delta(C4, E4) = 1 + 4 = 5$ and $\nabla(A_2) = \Delta(E4, F5) + \Delta(C4, E5) = 13 + 16 = 29$. Therefore, the first arrangement is the optimal one.

From now on, we will simplify and omit the square brackets for the pitch classes. The context will clearly disambiguate pitch classes from particular notes.

3   A Mathematical Model of Tonal Function

In this section, we introduced the **polynomial criterion** to classify a chord according to its tonal function given a tonal center. The chord will be classified into three categories, namely, tonic T, subdominant SD, and dominant D; the tonal center will be denoted by TC. For now, we will assume that the number of voices in all chords is the same and constant. Let us fix a tonal center TC for the rest of this discussion.

3.1   Voice Optimization

Given a chord X, there always exists an optimal voice leading linking X to TC. This optimal voice leading yields a **matrix link** $L = (\tilde{\Delta}_{ij})$, for $i, j = 1, \ldots, n$, where n is the number of voices (which is constant) and $\tilde{\Delta}_{ij}$ are the nabla distances between all the classes of X and TC. To fix ideas, let the number of voices n be equal to 4 and $TC = (C, E, G, B) \in \mathcal{P}(LC_k/\sim)$ be C major seventh. Suppose we wish to determine the tonal function of $V7$; in order to do so, we have to examine the chord progression $P = V7 \longrightarrow I\Delta$[1]. The corresponding matrix along with the matrix link are

---

[1] Δ here denotes the major seventh chord, not the nabla distance. The symbol is the same, but the context determines its meaning without confusion.

$$P = \begin{pmatrix} F & B \\ D & G \\ B & E \\ G & C \end{pmatrix}; \quad L = \begin{pmatrix} 6 & 2 & 1 & 5 \\ 3 & 5 & 2 & 2 \\ 0 & 4 & 5 & 1 \\ 4 & 0 & 3 & 5 \end{pmatrix}$$

We now apply a combinatorial optimization algorithm to matrix L. The algorithm is called the **Hungarian method** [16], which was developed by American mathematician Harold Kuhn in 1955. This algorithm solves the assignment problem in polynomial time. The reason to apply it here is because the optimization of a voice leading between two chords can be interpreted as an assignment problems of agents and tasks; it can furthermore be interpreted as a matching problem in weighted bipartite graphs in which the sum of the edges is minimum. Here follows a description of the Hungarian algorithm and an example of it; the input in our case is the link matrix $L = (\tilde{\Delta}_{ij})$.

Hungarian Algorithm

Step 1. For each row i in L, $i = 1, \ldots, n$, select the minimum value $\min_{j=1,\ldots,n}\{\tilde{\Delta}_{ij}\}$ and subtract it from each element in said row. This will make the minimum entry in each row equal to 0.

Step 2. For each column j in L, $j = 1, \ldots, n$, select the minimum value $\min_{i=1,\ldots,n}\{\tilde{\Delta}_{ij}\}$ and subtract it from each element in said column. At this point, the minimum value of each row and column is equal to 0.

Step 3. Draw lines through the rows and columns containing a 0 entry so the the minimum number of lines is drawn. The maximum amount of lines that can be drawn is n, the number of chord's voices.

Step 4. If in Step 3 n lines were drawn, then an optimal assignment is possible. Select the 0 entry for each zero and replace it with the original values. This assignment yields the optimum assignment of voice leadings between the two chords. Exit the algorithm. Otherwise, if fewer than n lines were drawn, then go to Step 5.

Step 5. This step is divided into three cases: Find the entries in the matrix not covered by any line and compute minimum entry. Then, (1) Subtract the minimum entry from those entries that are not crossed out; (2) Find the entries in the matrix covered by one line. Leave them unchanged; (3) Find the entries in the matrix covered by two lines. Sum the minimum entry to each entry. Then, go back to Step 3.

The output of the Hungarian algorithm on matrix L is a new optimized link matrix L^*. Applying this algorithm to the previous chord progression, we arrive at the matrix shown on the next page

$$\begin{pmatrix} 6\ 2\ 1\ 5 \\ 3\ 5\ 2\ 2 \\ 0\ 4\ 5\ 1 \\ 4\ 0\ 3\ 5 \end{pmatrix} \xrightarrow{\text{Step 1}} \begin{pmatrix} 5 & 1 & \boxed{0} & 4 \\ 1 & 3 & \boxed{0} & \boxed{0} \\ \boxed{0} & 4 & 5 & 1 \\ 4 & \boxed{0} & 3 & 5 \end{pmatrix} \xrightarrow{\text{Step 2}} \begin{pmatrix} 5 & 1 & \boxed{0} & 4 \\ 1 & 3 & \boxed{0} & \boxed{0} \\ \boxed{0} & 4 & 5 & 1 \\ 4 & \boxed{0} & 3 & 5 \end{pmatrix}$$

$$\xrightarrow{\text{Steps 3}} \begin{pmatrix} 5 & 1 & \boxed{0} & 4 \\ \cancel{1} & \cancel{3} & \cancel{0} & \cancel{0} \\ \boxed{0} & 4 & 5 & 1 \\ 4 & \boxed{0} & 3 & 5 \end{pmatrix} \xrightarrow{\text{Step 4}} \begin{pmatrix} - & - & \boxed{0} & -- \\ - & - & - & \boxed{0} \\ \boxed{1} & - & - & - \\ - & \boxed{2} & - & - \end{pmatrix}$$

In Step 3, we obtained 4 zeros, which is the number of voices; of the two possible zeros in column 3, the horizontal line correspond to the first one . After Step 4 is executed, the optimal voice leading corresponds to those zeros. In this case, that voice leading is $B \longrightarrow B, G \longrightarrow G, F \longrightarrow E$ and $D \longrightarrow C$.

In general, the solution to this optimization problem need not be unique and the Hungarian algorithm returns all of the solutions. The correctness of the algorithm ensures that the order and the particular zeros chosen in Step 3 always lead to an optimal solution.

Out of the solutions output by the Hungarian algorithm, an optimal voice leading P^* between the chord X and the tonal center TC is constructed. The notes of the voice leading can be rearranged so that the optimal values for the voice leading appear on the main diagonal of the associated link matrix. Such link matrix is denoted by L^*. Carrying on with our example, the rearrangement of the voice leading and the link matrix are

$$P^* = \begin{pmatrix} F & E \\ D & C \\ B & B \\ G & G \end{pmatrix}; \quad L^* = \begin{pmatrix} \boxed{1} & 5 & 6 & 2 \\ 2 & \boxed{2} & 3 & 5 \\ 5 & 1 & \boxed{0} & 4 \\ 3 & 5 & 4 & \boxed{0} \end{pmatrix}$$

Notice that the rearrangement is not unique.

It is straightforward to conclude that the trace of L^* is precisely $\nabla(\psi(P^*))$, that is, the distance of the optimal voice leading.

3.2   Cadence Endomorphisms and Chord Classification

For a chord X, let us call $v(X)$ a voicing. The mapping that takes any voicing of X to the optimal voice leading to the tonal center is a rearrangement of the notes in X. Let us consider $L^* = (a_{ij}), i, j = 1, \ldots, n$, where n is the number of voices in chord X. We can define an endomorphism—to be called **cadence endomorphism**—, $Cad : \Phi^n \longrightarrow \Phi^n$ that transform the original voicing into the optimal one, that is, $Cad(v(X)) = TC$, where TC is the given tonal center. In order to define Cad, for each note i in X, let b_{ij} be defined as follows: (1)

$b_{ij} = a_{ij}$ if note i is lower or equal than note j; (2) $b_{ij} = -a_{ij}$ if note i is higher than note j. In case (1) the voice leading goes down or does not change whereas in case (2) the voice leading goes up. Denote by s the quantity $2^{\frac{1}{12}}$, a semitone or half-tone in an equal division of the octave. Then, the endomorphism matrix for Cad, M_{Cad}, is given by

$$M_{Cad} = \begin{pmatrix} s^{b_{11}} & 0 & \cdots & 0 \\ 0 & s^{b_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s^{b_{nn}} \end{pmatrix} = \begin{cases} s^{b_{ij}}, & \text{if } i = j \\ \\ 0, & \text{if } i \neq j \end{cases}$$

Let $P_{Cad}(\lambda)$ be the characteristic polynomial associated to this endomorphism. The roots of $P_{Cad}(\lambda)$ are precisely $s^{b_{11}}, \ldots, s^{b_{nn}}$ and such polynomial can be written as $P_{Cad}(\lambda) = \prod_{i=1}^{n}(s^{b_{ii}} - \lambda)$. The exponents of those roots can be negative, positive, or zero. Below there is a classification of cadence endomorphisms according to the sign of the exponents. This classification serves as a means to define the tonal function in very general settings such as musical spaces with chords with a high number of voices; more will be discussed in the second paper.

Cadence endomorphism classification criteria:

- **Positive-definite endomorphisms:** when all the exponents are positive;
- **Negative-definite endomorphisms:** when all the exponents are negative;
- **Mixed-definite endomorphisms:** when some exponents are positive and others are negative, and there is no zero exponents;
- **Positive-semidefinite endomorphisms:** when all the exponents are either positive or zero;
- **Negative-semidefinite endomorphisms:** when all the exponents are either negative or zero;
- **Mixed-semidefinite endomorphisms:** when some exponents are either positive or negative or zero, and the three cases occur.

When the exponents are either all positive or all negative, we say that the exponents are **polarized**.

Examining classical music in the common practice period and assuming that chords have a constant number of voices equal to three or four, we are ready to put forward our definition. Our definition captures the voice movement to move from a given chord to a fixed tonal center. By the model we have built, such movement conforms with the signs of the exponents of the characteristic polynomial of the cadence endomorphism. According to the sign of the exponents, the tonal function is determined as follows.

Chord classification criteria

- **Dominant chords:** A chord X is a dominant chord if all the exponents are negative or zero having at least two negative exponents;
- **Weak dominant chords:** A chord X is a weak dominant chord if there is just one positive exponent, two or more negative exponents, and the rest of the exponents are zero;
- **Subdominant chords:** A chord X is a subdominant chord if all the exponents are positive or zero having at least two positive exponents;
- **Weak subdominant chords:** A chord X is a weak subdominant chord if there is just one negative exponent, two or more positive exponents, and the rest of the exponents are zero;
- **Tonic chords:** A chord X is a tonic chord if all the exponents are zero except one, which can be either negative or positive.

In the example above, the main diagonal of the cadence endomorphism matrix is s^{-1}, s^0, s^0, s^{-2} since F goes to E and D to C in the optimal voice leading. Therefore, the chord (G, B, D, F) is a dominant chord with respect to the tonal center (C, E, G, B).

If the tonal center is set as the major seventh chord $C\Delta$, then Table 2 shows the tonal function for the main 4-note chords in the scale of C major computed by our method.

Table 2. Four-note chords and their tonal function

Function	Chords
Tonic	$C\Delta$, Am7 Em7, $C\Delta$sus2, $C\Delta$sus4
Dominant	G7, B°7, G7sus4, Dm7sus2, Em7sus4
Subdominant	Dm7, $F\Delta$, Dm7sus4, Am7sus2

4   Conclusions

A mathematical model of tonal function has been presented in this paper. It is based on voice leadings and captures the way a chord moves towards the tonal center. The movement has to be optimal (hence, the Hungarian algorithm) and the voices have to move in the same direction (either up or down). We contend this definition explains more precisely and deeply the tonal function, especially as found in classical music during the common practice period. Our model allows for generalization and can also be used for other kinds of music traditions such as jazz music or music from the extended common practice (further development is given in the second paper of this series).

Some of the definitions of tonal function reviewed in the introduction are not actually such insofar as they describe the tonal function and classify chords, but

they do not explain or justified satisfactorily the classification. Some authors resort to the "ear" (this chord as a dominant chord because the ear feels it that way) while others—like Schenberg himself—rely on a theory of consonance and dissonance, which it is known it cannot explain tonal function in all music genres, especially jazz music and extended common practice.

In the second part of this series we address the problem of having chords with different number of voices and also how modulation works within this model.

References

1. Agustín-Aquino, O.A., Mazzola, G.: Modulation in tetradic harmony and its role in jazz. J. Math. Music **14**(1), 58–65 (2020). https://doi.org/10.1080/17459737.2019.1655673
2. Benson, D.: Music: A Mathematical Offering. Cambridge University Press, Cambridge (2006)
3. Bry, H., et al.: Harmonic function. http://openmusictheory.com/harmonicFunctions.html, Access Dec 2021
4. Caplin, W.: Tonal function and metrical accent: a historical perspective. Music Theory Spect. **5**, 1–14 (1983)
5. Carpenter, P.: Grundgestalt as tonal function. Music Theory Spect. **5**, 15–38 (1983)
6. Cohn, R.: Neo-riemannian operations, parsimonious trichords, and their tonnetz representations. J. Music Theory **41**(1), 1–66 (1997)
7. Cohn, R.: An introduction to neo-riemannian theory: a survey and historical perspective context and harmonic function in tonal music. J. Music Theory **42**(2), 167–180 (1998)
8. Cohn, R.: Square dances with cubes. J. Music Theory **42**(2), 283–296 (1998). http://www.jstor.org/stable/843879
9. Dudeque, N.: Schoenberg on tonal function (1997). http://www.rem.ufpr.br/_REM/REMv2.1/vol2.1/Schoenberg/Schoenberg_on_Tonal.html
10. Forte, A.: The Structure of Atonal Music. Yale University Press, London (1973)
11. Gollin, E., Rehding, A.: Oxford Handbook of Neo-Riemannian Music Theories. Oxford University Press, New York (2011)
12. Harrison, D.G.: Harmonic Function in Chromatic Music: A Renewed Dualist Theory and an Account of Its Precedents. University of Chicago Press, Chicago (1994)
13. Hyer, B.: Reimag(in)ing riemann. J. Music Theory **39**(1), 101–138 (1995)
14. Hyer, B.: Tonality. Grove Music (2001). https://doi.org/10.1093/gmo/9781561592630.article.28102
15. Kopp, D.: On the function of function. Music Theory Online **1**(3), 1–8 (1955)
16. Kuhn, H.W.: The Hungarian method for the assignment problem. Naval Res. Logist. Q. **2**, 83–97 (1955)
17. Lewin, D.: Generalized Musical Intervals and Transformations. Yale University Press, New Haven (1987)
18. Piston, W.: Harmony. Gollancz, London (1950)
19. Pozo, I., Gómez, F.: Formalization of voice-leadings and the nabla algorithm. In: Proceedings of Mathematics and Computation in Music 2019, Madrid, Spain (2019)

20. Pozo, I., Gómez, F.: A mathematical model of tonal function (I): voice leadings. In: Proceedings of Mathematics and Computation in Music 2019, Atlanta, USA (2022)
21. Quinn, I.: Harmonic function without primary triads. In: Paper Delivered at the Annual Meeting of the Society for Music Theory in Boston (2005)
22. Riemann, H.: Harmony simplified: or, The theory of the tonal functions of chords. Augener Ltd., London (1896). Translation of Vereinfachte Harmonielehre (1893). Republished by Alpha Editions, September 2019, London
23. SMCM: Proceedings of mathematics and computation in music (2022). http://mcm2022.org/
24. Wason, R.E.: Viennese Harmonic Theory from Albrecthsberger to Schenker and Schoenberg, Ann Arbor, London (1985)

A Mathematical Model of Tonal Function (II): Modulation

Isaac del Pozo[ORCID] and Francisco Gómez-Martín[(✉)][ORCID]

Universidad Politécnica de Madrid, Madrid, Spain
`francisco.gomez@upm.es`

Abstract. This work is the second part of a research into mathematical models of tonal function. In the first paper, we presented a mathematical model based on optimal voice leadings for chords with the same number of notes. We gave several examples for classical music in the common practice period. In this second paper, we generalize the model to jazz music and music from the extended common practice and we remove the constraint on the number of voices, allowing the classification of chords with a different number of notes. Furthermore, modulation within this mathematical model is also explored and a classification of modulation between major and minor scales is constructed.

Keywords: Tonal function · Voice-leadings · Nabla distance · Optimal voice-leadings · Cadence endomorphisms · Characteristic polynomials · Modal interchange · Modulation

1 Introduction

This is the second paper of a series of two where mathematical models of tonal function are examined. In the first paper [5], a review of the tonal function concept was carried out, followed by a mathematical formalization of musical elements, which was needed for the designing of a mathematical model of the tonal function. Contrary to many definitions of tonal function based on theories of consonance and dissonance, ours is grounded on optimal voice leadings between chords and the tonal center. Given a tonal center, the function of a chord is determined by the length of the optimal voice leading between the given chord and the given tonal center and how the voices in that voice leading move (the polarization of the exponentes in the characteristic polynomial). In the previous paper, we applied our model to music having chords with the same number of notes, which mainly corresponds to diatonic music such as classical music from the common practice, folk music, among others. In this work, we will generalize our mathematical model to chords with a variable number of voices and will apply to music traditions such as jazz music and music from the extended common practice.

In the first paper, we reviewed the definition of tonal function in classical music mainly and found that in general the definitions available were vague. How is the conceptualization of tonal function in other music traditions such as jazz music, for example? In jazz music—one of the musical styles where harmony is central—, it seems that a conceptual, comprehensive definition of tonal function also lacks. Certainly, there are many definitions in the literature and textbooks, but most of them establish the tonal function without elucidating its roots. In his 2004 paper [8] *Jazz harmony: a progress report*, Stover [8] reviews three important jazz textbooks, namely, Mulholland and Hojnacki's *The Berklee Book of Jazz Harmony* [4], David Berkman *The Jazz Harmony Book* [1], and Dariusz Terefenko's *Jazz Theory: From Basic to Advanced Study* [9]. In the first book, tonal function is described as "the relationship of a chord to its tonal center" (page 1). In Berkman's book, the three classical tonal functions are first described and then a series of concentric circles are built out of this core, from which all the rest of chords are classified. Moving through these circles is done through chord transformations (this patently resonates with neo-Riemannian theory). In the third book, Terefenko provides a definition of tonal function in terms of hierarchy of chords (page 54), but again that hierarchy is described in detailed, but not its derivation. Worth mentioning is the book by Russell [7] *The Lydian Chromatic Concept of Tonal Organization* where the author describes the tonal function as an interaction of harmonic forces and tonal gravity (page 3 and following). Russell's theory implies is a deeper and more comprehensive of tonal function.

The paper is organized as follows. In Sect. 2, the theory about how to deal with chords with different number of voices is developed. In Sect. 3, we examined an interesting phenomenon that arises in our model: chords may have two functions depending on their voice movement; this chords are said to have a dual tonal function. In Sect. 4, chord classification by our model is discussed; in particular, it is examined how the direction of the voice may influence the tonal function. In Sect. 5, the idea of optimal voice leading applied to determining the tonal function is now extended to modulation. Two tonalities are related through the optimal voice leading linking them. A classification of modulation is thus proposed. The paper ends with a conclusion section where we discuss the implications and reach of our model.

2   Tonal Function in Chords with Different Number of Voices

In this section, we will examine the case in which the number of voices is not the same in both chords. To that end, we will have to extend the Hungarian algorithm to deal with infinity arithmetic. To fix ideas, assume that X is a chord with fewer voices than the tonal center TC. The point where the previous ideas fail is when computing the distance between both chords. Let us now define

the **empty class** in the space of pitches as the class without pitches; it will be denoted by $[\emptyset]$. The distance from the empty class to any other class $[\alpha]$ is obtained by taking the limit

$$\lim_{\beta \to 0} \left| \int_{\alpha}^{\beta} \frac{\Omega}{\phi} d\phi \right| = +\infty$$

where $\Omega = 12 \cdot |\log_2(e)|$. Consequently, $\tilde{\Delta}([\alpha], [\emptyset]) = \infty$. For our purposes, we assume the following **convention** regarding operations with infinity

Convention: Given any real number k, the result of the operation $\infty - (\infty - k)$ is k.

Assuming this convention, we have the following expression for the difference of distances $\tilde{\Delta}([\alpha], [\emptyset]) - \tilde{\Delta}([\beta], [\emptyset])$ is:

$$\tilde{\Delta}([\alpha], [\emptyset]) - \tilde{\Delta}([\beta], [\emptyset]) = \infty - \infty = 0$$

The Hungarian algorithm with this infinity arithmetic will be called the **extended Hungarian algorithm**.

To illustrate how the tonal function in this new context, we will next calculate the tonal function of the chord $X = (G, B, D)$ being $TC = (C, E, G, B)$ the tonal center. Note that the first chord have fewer classes than the second chord.

$$P = \begin{pmatrix} [\emptyset] & B \\ D & G \\ B & E \\ G & C \end{pmatrix} \longrightarrow L = \begin{pmatrix} \infty & \infty & \infty & \infty \\ 3 & 5 & 2 & 2 \\ 0 & 4 & 5 & 1 \\ 4 & 0 & 3 & 5 \end{pmatrix} \xrightarrow{\text{Step 1}} \begin{pmatrix} \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} \\ 1 & 3 & \boxed{0} & \boxed{0} \\ \boxed{0} & 4 & 5 & 1 \\ 4 & \boxed{0} & 3 & 5 \end{pmatrix}$$

$$\xrightarrow{\text{Step 2}} \begin{pmatrix} \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} \\ 1 & 3 & \boxed{0} & \boxed{0} \\ \boxed{0} & 4 & 5 & 1 \\ 4 & \boxed{0} & 3 & 5 \end{pmatrix} \xrightarrow{\text{Step 3}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 4 & 5 & 1 \\ 4 & 0 & 3 & 5 \end{pmatrix} \xrightarrow{\text{Step 4}} \begin{pmatrix} - & - & - & \boxed{\infty} \\ - & - & \boxed{2} & - \\ \boxed{0} & - & - & - \\ - & \boxed{0} & - & - \end{pmatrix}$$

The voice leading between both chords turns out to be $D \longrightarrow E, B \longrightarrow B, G \longrightarrow G$ and $[\emptyset] \longrightarrow C$. Notice that in this optimal voice leading the class $[\emptyset]$ lands on C. According to the classification given earlier, the chord $X = (G, B, D)$ is a subdominant chord with respect to the tonal center $TC = (C, E, G, B)$ since two voices are kept constant and two voices move up. Notice that the appearance of a voice as in $[\emptyset] \longrightarrow C$ is considered as a voice moving up.

Having taken the presence of infinite into account, some formal issues arise. In the previous example, one of the exponents is $+\infty$. The associated cadence endomorphism Cad will have an entry $s^{+\infty}$, which makes no sense. It would seem that there is no cadence endomorphism in this case, but that is not so. It is possible to construct a cadence endomorphism M_{Cad} that maps X onto TC. In our example such endomorphism is

$$M_{Cad} : \Phi^4 \longrightarrow \Phi^4, \quad M_{Cad} \cdot \psi(X) = \psi(TC)$$

$$M_{Cad} = \begin{pmatrix} w_1 & \phi_4/\phi_1 & 0 & 0 \\ w_2 & s^0 & 0 & 0 \\ w_3 & 0 & s^0 & 0 \\ w_4 & 0 & 0 & s^2 \end{pmatrix}$$

where $\psi(X)$ and $\psi(TC)$ are voicings of X and TC, respectively, and $w_1 \ldots, w_4$ are arbitrary parameters to be determined. Let ϕ_1, ϕ_2, ϕ_3 be the notes in the arrangement corresponding to the chords in the voice leading; in our example, those notes correspond to the classes G, B, D, respectively. Furthermore, let ϕ_4, ϕ_5 be the notes corresponding to C and E, respectively. If M_{Cad} is the endomorphism taking $(0, \phi_3, \phi_2, \phi_1)$ to $(\phi_4, \phi_5, \phi_2, \phi_1) = (\phi_4, s^2\phi_3, s^0\phi_2, s^0\phi_1)$, then the matrix expression of the cadence endomorphism is:

$$\begin{pmatrix} \phi_4 \\ \phi_1 \\ \phi_2 \\ \phi_5 \end{pmatrix} = \begin{pmatrix} w_1 & \phi_4/\phi_1 & 0 & 0 \\ w_2 & s^0 & 0 & 0 \\ w_3 & 0 & s^0 & 0 \\ w_4 & 0 & 0 & s^2 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix} = \begin{pmatrix} \phi_4 \\ s^0\phi_1 \\ s^0\phi_2 \\ s^2\phi_3 \end{pmatrix}$$

Therefore, there are infinite endomorphisms that output the same optimal voice leading.

Let us examine what happens when the number of voices decreases. Take, for example, the chord $X = (C, E, G, B\flat)$ and set $TC = (F, A, C)$ as tonal center. Below we have the execution of the Hungarian algorithm on this chord and tonal center.

$$P = \begin{pmatrix} B\flat & [\o] \\ G & C \\ E & A \\ C & F \end{pmatrix} \longrightarrow L = \begin{pmatrix} \infty & 2 & 1 & 5 \\ \infty & 5 & 2 & 2 \\ \infty & 4 & 5 & 1 \\ \infty & 0 & 3 & 5 \end{pmatrix} \xrightarrow{\text{Step 1}} \begin{pmatrix} \infty & -1 & 1 & 0 & 4 \\ \infty & -2 & 3 & 0 & 0 \\ \infty & -1 & 3 & 4 & 0 \\ \infty & & 0 & 3 & 5 \end{pmatrix}$$

$$\xrightarrow{\text{Step 2}} \begin{pmatrix} 1 & 1 & 0 & 4 \\ 0 & 3 & 0 & 0 \\ 1 & 3 & 4 & 0 \\ 2 & 0 & 3 & 5 \end{pmatrix} \xrightarrow{\text{Step 3}} \begin{pmatrix} 1 & 1 & \boxed{0} & 4 \\ \boxed{0} & 3 & 0 & 0 \\ 1 & 3 & 4 & \boxed{0} \\ 2 & 0 & 3 & 5 \end{pmatrix} \xrightarrow{\text{Step 5}} \begin{pmatrix} 1 & 1 & \boxed{0} & 4 \\ \boxed{0} & 3 & 0 & 0 \\ 1 & 3 & 4 & \boxed{0} \\ 2 & \boxed{0} & 3 & 5 \end{pmatrix}$$

$$\xrightarrow{\text{Step 4}} \begin{pmatrix} - & - & \boxed{0} & - \\ \boxed{\infty} & - & - & - \\ - & - & - & \boxed{0} \\ - & \boxed{0} & - & - \end{pmatrix}$$

Reading off the voice leading from the last matrix above, it yields the voice leading $B\flat \longrightarrow A, E \longrightarrow F, C \longrightarrow C$ and $G \longrightarrow [\o]$. This last voice leading makes the note G in the chord X disappear. Similarly to the previous case, the disappearance of a voice is considered as a voice moving down. As for the tonal function, two voices move down, two voices stay the same, and just one voice

move up. Therefore, $(C, E, G, B\flat)$ is a weak dominant chord with respect to the tonal center (F, A, C).

In the case of fewer voices in the tonal center, the issue of infinity does not occur. The infinity in the last matrix in the example actually corresponds to $-\infty$ since the C disappears in the voice leading. Therefore, the entry $s^{-\infty}$, which is interpreted as $\lim_{K \to -\infty} s^K$, is zero. The characteristic polynomial of the above endomorphism is $P_{Cad}(\lambda) = -\lambda(s^{-1} - \lambda)(s^1 - \lambda)(s^0 - \lambda)$.

3   Dual Tonal Function

The mathematical model proposed here is based on optimal voice leadings. In all models of tonal function—mathematical or otherwise—we are aware of, the tonal function is always unique. A chord cannot take on two different functions. However, in our model that is not always the case. Certain chords may have a dual tonal function that will depend on whether the voices move up or down in the optimal voice leading. Some of the definitions found in the literature do not take into account how the voices move and therefore miss this subtle point. The lack of uniqueness in our model is due to the fact that the optimal voice leading need not be unique. Let us illustrate this point with an example. Consider the scale of C major and within it the 4-note chord D minor seventh (D, F, A, C); the tonal center is $C\Delta$, the chord (C, E, G, B). Let us apply the Hungarian algorithm to find the optimal voice leading.

$$P = \begin{pmatrix} C & B \\ A & G \\ F & E \\ D & C \end{pmatrix} \longrightarrow L = \begin{pmatrix} 1 & 5 & 4 & 0 \\ 2 & 2 & 5 & 3 \\ 6 & 2 & 1 & 5 \\ 3 & 5 & 2 & 2 \end{pmatrix} \xrightarrow{\text{Step 1}} \begin{pmatrix} 1 & 5 & 4 & 0 \\ 0 & 0 & 3 & 1 \\ 5 & 1 & 0 & 4 \\ 1 & 3 & 0 & 0 \end{pmatrix} \xrightarrow{\text{Step 2}} \begin{pmatrix} 1 & 5 & 4 & 0 \\ 0 & 0 & 3 & 1 \\ 5 & 1 & 0 & 4 \\ 1 & 3 & 0 & 0 \end{pmatrix}$$

$$\xrightarrow{\text{Steps 3 and 4}} \begin{pmatrix} 1 & 5 & 4 & 0 \\ 0 & 0 & 3 & 1 \\ 5 & 1 & 0 & 4 \\ 1 & 3 & 0 & 0 \end{pmatrix} \xrightarrow{\text{Step 5}} \begin{pmatrix} 0 & 4 & 4 & 0 \\ 0 & 0 & 4 & 2 \\ 4 & 0 & 0 & 4 \\ 0 & 2 & 0 & 0 \end{pmatrix} \xrightarrow{\text{Step 3 and 4}} \begin{cases} S_1 = \begin{pmatrix} 0 & 4 & 4 & \boxed{0} \\ \boxed{0} & 0 & 4 & 2 \\ 4 & \boxed{0} & 0 & 4 \\ 0 & 2 & \boxed{0} & 0 \end{pmatrix} \\[20pt] S_2 = \begin{pmatrix} \boxed{0} & 4 & 4 & 0 \\ 0 & \boxed{0} & 4 & 2 \\ 4 & 0 & \boxed{0} & 4 \\ 0 & 2 & 0 & \boxed{0} \end{pmatrix} \end{cases}$$

There are two optimal solutions, S_1 and S_2. The voice leadings associated to those solutions are:

$$S_1 = C \longrightarrow C, A \longrightarrow B, F \longrightarrow G, D \longrightarrow E$$
$$S_2 = C \longrightarrow B, A \longrightarrow G, F \longrightarrow E, D \longrightarrow C$$

The first voice leading yields a positive-definite endomorphism, that is, all the voices move up and therefore the tonal function of that chord with this voice leading is subdominant. However, all the voices in the second voice leading go down and according to our classification this chord is a dominant chord. Chords with two tonal functions are called **dual chords**.

The optimal voice leading need not be unique also in the case of a different numbers of voices. In the example given above, where the number of voices increased, we computed the optimal voice leading between the chord $X = (G, B, D)$ and the tonal center $TC = (C, E, G, B)$. The solution obtained then was $D \longrightarrow E, B \longrightarrow B, G \longrightarrow G$ and $[\emptyset] \longrightarrow C$. The algorithm actually returns another solution, which is $D \longrightarrow C, B \longrightarrow B, G \longrightarrow G$ and $[\emptyset] \longrightarrow E$. Similar example can be found for the case in which the number of voices decreases.

4 Chord Classification

For the first example, let us fix the number of voices in the chord as 3 (the case is many musical traditions). Let C major be the tonal center. Table 1 shows the tonal functions of the chords in the scale computed by our model.

Table 1. Three-note chords and their tonal function

Function	Chords
Tonic	C, Am Em, Csus2, Csus4
Subdominant	G, B°
Dominant	Dm, F

Surprisingly enough, the functions of dominant and subdominant seemed to be swapped, but there is no mistake. For example, in the case of the voice leading $(G, B, D) \longrightarrow (C, E, G)$ the optimal voice leading consists of two voices moving up, $B \longrightarrow C, D \longrightarrow E$, which is subdominant function. However, if we consider the dominant seventh chord instead of just the chord on the fifth degree of the scale, then the voice leading $(G, B, D, F) \longrightarrow (C, E, G, B)$ does have dominant function. Its optimal voice leading is

$$G \longrightarrow G, B \longrightarrow B, D \longrightarrow C, F \longrightarrow E$$

where all the voices either stay constant or move down. Compare this table to the table of 4-note chords given in the first paper of this series [5].

5 Modulation

The study of tonal function through optimal voice leadings can be extended to the study of modulation. Instead of dealing with a variable number of voices in

the chords, we simply consider the whole scale and apply the same techniques to compute the distance between the original tonality and the target tonality. As we will see later, this procedure will produce a classification of the modulations similar to that of chords. To illustrate this idea, suppose we wish modulate from C major to F major. By applying the same procedure as in obtaining the tonal function, we arrive at:

$$P = \begin{pmatrix} B & E \\ E & A \\ A & D \\ D & G \\ G & C \\ C & F \\ F & B\flat \end{pmatrix} \longrightarrow L = \begin{pmatrix} 5\,2\,3\,4\,1\,6\,1 \\ 0\,5\,2\,3\,4\,1\,6 \\ 5\,0\,5\,2\,4\,4\,1 \\ 2\,5\,0\,5\,2\,4\,4 \\ 3\,2\,5\,0\,5\,2\,3 \\ 4\,3\,2\,5\,0\,5\,2 \\ 1\,4\,3\,2\,5\,0\,5 \end{pmatrix} \xrightarrow{\text{Step 1}} \begin{pmatrix} 4\,1\,2\,3\,0\,5\,0 \\ 0\,5\,2\,3\,4\,1\,6 \\ 5\,0\,5\,2\,4\,4\,1 \\ 2\,5\,0\,5\,2\,4\,4 \\ 3\,2\,5\,0\,5\,2\,3 \\ 4\,3\,2\,5\,0\,5\,2 \\ 1\,4\,3\,2\,5\,0\,5 \end{pmatrix}$$

$$\xrightarrow{\text{Step 2}} \begin{pmatrix} 4\,1\,2\,3\,0\,5\,0 \\ 0\,5\,2\,3\,4\,1\,6 \\ 5\,0\,5\,2\,4\,4\,1 \\ 2\,5\,0\,5\,2\,4\,4 \\ 3\,2\,5\,0\,5\,2\,3 \\ 4\,3\,2\,5\,0\,5\,2 \\ 1\,4\,3\,2\,5\,0\,5 \end{pmatrix} \xrightarrow{\text{Step 3}} \begin{pmatrix} 4\,1\,2\,3\,0\,5\,0 \\ 0\,5\,2\,3\,4\,1\,6 \\ 5\,0\,5\,2\,4\,4\,1 \\ 2\,5\,0\,5\,2\,4\,4 \\ 3\,2\,5\,0\,5\,2\,3 \\ 4\,3\,2\,5\,0\,5\,2 \\ 1\,4\,3\,2\,5\,0\,5 \end{pmatrix} \xrightarrow{\text{Step 4}} \begin{pmatrix} -\,-\,-\,-\,-\,-\,\boxed{0} \\ \boxed{0}\,-\,-\,-\,-\,-\,- \\ -\,\boxed{0}\,-\,-\,-\,-\,- \\ -\,-\,\boxed{0}\,-\,-\,-\,- \\ -\,-\,-\,\boxed{0}\,-\,-\,- \\ -\,-\,-\,-\,\boxed{0}\,-\,- \\ -\,-\,-\,-\,-\,\boxed{0}\,- \end{pmatrix}$$

The voice leading keeps all the voices except B, which is taken to $B\flat$, as one would expect.

As in the case of chords, an endomorphism mapping the original tonality onto the target tonality can be defined. The endomorphism matrix will consist of a diagonal matrix with entries of the form s^λ where λ are the minima output by the Hungarian algorithm along with sign given by the direction of the voice movement (positive if it moves up; negative otherwise).

Let us give a more complicated example to illustrate this model of modulation. Take C major (C, D, E, F, G, A, B) and $G\flat$ major $(G\flat, A\flat, B\flat, C\flat, D\flat, E\flat, F)$. The Hungarian algorithm outputs two solutions S_1, S_2 (we save the computations at this moment as they are not needed for our point):

$$S_1 : C \longrightarrow C\sharp, D \longrightarrow E\flat, E \longrightarrow F\sharp, F \longrightarrow F,$$
$$G \longrightarrow G\sharp, A \longrightarrow B\flat, B \longrightarrow B$$
$$S_2 : C \longrightarrow B\flat, D \longrightarrow C\sharp, E \longrightarrow E\flat, F \longrightarrow F,$$
$$G \longrightarrow F\sharp, A \longrightarrow G\sharp, B \longrightarrow B$$

In S_1 all the voices move up whereas in S_2 all the voices move down. Again, we can state that the modulation from C major to $G\flat$ major possess a dual character as it is possible to go from the former to the latter in two manners.

Mimicking the chord classification given in the first paper, it is possible to classify modulations according to how the original tonality is converted into the target tonality. This is given in terms of the exponents of the endomorphism

associated to the optimal voice leading. The modulation classification criteria are as follows.

Modulation classification criteria

- **Dominant modulation:** A modulation is said a tonic modulation if all the exponents are negative or zero having at least two negative exponents;
- **Weak dominant modulation:** A modulation is weak dominant if there is just one positive exponent, two or more negative exponents, and the rest of the exponents are zero;
- **Subdominant modulation:** A modulation is said a subdominant modulation if all the exponents are positive or zero having at least two positive exponents;
- **Weak subdominant modulation:** A modulation is weak subdominant if there is just one negative exponent, two or more positive exponents, and the rest of the exponents are zero;
- **Tonic modulation:** A modulation is said a tonic modulation if all the exponents are zero except one, which can be either negative or positive.

6   Conclusions

In these two papers we have presented a mathematical model of tonal function based on optimal voice leadings. Unlike other tonal function models based on theories of consonance/dissonance (Riemann's models [6]) or on operations on chords (neo-Riemannian theory [2,3]), our model is based on the optimal voice leading from a chord to a fix tonal center. Once the optimal leading is constructed (via the Hungarian algorithm), the manner in which the voices move towards the tonal center is key to deciding the final tonal function of the chord. Essentially, if most of the voices stay constant, then the tonal function assigned is tonic. If most of the voices move down, then the chord has dominant function; and finally, if most of the voices move up, then the chord has subdominant. Certain chords has no tonal function in the sense they do not fall in the categories for tonal function defined here. Those chords without a clear tonal function have several voices moving down and up at the same time (in our terminology, their associate endormorphisms are either mixed-definite or mixed-semidefinite). Moreover, some chords may have two functions depending whether the voices move up or down when approaching the tonal center, the so-called dual chords. This subtle distinction is new in tonal function classification (to the best of our knowledge).

Furthermore, our model of tonal function allows broader generalization and applicability. For instance, the tonal function of chords of different numbers of voices can always be determined (via the extended Hungarian algorithm) and therefore the chord classification criteria can be extended to an arbitrary number of voices. This is due to the fact that the model is purely based on voice leadings.

Chords such as polychords, quartal chords, or 9th, 11th, 13th chords can be classified under our model. The model presented here can easily be applied to music from the extended common practice such as pandiatonicism or twelve-tone serialism. In particular, we contend that our model is especially suitable for the analysis of jazz music.

The idea of classifying chords according to their optimal voice leadings can be abstracted away to scales and modes. This classification provides an insight into the relationship between tonalities. The modulation between two tonalities is understood both in terms of the new tonal center but also in terms of the optimal voice leading linking both tonalities. In this paper, we only looked into major and minor scales, but as future work we intend to consider modulation between two arbitrary scales, even if they have different number of notes (which our model can handle).

References

1. Berkman, D.: The Jazz Harmony. Sher Music Co., Petaluma (2013)
2. Gollin, E., Rehding, A.: Oxford Handbook of Neo-Riemannian Music Theories. Oxford University Press, New York (2011)
3. Lewin, D.: Generalized Musical Intervals and Transformations. Yale University Press, New Haven (1987)
4. Mulholland, J., Hojnacki, T.: The Berklee Book of Jazz Harmony. Berklee Press, Boston (2013)
5. Pozo, I., Gómez, F.: A mathematical model of tonal function (I): voice leadings. In: Proceedings of Mathematics and Computation in Music 2019, Atlanta, USA, June 2022
6. Riemann, H.: Harmony simplified: or, the theory of the tonal functions of chords. Augener Ltd., London (1896). Translation of Vereinfachte Harmonielehre (1893). Republished by Alpha Editions, London, September 2019
7. Russell, G.: The Lydian Chromatic Concept of Tonal Organization. Concept Publishing Company, Delhi (1959)
8. Stover, C.: Jazz harmony: a progress report. J. Jazz Stud. **10**(2), 157–197 (2014)
9. Terefenko, D.: Jazz Theory: From Basic to Advanced Study. Routledge, New York (2014)

Hypercube + Rubik's Cube + Music = HyperCubeHarmonic

Maria Mannone[1,2(✉)], Takashi Yoshino[3], Pascal Chiu[4],
and Yoshifumi Kitamura[4]

[1] Department of Engineering, University of Palermo, Palermo, Italy
`mariacaterina.mannone@unipa.it`
[2] Dipartimento di Scienze Ambientali, Informatica e Statistica (DAIS)
and European Centre for Living Technology (ECLT),
Ca' Foscari University of Venice, Venice, Italy
`maria.mannone@unive.it`
[3] Department of Mechanical Engineering, Toyo University, Kawagoe, Japan
`tyoshino@toyo.jp`
[4] Research Institute of Electrical Communication, Tohoku University, Sendai, Japan
`kitamura@riec.tohoku.ac.jp`

Abstract. Musical chords and chord relations can be described through mathematics. Abstract permutations can be visualized through the Rubik's cube, born as a pedagogical device [7,21]. Permutations of notes can also be heard through the CubeHarmonic, a novel musical instrument. Here, we summarize the basic ideas and the state of the art of the physical implementation of CubeHarmonic, discussing its conceptual lifting up to the fourth dimension, with the HyperCubeHarmonic (HCH). We present the basics of the hypercube theory and of the 4-dimensional Rubik's cube, investigating its potential for musical applications. To gain intuition about HCH complexity, we present two practical implementations of HCH based on the three-dimensional development of the hypercube. The first requires a laptop and no other devices; the second involves a physical Rubik's cube enhanced through augmented and virtual reality and a specifically-designed mobile app. HCH, as an augmented musical instrument, opens new scenarios for STEAM teaching and performing, allowing us to hear the "sound of multiple dimensions."

Keywords: Rubik's cube · Hypergeometry · Permutations · Chords · Tonnetz · Mobile

1 Introduction

Pythagoras advanced joint knowledge of mathematics and music. In the Middle Ages, they were part of the Quadrivium. Today, their exchanges are leading to a flourishing research field [20]. Classic examples of interactions between mathematics and music are the permutations of pitches as well as entire musical sequences, as in Mozart's dice game. Here, we connect (hyper)geometry,

combinatorics, and music, presenting the general idea of the HyperCubeHarmonic (HCH). It is a development of the CubeHarmonic (CH) [11], that generates chords according to faces' rotations. The CH is a novel musical instrument inspired by the Rubik's cube and the abstract pitch spaces (a family of representations including the tonnetz).[1] The core idea of CH is the following:

– each facet is a note within an octave;
– each face is a chord (with nine notes, and some of them can be repeated);
– rotating the cube, the initial chords are scrambled.

For reasons of clarity of sound and simplicity, the reproduced sounds correspond to the notes on the top face of the cube. A cycle of rotations corresponds to a cycle of chords. In fact, we can investigate chord-preserving symmetries of the Rubik's cube. For example, the cube's rotational symmetry allows cyclical chord progressions. We notice that the symmetry of harmonically-played chords is bigger than the symmetry of melodically-played chords. In fact, a rotated face corresponds to the same sound of an un-rotated face if the notes are played simultaneously.

A simplification of the unscrambled cube has one pitch for all the facets in each face; twisting the cube, new chords are obtained. From the point of view of music theory, we can see CH as a tangible application of slot-machine transformations.[2] CH involves a topic of mathematical music theory, the *tonnetz* [20], a lattice of pitches and musical chords, which is first cut and glued to be adapted to the cube, and then twisted (and cut again) according to the performed rotations.

The first concept of CubeHarmonic (called CubHarmonic in [12, p. 20.1.2]) is a $4 \times 4 \times 4$ cube, used for a 4-part harmony. Twisting this cube, new chord sequences are obtained. However, all current physical CH prototypes [10,11] use a $3 \times 3 \times 3$ cube. With a 3-part harmony, the performer can still enjoy a considerable amount of combinations, with the advantages of easiness in cube's scrambling.

The very first implementation of the CH was realized by M. Mannone through a giant Rubik's cube and sound modules, allowing the user/performer to play each note separately or simultaneously,[3] see Fig. 1. It is described in [10]. The disadvantage of size was overcome by IM3D+ use [8], as described below.

The $3 \times 3 \times 3$ prototype built at the Tohoku University,[4] see Fig. 2, needs only a few markers (LC coils) for 3D motion tracking [11], while a $4 \times 4 \times 4$ cube would present more technical issues regarding physical motion capture.

---

[1] The tonnetz is a lattice constituted by notes and their connections as chords [20]. The CH has been thought by M. Mannone during her studies at IRCAM in 2013, and then first described in [12].

[2] In music theory, slot-machine transformations are permutations. If we have three discs with three notes in each of them, they give a sequence of 3-note chords. Rotating the discs, the chord changes. For example, the vertical sequence $0 - 1 - 2$ becomes $1 - 2 - 0$ after a rotation of one of the discs [1].

[3] Video: https://tinyurl.com/3j5csh36.

[4] Video: https://www.youtube.com/watch?v=r_wNpQnsWhg.

In fact, in the current CH prototype, the IM3D+ platform [8] allows motion tracking through battery-less and wireless small lightweight identifiable passive sensors modifying a magnetic field (Fig. 2). If these sensors are embedded in HCH, more degrees of freedom are allowed: seven more Rubik's cubes and all their intertwined rotations.

The CH and HCH have a good potential not only as music instruments but also as teaching tools for STEAM (Science, Technology, Engineering, Art, and Mathematics) education [3] in the domain of mathematics [9,14,15], because of their specific features described in Sects. 2 and 3. In fact, HCH joins the abstraction of hypergeometry with the tangibility of a physical tool and the resources of computer visualization.
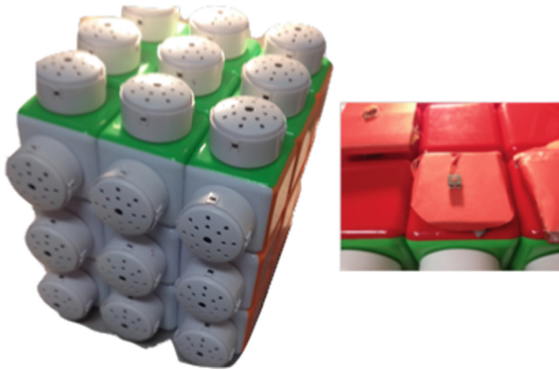
Fig. 1. First implementation of CubeHarmonic, from [10].
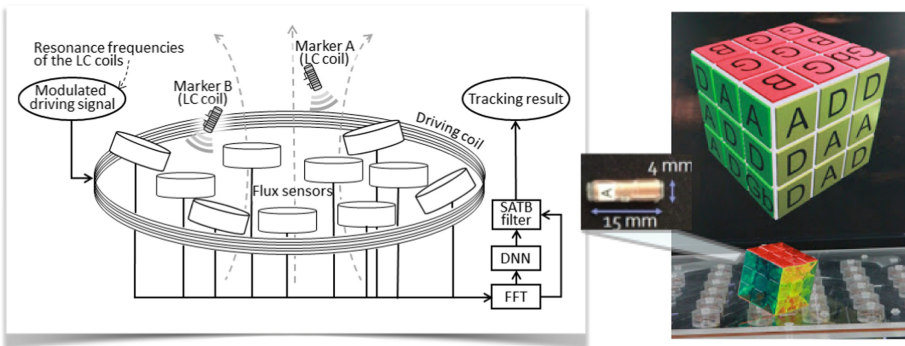
Fig. 2. IM3D+ (left) and CubeHarmonic (right) [11].

Other independently-developed applications of music to the Rubik's cube are described in a NIME article [16] and in a TED tutorial [19]. The TED tutorial concerns a similar theoretical idea, without however mentioning chord sequences

or proposing the development of a physical instrument. Moreover, the tutorial dates back to November 2015, while the book where the CubeHarmonic had first been described [12] was submitted to Springer in September 2015. The article [16], published one year after the book [12], proposes a new device, with a random choice of pitch, to be used as a controller. However, the starting idea of CH is theoretical, and its implementation is physical and tangible. Also, in CH [10,11] some musical features, such as the possibility to manipulate overall loudness and pitch changes according to hands' movements, are all implemented, creating crossmodal effects. These features will be added to future versions of HCH as well.

HCH is inspired by the multidimensional geometry, and it will allow the transformation of the instrument's structure during performance, besides adding more parameters to the existing system. This results in an innovative instrument which enables rich manipulations of tones and other musical parameters while making notes, melodies, chords, and timbre transitions interactive and appreciably tangible. Thus, the HCH should allow musicians to hear the "sound of multiple dimensions," by extending a customizable physical cube into additional dimensions in virtual reality. An object living in the 4 dimensions cannot be completely realized in our 3-dimensional world. However, augmented reality will help us gain intuition about this research.

Thus, we discuss the theory behind 4-dimensional Rubik's cube and its potential for musical application. Then, we present two first implementations. The first one implements a real hyper-Rubik's cube and allows the user/musician to play it; it is based on Mathematica^{TM}, and requires only a laptop. The second one exploits a physical cube and a mobile device. Thus, HCH can be totally mobile, because it does not require any platform. However, if small sensors are embedded in the corners of the cube, it can be used jointly with the former technology for CH, adding more degrees of freedom—and thus, more potential features. HCH can exploit all resources of musical mobile connectivity (both via cables, MIDI to computer or to sound interface, and wireless with bluetooth), and it has a potential as a controller.

The mathematical idea is described in Sect. 2; the implementations in Sect. 3; discussions on potentialities and further developments are proposed in Sect. 4.

2   Mathematical Concepts

2.1   *Ars Combinatoria*, Permutations, and the Rubik's Cube

Gottfried Wilhelm von Leibniz defined as *ars combinatoria*[5] the technique of ideas' symbolization through geometric and algebraic signs and their recombination and organization in all possible ways, finalized to create a universal map of concepts [4]. Then, the computational resources of *ars combinatoria* inspired artistic applications through the centuries, in music, visual arts, and literature [27]. Well-known examples include musical dice games (Musikalische

---

[5]   It had been called *ars magna* by Ramon Llull [4].

Würfelspiele), which allowed to compose musical pieces from random combinations of pre-composed musical fragments—see some works by Llull, Haydn, and Mozart [1,17]. Mathematically, combinatorics is based on group theory. A *group* is constituted by a set and an operation such as the combination of two elements of a group gives an element of the same group. The operation verifies closure, associativity, identity, and invertibility [18]. A *permutation group* is a group whose elements are the permutations, and whose operations are the permutation compositions.

Invented as a pedagogical device to teach permutations, the Rubik's cube, thanks to its colorful aesthetic appeal and manipulative attractiveness, became one of the best-sold toys and games. The design of the Rubik's cube is based on permutation groups. The facets are the set, and their possible permutations (moves) constitute the group operations. Sets of moves are the subgroups of the Rubik's cube, and all possible moves belong to this group [26].

2.2   The Hypercube

Our hypercube has four dimensions; it is a tesseract and it can be developed with eight cubes in three dimensions [23,24]. The cubes are regarded to be located oppositely on the four axes of the Cartesian coordinates in a hyperspace. Hypercubes sparkled the interest of music theorists; hypergeometry led to dynamic visualizations of symmetries in musical pieces [2,5]. The famous way to understand the hypercube is extending the relation between 2D and 3D to 3D and 4D. For example, a square (2D) moving in the space (3D) for a length equal to its side and reaching another square, spans the space of a cube (3D). A cube, moving in space (4D) and reaching another cube, spans the space of a hypercube (4D). In short, a dimensionally-lifted square gives a cube, and a dimensionally-lifted cube gives a hypercube. Figure 3a shows a cube and its central projection, and Fig. 3b presents a hypercube through its center projection. Fig. 3c illustrates the expansion of a hypercube with eight cubes on a hypersurface; this structure
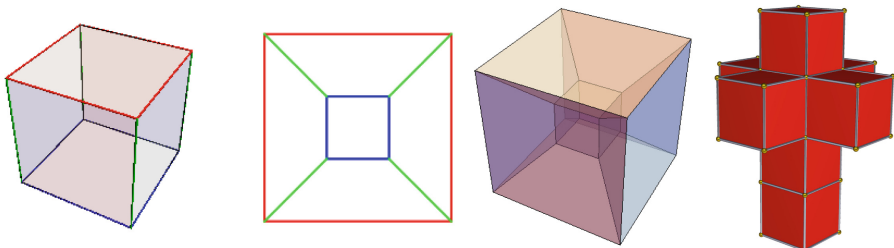
(a) A cube and its representation via a central projection [25].

(b) Hypercube center projection [25].

(c) Hypercube 3D development [22].

Fig. 3. Some representations of cube and hypercube.

has been used in Dalí's painting *Crucifixion (Corpus Hypercubus)* [6]. The idea of hypercube also inspired music-tech applications, not related with the Rubik's cube; as an example, see the *Sound-reactive cube.*[6]

2.3   The Hyper-Rubik's Cube

As the next step, we imagine that each cube of the hypercube is a Rubik's cube. Fig. 4 shows a 3D projection of a 4D Rubik's cube, and its constitutive eight cubes. In this subsection, we assume that all faces of each cube are painted the same color, for simplicity. However, we can paint them more generally: different colors to different subcubes or to different facets.
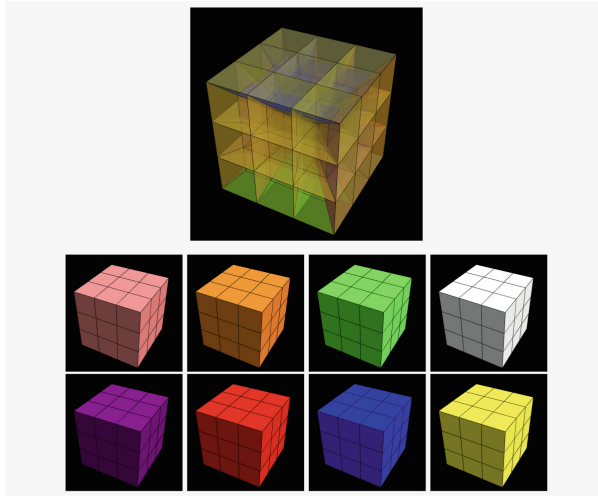
Fig. 4. Development of the Hyper Rubik's cube. Image from [25].

The traditional Rubik's cube has 27 subcubes; the center cube is not colored (ideally, it is in the inner core of the cube, where there is the rotational mechanism), and the other cubes are partially colored (according if they are corner, lateral, or central pieces). The colored pieces are "faces × facets" = $6 \times 9 = 54$. It can be shown [25] that the hyper-Rubik's cube contains overall 81 small 4-cubes; one of the cubes is at the center and it is not colored, while the remaining 80 are partially colored. The colored subcubes in a 4-Rubik's cube are: cubes × subcubes = $8 \times 27 = 216$. For the people in 3D, the subcube at the center of each cube is hidden so the value is 26. However, for the people in 4D, the center can be observed, maybe. This is the reason why we are considering 27 rather than 26. A rotation on one cube provokes some rotation of the adjacent cubes. According to [13,25], while in 3D an axis of rotation is a line, in 4D it

---

is a plane. E.g., the rotation matrix $R_{x,y}(\theta)$ with respect to the $(x - y)$-plane is given in Eq. (1); because there are six planes of rotations, there are six rotation matrices.

$$R_{x,y}(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\theta & -\sin\theta \\ 0 & 0 & \sin\theta & \cos\theta \end{pmatrix} \tag{1}$$

We refer to the smallest cluster of subcubes to be rotated simultaneously as a *block*. In the case of Rubik's cube, a block is a prism of $3 \times 3 \times 1$ subcubes. Similarly, for hyper-Rubik cube, a block is a hyperprism consisting of $3 \times 3 \times 3 \times 1$ small hypercubes. The operation to rotate a block is described in [25].

2.4   The Hyper-Rubik's Cube with Music: The HyperCubeHarmonic

If we have 8 different Rubik's cubes, we have more degrees of freedom to be used musically. For example, all cubes could contain the same notes, but with different timbres. Alternatively, each cube can play the same note, or there can be different notes for different squares. A different choice of musical layers could lead, for example, to rhythm-combination controllers. We can choose to play all cubes together or just one after the other. In any case, eight cubes remind one of the mentioned 3D-development of the hypercube.

   If each cube is a CH, adding more cubes means adding degrees of freedom, that is, more musical resources to the instrument. Figure 5 shows the concept of HCH. For the geometric discussion, see [24]. In the $3 \times 3 \times 3$ Rubik's cube used for CH, each facet is a single note; in HCH each cube of the tesseract is a CH. Thus, higher geometric dimensions allow higher musical complexity.
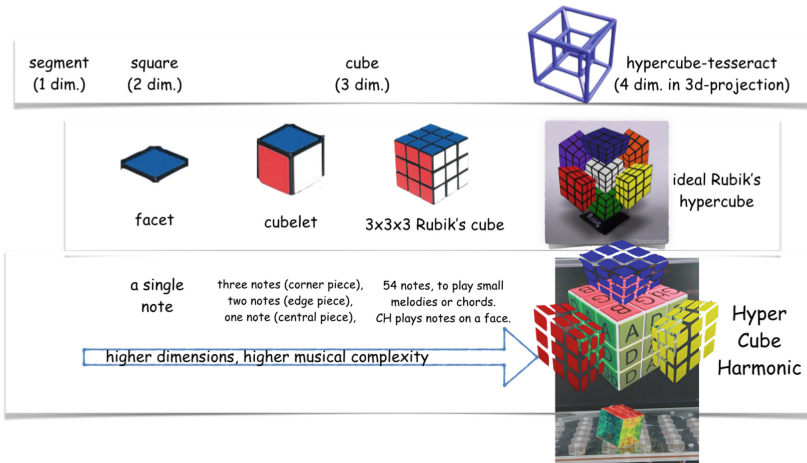
Fig. 5. The concept of HyperCubeHarmonic

3 First Implementations

To have a grasp of the musical potentialities of the Rubik's hypercube, we present here two implementations.

3.1 Implementation 1

As a first computational implementation, for the sake of simplicity we put the same note on each one of the eight cubes (e.g., creating a scale C-D-E-F-G-A-B-C octave), visually identifying each note and cube with a color. In this way, the result of each rotation will straightforwardly be visualized and heard. This implementation, coded in Mathematica^{TM} by T. Yoshino, can be accessed online.[7]
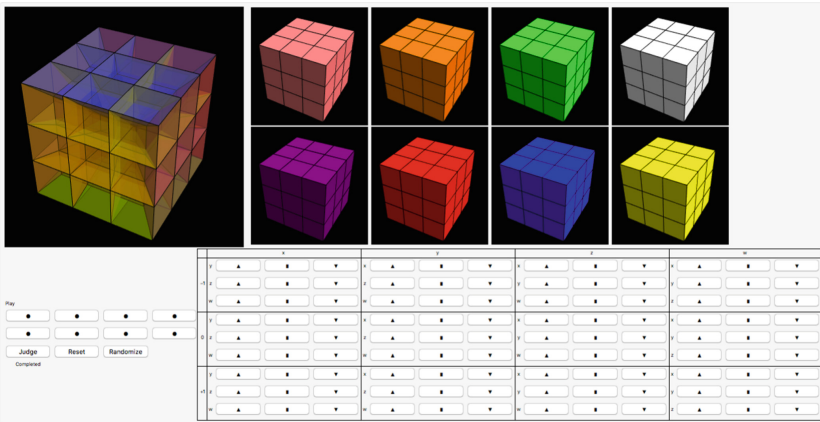
Fig. 6. Mathematica^{TM} rendition of the unscrambled HyperCubeHarmonic. Pink: C4; Orange: D4; Green: E4; White: F4; Yellow: G4; Blue: A4; Red: B4; Violet: C5. (Color figure online)

Figure 6 shows a screenshot of this implementation. The sound(s) of each cube can be heard pressing the round buttons. The table on the right contains the commands for the rotations: the black triangle indicates a rotation angle of $\pi/2$, the black square of π, and the black inverted triangle of $-\pi/2$. As an example, Figs. 7 and 8 show the result of a rotation of $-\pi$ of the y-column and x-row. Our video shows how this app works.[8]

---

[7] http://random-walk.org/PrototypeHyperCH.nb.

[8] https://youtu.be/wB8VoCKHrmc.
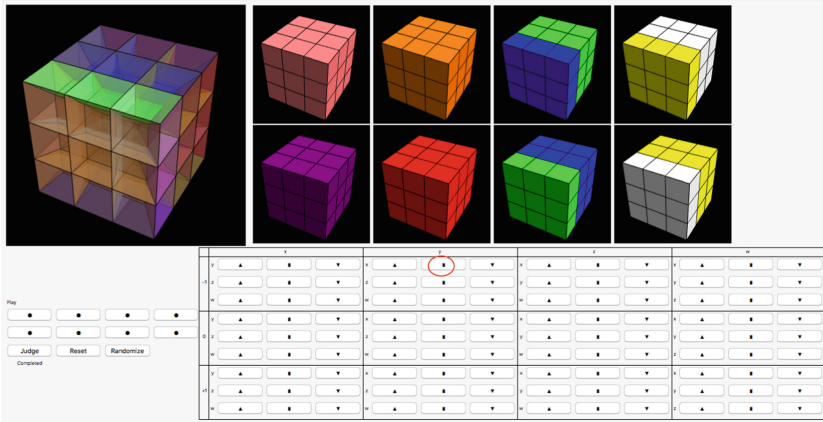
Fig. 7. HyperCubeHarmonic with the highlighted rotation.
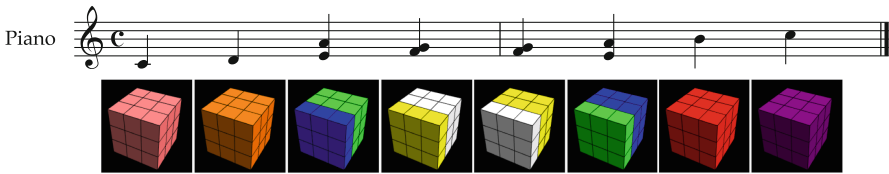
Fig. 8. Musical effect of the rotation of Fig. 7.

3.2   Implementation 2

We can also simplify the HCH 3D projection and develop a manipulative device used together with an app. The user/performer has a physical device, a Rubik's cube with embedded sensors, and a mobile with the specific app coded in Unity. The physical cube is the "Go Cube,"[9] which contains embodied, inner sensors transmitting information about the position to a mobile app. The mobile app (Fig. 9a) has been specifically coded by P. Chiu for HCH, working for Android and iOS mobile operative systems. In this app, the cubes are considered as independent; this is why they are not distinguished by colors.

The Go Cube + app system implements an IoT (internet of things, connected smart-objects) approach. The app screen shows, in real time, the position of the facets of the real cube, and other additional seven cubes. The user can select specific cubes to create a local rotation. In doing so, we do the simplification of independent cubes. In future releases, these rotations will be related between them according to the hypercube's geometrical constraints [25].

Even if the physical cube is scrambled, the cube(s) in the app can be reset to the unscrambled state. Thus, HCH does not require the ability to solve the Rubik's cube to be played. The musical parameters in the current prototype are

---

[9] https://getgocube.com/.

(a) Opening the app.

(b) Selection of one cube.

(c) All cubes are selected and partially scrambled.

(d) Long-tapping on it, we can scramble only one cube.
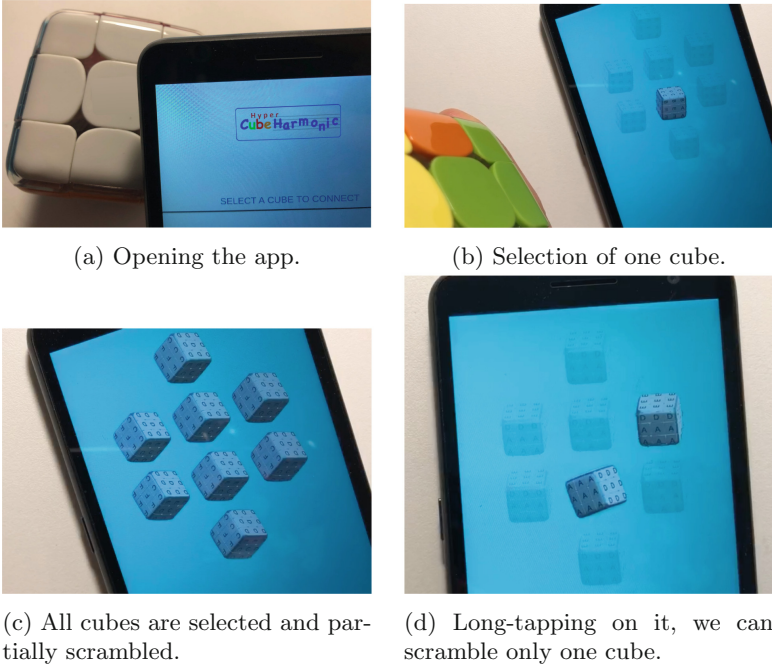
Fig. 9. HCH mobile app

the same pitches with different timbres: each cube has a different timbre. Sounds come from the sampled orchestral sounds.[10]

The top face of each cube is the one which sounds. If the rotation does not involve the top face, the sound does not change. HCH development is still ongoing, to enhance sound-movement correspondence and use easiness. Next developments will include rhythms and sound effects, to make the performance richer. Our video[11] shows how the app prototype works, with the original (unedited) sound. A single tap enables a cube to play (Fig. 9b). As default, all cubes are moving (Fig. 9c), but a long-tap on one cube disables scrambling for the other cubes, letting them keep the same notes (Fig. 9d).

4 Discussion and Conclusions

From a theoretical idea, the CubeHarmonic has been little by little shaped into a working musical instrument. We described the idea of CH and its 4-dimensional development, including two computational implementations. The presented extension to the fourth dimension broadens the scenarios in musical applications and conceptual developments. One of the proposed implementations

---

[10] https://philharmonia.co.uk/resources/sound-samples/.
[11] https://youtu.be/p5i6_uzIips.

of HCH has already been used in musical setups such as multi-instrument impro-
visations.[12] We are planning to join both prototypes to use a physical cube as a
controller for the Mathematica app with the real HyperCubeHarmonic (Fig. 10).

Fig. 10. The logo of HyperCubeHarmonic project.

The main advantage of the presented simplified implementation of HCH with
respect to CH is the portability. Also, it can be connected (through the mobile)
wireless with bluetooth, or through Sonobus software, allowing a stable stream
transmission, ideal for remote collaborative performances. HCH could also be
used as a MIDI controller if the input is transmitted to an audio interface. The
use of sequencers such as Bitwig, Mulab, or Ableton will allow HCH playing
recordings. Further developments may include a detailed user study, and the
application of HCH concept to other Rubik's Platonic Solids.

Connectivity of HCH makes it suitable for remote STEAM teaching [3,9]. In
fact, HCH might have a potential in both mathematical and musical education,
as it makes complex abstract objects tangible, by manipulating their parameters
interactively. HCH can be creatively used to generate chords in correspondence
of non-intuitive Rubik's hypercube rotations. In particular, students could recog-
nize and master combinations of rotations along with their inverse through their
musical effect. A classroom network of connected HCH might allow the teacher
to give feedback on students' learning. The hour of math could be turned into
an HCH orchestra rehearsal!

The progressive addition of multiple cubes as in HCH could help enhance
creativity, mixing timbres and developing motor abilities, through the com-
mon ground of hearing and movement. Finally, HCH, through its multi-sensory
modalities, might hopefully help students with visual or hearing impairments
learn mathematics and solve the Rubik's cube.

References

1. Alegant, B., Mead, A.: Having the last word: Schoenberg and the ultimate cadenza.
   Music Theor. Spectr. **34**(2), 107–136 (2012)
2. Amiot, E., Baroin, G.: Old and new isometries between pc sets in the planet-4D
   model. Music Theor. Online **21**(3) (2015)
3. Andreotti, E., Frans, R.: The connection between physics, engineering and music
   as an example of STEAM education. Phys. Educ. **54**, 045016 (2019)

---

[12] *Night Forest*: https://www.youtube.com/watch?v=1OClm1pn7-g&t=1s, *Blues and Grays*: https://www.youtube.com/watch?v=oMldYLbfIRE, Female Laptop Orches-
tra (FLO).

4. Ars combinatoria. Treccani Enciclopedia (in Italian). https://www.treccani.it/enciclopedia/ars-combinatoria/

5. Baroin, G.: The planet-4D model: an original hypersymmetric music space based on graph theory. In: Agon, C., Andreatta, M., Assayag, G., Amiot, E., Bresson, J., Mandereau, J. (eds.) Mathematics and Computation in Music, MCM 2011. LNCS (LNAI), vol. 6726, pp. 326–329. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21590-2_25

6. D'amore, B.: Salvador Dalì e la geometria a quattro dimensioni (2015). http://www.saperescienza.it/biologia/salvador-dali-e-la-geometria-a-quattro-dimensioni-18-2-15/

7. Hofstadter, D.: Metamagical Themas, March 1981. The Magic Cube's cubies are twiddled by cubists and solved by cubemeisters. Scientific American (1981). https://www.scientificamerican.com/article/metamagical-themas-1981-03/

8. Huang, J., Sugawara, R., Chu, K., Komura, T., Kitamura, Y.: Reconstruction of dexterous 3D motion data from a flexible magnetic sensor with deep learning and structure-aware filtering. IEEE Trans. Vis. Comput. Graph. (2020). https://doi.org/10.1109/TVCG.2020.3031632

9. Mannone, M.: Have fun with math and music! In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O.A. (eds.) Mathematics and Computation in Music, MCM 2019. LNCS (LNAI), vol. 11502, pp. 379–382. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_33

10. Mannone, M., Kitamura, E., Huang, J., Sugawara, R., Kitamura, Y.: Musical combinatorics, tonnetz, and the cubeharmonic. Collect. Pap. Acad. Arts Novi Sad. **6**, 137–153 (2018). Zbornik Radova Akademije Umetnosti

11. Mannone, M., Kitamura, E., Huang, J., Sugawara, R., Chiu, P., Kitamura, Y.: CubeHarmonic: a new musical instrument based on Rubik's cube with embedded motion sensor. ACM SIGGRAPH Posters, Article no. 53 (2019). https://doi.org/10.1145/3306214.3338572

12. Mazzola, G., Mannone, M., Pang, Y.: Cool Math for Hot Music. CMS. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42937-3

13. Miyazaki, K., Ishii, M., Yamaguchi, S.: Science of Higher-Dimensional Shape and Symmetry. Kyoto University Press, Kyoto (2005). (in Japanese)

14. Montiel, M., Gómez, F.: Music in the pedagogy of mathematics. J. Math. Music **8**(2), 151–166 (2014)

15. Montiel, M., Gómez, F. (eds.): Theoretical and Practical Pedagogy of Mathematical Music Theory: Music for Mathematics and Mathematics for Music. World Scientific, From School to Postgraduate Levels, Singapore (2018)

16. Polfreman, R., Oliver, B.: Rubik's cube, music's cube. In: NIME Conference Proceedings (2017)

17. Riepel, J.: Grundregeln zur Tonordnung Insgemein. Nabu Press, Charleston (1755, 2014)

18. Rotman, J.: An Introduction to the Theory of Groups. Springer, New York (1991). https://doi.org/10.1007/978-1-4612-4176-8

19. Staff, M.: How to play a Rubik's cube like a piano (2015). https://ed.ted.com/lessons/group-theory-101-how-to-play-a-rubik-s-cube-like-a-piano-michael-staff

20. Tymoczko, D.: The geometry of musical chords. Science **313**, 72–74 (2006)

21. Undark, H.P.: A brief history of the Rubik's cube. Smithsonian Mag. (2020) https://www.smithsonianmag.com/innovation/brief-history-rubiks-cube-180975911/

22. Webb, R.: Image and stella software. http://www.software3d.com/Stella.php

23. Weisstein, E.W.: Hypercube. From MathWorld-A Wolfram Web Resource (2014). https://mathworld.wolfram.com/Hypercube.html
24. Williams, R.: The Geometric Foundation of Natural Structure. Dover, New York (1979)
25. Yoshino, T.: Rubik's 4-cube. Math. J. (2017). https://www.mathematica-journal.com/2017/12/31/rubiks-4-cube/
26. Zassenhaus, H.: Rubik's cube: a toy, a galois tool, group theory for everybody. Phys. A **114**, 629–637 (1982)
27. Zweig, J.: Ars Combinatoria. Art J. **56**(3), 20–29 (2014)

Applications of Mathematics to Musical Analysis

Mathematical Morphology Operators for Harmonic Analysis

Gonzalo Romero-García[1(✉)] , Isabelle Bloch[2] , and Carlos Agón[1]

[1] Sorbonne Université, CNRS, IRCAM, STMS, Paris, France
{romero,agonc}@ircam.fr
[2] Sorbonne Université, CNRS, LIP6, Paris, France
isabelle.bloch@sorbonne-universite.fr

Abstract. Mathematical Morphology provides powerful tools for image processing, analysis and understanding. In this paper, we apply these tools to analyze scores, that are image-like representations of Music. To do that, we consider *chroma rolls*, a representation of scores similar to piano rolls that use chromas instead of pitches. Endowing this representation with a lattice structure, one can define Mathematical Morphology operators, and setting a group structure to the Time-Frequency plane allows us to use the notion of *structuring element*. We show throughout some examples that this relates with the notion of pitch-class set and chord progressions, and we analyze two Chopin's Nocturnes with this technique.

Keywords: Mathematical Morphology · Harmonic analysis ·
Time-frequency · Pitch-class set · Chord progression · Chroma roll

1 Introduction

Mathematical Morphology is a theory for the analysis and processing of geometrical structures that intersects with several domains such as Topology, Lattice Theory or Integral Geometry [6,9]. It has been developed in the second half of the 20th century and it has been extensively used for analyzing images. Following this approach, we will show that Mathematical Morphology operators may be useful for analyzing image-like representations of Music such as scores. In particular, we will focus on harmonic analysis and we will take as support a two-dimensional representation of the score closely related to the standard piano roll: the *chroma roll*.

There are already a few applications of Mathematical Morphology to Music [3,4], and [5] on this topic that use piano rolls. This paper extends these earlier works in terms of both representations of music and types of operators.

There are several frameworks where Mathematical Morphology may be applied; in its deterministic form, we need to have a complete lattice as algebraic structure. We will ask a bit more, intending to make it possible to work with structuring elements: the lattice will then consist of functions with a group as domain and a complete lattice as codomain. Moreover, since we will make abstraction of the dynamics of the score, we will select as codomain the lattice $(\{0, 1\}, \leq)$, and then we will be able to identify functions with subsets of the group through the characteristic function.

In this paper, we will start by presenting the algebraic framework of the Time-Frequency plane based on Group Theory. Then, we will present the Mathematical Morphology basics that we will use in this work. Finally, we will propose some applications of Mathematical Morphology operators to extract harmonic information. As a particular case, we will focus on Chopin's Nocturnes since they are well adapted to this task.

2   Algebraic Framework

Since we want to model scores, our domain is the Cartesian product of a set modeling time and a set modeling frequency. We require these sets to be groups in order to have a notion of *translation*; this will lead to straightforward definitions of the morphological operators.

We recall that the translation action of a group $(G, +)$ on itself is the function

$$\begin{aligned} T : G \times G &\to G \\ (g, x) &\mapsto T_g x = x + g \end{aligned} \tag{1}$$

In the following sections, we present the corresponding groups we chose for modeling time and frequency, that lead to a time-frequency representation of a score.

2.1   Time Group

There are several groups that may model the time: if we express it in seconds we can choose $(\mathbb{R}, +)$; if we consider the time inside a MIDI file it will be useful to use $(\mathbb{Z}, +)$ where each unit is a tick; if we consider an audio signal we may also use $(\mathbb{Z}, +)$ where each unit is a sample.

When we measure the time inside a score, one can choose several units: among others, we may use either the bar, the beat or the *tatum*[1]. We choose to measure the time inside a score in bars. Letting $d \in \mathbb{N}$ be the ratio between the duration of a bar and the duration of the tatum, we model the time inside a score measured in bars by the group $(\frac{1}{d}\mathbb{Z}, +)$, where $\frac{1}{d}\mathbb{Z} = \{\frac{n}{d} \in \mathbb{Q} : n \in \mathbb{Z}\}$. This group is a subgroup of $(\mathbb{Q}, +)$ but is isomorphic to $(\mathbb{Z}, +)$ using the isomorphism

---

[1] This name was introduced in [1] for calling the maximal note duration such that all the note durations in the score are integer multiples of it. For a deeper discussion about it and its relation with the notion of GCD (Greatest Common Divisor) see [7].

$\phi : \left(\frac{1}{d}\mathbb{Z}, +\right) \to (\mathbb{Z}, +), \frac{n}{d} \mapsto n$. Nevertheless, it is interesting to use it rather than $(\mathbb{Z}, +)$ because measuring the time in multiples of tatum is far less practical that doing it in bars.

2.2 Frequency Groups

For considering the frequency inside a score, we may use the notion of pitch. Usually, pitch is measured in semitones and may be modeled by the set $\mathbb{Z}$ where each number corresponds to the MIDI number of a note (for instance, the number 60 corresponds to C_4).

This frequency group is interesting for analyzing melodies but, when considering harmony, it would be preferable to work with the set $\mathbb{Z}_{12}$ of *chromas*. Then, we use this set with the translation action of $(\mathbb{Z}_{12}, +)$. In order to make the distinction between chromas and translations, we will use letters for the chromas (C, C♯, ..., B) and numbers for the chroma shifts $(\overline{0}, \overline{1}, ..., \overline{11})$, even if chromas have an associated number for computations and are considered as elements of $\mathbb{Z}_{12}$.

2.3 Time-Frequency Groups

Now that we have set the models for time and frequency, we can couple them by means of the direct product and have a time-frequency representation of scores. We then work with the group $\left(\frac{1}{d}\mathbb{Z} \times \mathbb{Z}_{12}, +\right)$ and the action on itself. A score may then be conceived as a subset of the time-frequency plane; piano roll representations are a good way of visualizing this, but they are subsets of $\frac{1}{d}\mathbb{Z} \times \mathbb{Z}$; since we want to work with the time-frequency plane $\frac{1}{d}\mathbb{Z} \times \mathbb{Z}_{12}$, we call *chroma roll* a subset of it. The way of getting a chroma roll from a score is the following: we take the piano roll representation of the score that we call $S \subseteq \frac{1}{d}\mathbb{Z} \times \mathbb{Z}$ and we say that its chroma roll representation is the set $\pi(S) \subseteq \frac{1}{d}\mathbb{Z} \times \mathbb{Z}_{12}$, where $\pi : \mathbb{Z} \to \mathbb{Z}_{12}, n \mapsto \overline{n}$ is the canonical projection from $\mathbb{Z}$ to $\mathbb{Z}_{12}$. See Fig. 1 for an illustration of a chroma roll.

Now that we have an image-like representation of the music with a base group structure, we are able to define some Mathematical Morphology operators that allow us to extract harmonic features of the score.
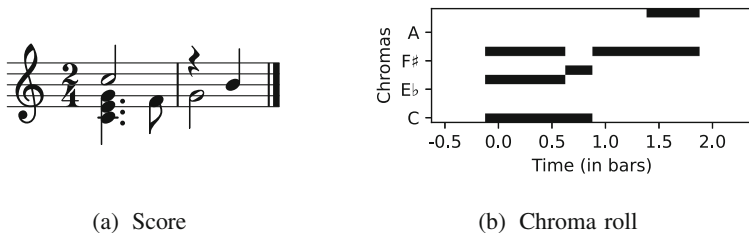

(a) Score   (b) Chroma roll

Fig. 1. Chroma roll representation of a score

3   Mathematical Morphology

Let us now introduce some basic but useful tools of Mathematical Morphology. We start by the two basic operators, which are the *erosion* and the *dilation*.

3.1   Erosion and Dilation

Erosion and dilation are defined in the context of complete lattices. A good introduction to these operators where all the subsequent definitions and propositions may be found is [2].

Definition 1. *Let $(L_1, \leq_1)$ and $(L_2, \leq_2)$ be two complete lattices, and $\wedge_1, \wedge_2$ (respectively $\vee_1, \vee_2$) the associated infimum (respectively supremum).*
    An operator $\varepsilon : L_2 \to L_1$ is called an **erosion** *if*

$$\forall X_2 \subseteq L_2 \,, \, \bigwedge_1 \varepsilon(X_2) = \varepsilon \left(\bigwedge_2 X_2 \right) . \tag{2}$$

An operator $\delta : L_1 \to L_2$ is called a **dilation** *if*

$$\forall X_1 \subseteq L_1 \,, \, \bigvee_2 \delta(X_1) = \delta \left(\bigvee_1 X_1 \right) . \tag{3}$$

An important property of these operators is that they are *increasing*[2] and then preserve the order.
    Whereas this is the most general definition for erosion and dilation, it is an implicit definition. In order to have an explicit type of erosion and dilation (the one with structuring elements), we choose as lattice the power set of a group $(G, +)$. The next proposition provides an explicit definition these types of operators.

Proposition 1. *Let $(G, +)$ be an additive group. Let $T : G \times G \to G, (x, g) \mapsto T_g x = x + g$ be the translation action of $(G, +)$ on itself. We note the translation of a set $A \subseteq G$ by an element $g \in G$ by $T_g A = \{T_g a \in G : a \in A\} \subseteq G$. Let $B \subseteq G$ and $\check{B} = \{-b \in G : b \in B\}$, where $-b$ is the inverse element of b.*
    We consider the complete lattice $(\mathcal{P}(G), \subseteq)$. Then, the operators

$$\varepsilon_B : \mathcal{P}(G) \to \mathcal{P}(G) \tag{4}$$
$$A \quad \mapsto \varepsilon_B(A) = \{x \in G : T_x B \subseteq A\}$$

$$and \quad \delta_B : \mathcal{P}(G) \to \mathcal{P}(G) \tag{5}$$
$$A \quad \mapsto \delta_B(A) = \{x \in G : T_x \check{B} \cap A \neq \emptyset\}$$

are respectively an erosion and a dilation.
    We use the following notations: $\forall A \subseteq G$,

$$\varepsilon_B(A) = A \ominus B \qquad\qquad \delta_B(A) = A \oplus B \,.$$

The set B is called structuring element. *The operators ε_B and δ_B are called respectively* binary erosion *and* binary dilation *with structuring element B.*

---

[2] $\psi : L_1 \to L_2$ is said to be increasing if $\forall X, Y \in L_1, \, X \leq_1 Y \Rightarrow \psi(X) \leq_2 \psi(Y)$.

We notice that for each choice of B we have a different erosion and a different dilation, so, what we actually have is a family of erosions and dilations.

The next property exposes an analogous way of defining these operators making use of union and intersection.

Proposition 2. *Using the notations from Proposition 1,*

$$A \ominus B = \bigcap_{x \in \check{B}} T_x A \qquad\qquad A \oplus B = \bigcup_{x \in B} T_x A . \qquad (6)$$

Some interesting properties of these erosions and dilations are shown in the next proposition.

Proposition 3. *Let $(G, +)$ be a group. We have*

1. $\forall A \subseteq G,\ \forall B_1, B_2 \subseteq G,\ B_1 \subseteq B_2,$
   (a) $\varepsilon_{B_1}(A) \supseteq \varepsilon_{B_2}(A)$
   (b) $\delta_{B_1}(A) \subseteq \delta_{B_2}(A)$.
2. $\forall A \subseteq G,\ \forall B \subseteq G,$

$$(A \oplus B)^c = A^c \ominus \check{B} \qquad\qquad (A \ominus B)^c = A^c \oplus \check{B} , \qquad (7)$$

where $A^c = \{x \in G : x \notin A\}$ *is the complementary set of A.*

Combining these operators, new morphological operators arise, in particular *opening* and *closing*.

3.2   Opening and Closing

As erosion and dilation, opening and closing have a general definition in the context of complete lattices. A good introduction to these operators is [8].

Definition 2. *Let $(L, \leq)$ be a lattice. Let $\psi : L \to L$ an operator. We say that*

1. ψ *is an* opening *if it is increasing, anti-extensive[3] and idempotent[4];*
2. ψ *is a* closing *if it is increasing, extensive[5] and idempotent (see footnote 4).*

Since they are idempotent and increasing, openings and closings are *morphological filters*. Idempotence is particularly useful since it ensures that several applications of the same operator do not change the result (as do a lot of other filters). This can be thought as a guarantee of filtering all that we want to filter at once.

A particular case of openings and closings are the compositions of erosions and dilations with respect to a structuring element, as shown in the next proposition.

---

[3] $\forall X \in L,\ \psi(X) \leq X$.
[4] $\psi^2 = \psi$.
[5] $\forall X \in L,\ X \leq \psi(X)$.

Proposition 4. *Let $(G, +)$ be a group. Let $B \subseteq G$. Let ε_B and δ_B be respectively the erosion and dilation with respect to the structuring element B. Then,*

1. $\gamma_B \triangleq \delta_B \circ \varepsilon_B$ *is an opening.*
2. $\varphi_B \triangleq \varepsilon_B \circ \delta_B$ *is a closing.*

We use the following notations: $\forall A \subseteq G$,

$$\gamma_B(A) = A \circ B \qquad\qquad \varphi_B(A) = A \bullet B\,.$$

There are many other morphological operators, but in this paper we will only expose a last one: the *hit-or-miss transform*.

3.3   Hit-or-miss Transform

The hit-or-miss transform is defined as follows.

Definition 3. *Let $(G, +)$ be a group. Let $C, D \subseteq G$ with $C \cap D = \emptyset$. Then, the hit-or-miss transform of $A \subseteq G$ with respect to the structuring elements C and D is defined by*

$$A \odot (C, D) = (A \ominus C) \cap (A^c \ominus D). \tag{8}$$

An interesting way of seeing the hit-or-miss transform is given in the next remark and will be the one in which we will think when applying it to Music.

Remark 1. Using the notations from Definition 3,

$$A \odot (C, D) = \{p \in G : T_p C \subseteq A \subseteq T_p D^c\}\,. \tag{9}$$

We may then think the hit-or-miss transform as a pattern matching transformation with an internal condition ($T_p C \subseteq A$) and an external condition ($A \subseteq T_p D^c$); we want to find structures that are bigger than C but smaller than D^c.

4   Applications

Now that we have some Mathematical Morphology operators, let us give some applications. We propose the two following linked steps: first, we consider the lattice formed by the power set of the group $(\mathbb{Z}_{12}, +)$ ordered by inclusion. Then, we extend it with time information to be able to analyze a score. We show that $\mathbb{Z}_{12}$ structuring elements correspond to pitch-class sets, and $\frac{1}{d}\mathbb{Z} \times \mathbb{Z}_{12}$ structuring elements correspond to chord progressions.

4.1   Mathematical Morphology on $\mathbb{Z}_{12}$

Musical chords may be identified with subsets of $\mathbb{Z}_{12}$. We can identify two subsets by the fact that they are both major chords and this is modeled in mathematical terms by the concept of equivalence relation, in this case $\forall A, B \subseteq \mathbb{Z}_{12}, A \sim B \Leftrightarrow \exists p \in G$ such that $T_p A = B$. Mathematical Morphology comes in handy for analyzing this fact as shown in the following.

Erosion as a Structure Detector. Let us focus first in the way we can use erosion to detect chord types; if we want to detect, for instance, major chords, we may erode a score with the major chord structure as illustrated in the following example.

Example 1. Let us consider the following chords:

$$\mathrm{Fmaj} = \{\mathrm{C}, \mathrm{F}, \mathrm{A}\} \qquad \mathrm{Fmaj}^7 = \{\mathrm{C}, \mathrm{E\flat}, \mathrm{F}, \mathrm{A}\} \qquad \mathrm{F}^\circ = \{\mathrm{F}, \mathrm{A\flat}, \mathrm{B}\}\,.$$

We consider the structuring element $\mathrm{maj} = \{\bar{0}, \bar{4}, \bar{7}\}$. Then,

$$\mathrm{Fmaj} \ominus \mathrm{maj} = \{\mathrm{F}\} \qquad \mathrm{Fmaj}^7 \ominus \mathrm{maj} = \{\mathrm{F}\} \qquad \mathrm{F}^\circ \ominus \mathrm{maj} = \emptyset\,.$$

We see that erosion may be used as a detector: it detects the presence of a major chord with root F in Fmaj and Fmaj^7 and shows that there is no major chord in F°. The fact that we got the chroma F after the erosion is because the zero corresponds to the root in the structuring element; it serves as a reference.

An interesting case is when we deal with chords that have non trivial stabilizer[6] in $\mathbb{Z}_{12}$. An example is the diminished seventh chord: if we erode the chord $\mathrm{D}^{\circ 7} = \{\mathrm{D}, \mathrm{F}, \mathrm{A\flat}, \mathrm{B}\}$ by the diminished seventh structuring element $\mathrm{dim7} = \{\bar{0}, \bar{3}, \bar{6}, \bar{9}\}$ it remains unchanged, *i.e.*:

$$\mathrm{D}^{\circ 7} \ominus \mathrm{dim7} = \mathrm{D}^{\circ 7}\,. \tag{10}$$

This points out the fact that the diminished seventh chord may be built on each of its notes or, equivalently, that any of its notes may be considered the root. In Messiaen's terminology, it has a limited number of transpositions equal to the cardinal of the quotient group

$$\mathbb{Z}_{12} \big/ \mathrm{Stab}(\mathrm{D}^{\circ 7}) = \{\mathrm{C} + \mathrm{dim7}, \mathrm{C\sharp} + \mathrm{dim7}, \mathrm{D} + \mathrm{dim7}\}\,, \tag{11}$$

which in this case is 3.

Opening as a Simplification. Let us consider now opening in the same case that in Example 1; we have:

$$\mathrm{Fmaj} \circ \mathrm{maj} = \mathrm{Fmaj} \qquad \mathrm{Fmaj}^7 \circ \mathrm{maj} = \mathrm{Fmaj} \qquad \mathrm{F}^\circ \ominus \mathrm{maj} = \emptyset\,.$$

We see that opening simplifies chords containing a major chord (removing the seventh in the case of Fmaj^7) and removes all the chords that do not contain a major chord. This illustrates why they are called filters (Fig. 2).

---

[6] The stabilizer of a subset $A \subseteq \mathbb{Z}_{12}$ is defined by $\mathrm{Stab}(A) = \{n \in \mathbb{Z}_{12} : T_n A = A\}$ and is a subgroup of $\mathbb{Z}_{12}$.
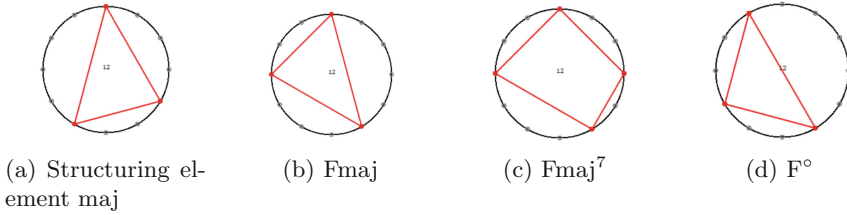
(a) Structuring element maj   (b) Fmaj   (c) Fmaj^7 (d) F$^\circ$

Fig. 2. Chords and structuring element represented in $\mathbb{Z}_{12}$.

Hit-or-miss Transform as a Pattern Detector. Hit-or-miss transform is closely related to erosion; indeed, $\forall A, B \subseteq G$,

$$A \ominus B = A \odot (B, \emptyset). \tag{12}$$

In addition, the second structuring element plays an interesting role: if we want to know to which major and minor scales does the Fmaj chord belong, we may proceed as in the following example.

Example 2. Let us consider the structuring elements

$$\text{majorScale} = \{\overline{0}, \overline{2}, \overline{4}, \overline{5}, \overline{7}, \overline{9}, \overline{11}\} \qquad \text{minorScale} = \{\overline{0}, \overline{2}, \overline{3}, \overline{5}, \overline{7}, \overline{8}, \overline{10}\}.$$

Then, we have

$$\text{Fmaj} \odot (\emptyset, \text{majorScale}^c)) = \{C, F, B\flat\} \quad \text{Fmaj} \odot (\emptyset, \text{minorScale}^c) = \{D, G, A\}.$$

This shows that the F major chord is present in the C, F and B$\flat$ major scales and in the D, G and A minor scales.

It is interesting to note that the major and (natural) minor scales are equivalent in the sense that one is the translation of the other. However, the reference note is not the same and thus gives different results.

Until now, we have used the hit-or-miss transform with one of its structuring elements being the empty set. Let us show an example where we use both constraints for detecting a range of chords that hold both bigger-than and smaller-than constraints.

To detect whether a chord X is a dominant-like chord of some major scale, *i.e.* it contains the dominant third $\text{Dom}^{3M} = \{\overline{7}, \overline{11}\}$ and all its notes are contained in the major scale, the hit-or-miss transform can be used as follows:

$$X \odot (\text{Dom}^{3M}, \text{majorScale}^c). \tag{13}$$

This is illustrated in the following example with three choices of X among the 4096 possible combinations[7].

---

[7] This number comes from the number of subsets of $\mathbb{Z}_{12}$ that is equal to 2^{12}.

Example 3. We consider the following chords:

$$\text{Emaj} = \{\text{E}, \text{G}\sharp, \text{B}\} \quad \text{Dmaj}^{7(5)} = \{\text{C}, \text{D}, \text{F}\sharp\} \quad \text{Amaj}^{7\flat 9} = \{\text{C}\sharp, \text{E}, \text{G}, \text{A}, \text{B}\flat\}.$$

Then, we have

- $\text{Emaj} \odot (\text{Dom}^{3\text{M}}, \text{majorScale}^c) = \{\text{A}\}$,
- $\text{Dmaj}^{7(5)} \odot (\text{Dom}^{3\text{M}}, \text{majorScale}^c) = \{\text{G}\}$,
- $\text{Amaj}^{7\flat 9} \odot (\text{Dom}^{3\text{M}}, \text{majorScale}^c) = \emptyset$,

which shows that Emaj is a dominant-like chord of the A major scale, $\text{Dmaj}^{7(5)}$ is a dominant-like chord of the G major scale and $\text{Amaj}^{7\flat 9}$ is not a dominant-like chord of any major scale.

As a last application of the hit-or-miss transform, we may use it to detect any chord type $C \subseteq \mathbb{Z}_{12}$ by choosing the structuring elements C and C^c.

4.2 Mathematical Morphology on Chroma Rolls

We now go one step further, and illustrate how Mathematical Morphology operators can be useful to analyze scores represented as chroma rolls by using the group $(\frac{1}{d}\mathbb{Z} \times \mathbb{Z}_{12}, +)$ and the lattice $(\mathcal{P}(\frac{1}{d}\mathbb{Z} \times \mathbb{Z}_{12}), \subseteq)$.

Supremum Windowing for Collapsing Harmony. When analyzing harmonically a score, we should make decisions about which notes do we keep and which ones do we discard. Also, we should decide how to handle notes spread in time: usually, we consider that several notes that do not coexist at the same time but are close enough to each other are related because they are part of the same chord, like in the case of arpeggios. To take such configurations into account, we propose a pre-processing step to modify the chroma rolls such that they become more representative of the harmony of the piece.

It is clear that the composer's own style and particularities could lead to different rules. In this paper, as an example, we focus on Chopin's Nocturnes since they are a good illustration of our purpose. The proposed pre-processing method consists in taking only the left hand, since the harmony is concentrated there, and applying what we call a *supremum windowing*.

The supremum windowing is a way of collapsing several notes into the same chord by taking the supremum. Its formula is presented in the next definition.
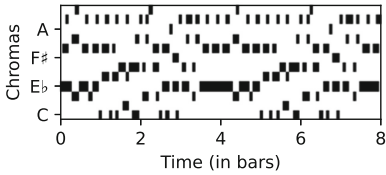
Definition 4. *Let $S \subseteq \frac{1}{d}\mathbb{Z} \times \mathbb{Z}_{12}$. Let $\chi_S \in \{0,1\}^{\frac{1}{d}\mathbb{Z} \times \mathbb{Z}_{12}}$ be the characteristic function of S. The* supremum windowing *of S with window length $L \in \mathbb{Q}$ and hop size $H \in \mathbb{Q}$ is the subset of $\frac{H}{d}\mathbb{Z} \times \mathbb{Z}_{12}$ with characteristic function*

$$W_\infty^{(L,H)}[S] : \frac{H}{d}\mathbb{Z} \times \mathbb{Z}_{12} \to \{0,1\} \qquad\qquad (14)$$
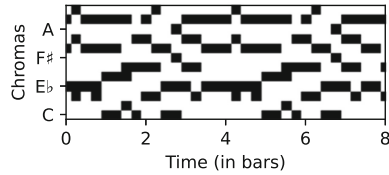$$(t,c) \mapsto \sup\{\chi_S(t+x,c) : x \in [0,L]\}$$

Figure 3 shows the supremum windowing of the left hand of Chopin's Nocturne Op. 9 No. 2.

(a) Original score

(b) Resulting chroma roll

(c) Windowed chroma roll

Fig. 3. Score, chroma roll and its supremum windowed version of the first 8 bars of the left hand of Chopin's Nocturne Op. 9 No. 2 (dynamics, expressions and other features are not considered).

Chord Type Detection. Let us now describe some applications of Mathematical Morphology operators for the detection of chord types in scores. We can extend all that we have done in the one-dimensional case to the two dimensional case; for instance, if we want to detect major chords, we may use an erosion with structuring element $\{0\} \times$ maj. This is done for the (supremum windowed) chroma roll of the Nocturne Op. 9 No. 2 of Chopin in Fig. 4.
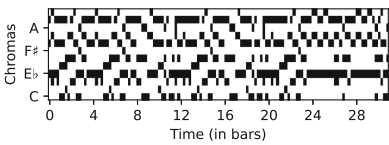
(a) Chroma roll
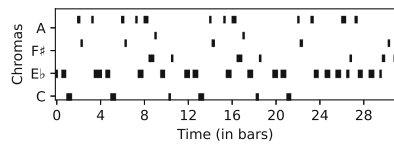
(b) Erosion of the chroma roll

Fig. 4. Chroma roll of the Nocturne Op. 9 No. 2 of Chopin and its erosion by the structuring element $\{0\} \times$ maj.

We may also apply a process analogous to the one in Example 2 to detect tonality and modulations throughout the score. This is illustrated in Fig. 5 in the case of the first 24 bars of the Chopin's Nocturne Op. 48 No. 1 after supremum windowing with length and hop size equal to $\frac{1}{2}$ of the bar. We apply a hit-or-miss transform with $C = \emptyset$ and $D = (\{0\} \times \text{minorHarmScale})^c$, where minorHarmScale $= \{\overline{0}, \overline{2}, \overline{3}, \overline{5}, \overline{7}, \overline{8}, \overline{11}\}$ is the structuring element corresponding

to the harmonic minor scale. We notice several parts that are clearly in C minor and a part around bars 6 to 8 that is in G minor.
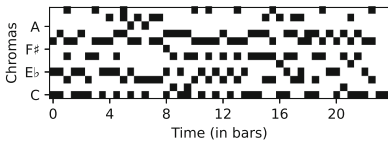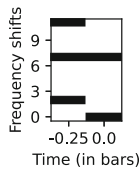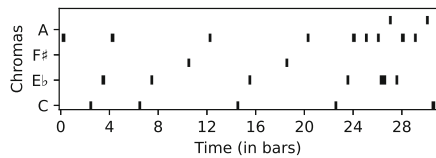

(a) Chroma roll
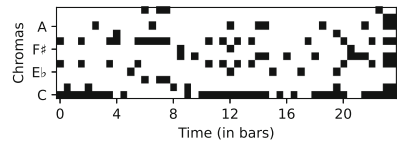
(b) Hit-or-miss transform with structuring elements $C = \emptyset$ and $D = (\{0\} \times \text{minorHarmScale})^c$

Fig. 5. Hit-or-miss transform of the chroma roll of the first 24 bars of the Chopin's Nocturne Op. 48 No. 1 for detecting potential harmonic minor scales.

Chord Progression Detection. Extending structuring elements with a time component allows us to detect chord progressions. For instance, we may detect authentic cadences using the structuring element V-I $= \left(\{-\frac{1}{h}\} \times \{\overline{2}, \overline{7}, \overline{11}\}\right) \cup \left(\{0\} \times \{\overline{0}, \overline{7}\}\right)$, where $\frac{1}{h}$ models the harmonic rhythm. Here, the interval $\{\overline{0}, \overline{7}\}$ guarantees that we recover both the major and minor I degree, and we set the timestamp of the dominant chord $\{\overline{2}, \overline{7}, \overline{11}\}$ to $-\frac{1}{h}$ because we want it to precede the tonic (and thus the minus sign). This way, the result of the erosion provides the time and the chroma corresponding to the I chord. Figure 6 illustrates this for the 24 first bars of the Nocturne Op. 9 No. 2 with $h = 4$.

(a) Structuring element of V-I

(b) Erosion of the chroma roll with structuring element V-I

Fig. 6. Erosion of the chroma roll of Chopin's Nocturne Op. 9 No. 2 by the V-I structuring element.

5   Conclusions

Throughout this paper, we have seen that Mathematical Morphology, based on Group Theory, may be very useful for analyzing Music scores. This framework provides efficient tools for harmonic analysis that are lightweight and well

adapted to harmonic considerations. The use of chroma rolls ensures that we do not have disposition problems in the chords. While simple, the proposed examples illustrate the power of the proposed approach.

It is interesting to notice that we have only used erosion-like operators (erosion, opening and hit-or-miss); this is because we focused on the analysis part. Other operators such as dilations and closing could be used for composition; for instance, a dilation reproduces a particular structure from a reference note.

Several paths can be followed for extending these operators to more complex lattices; for instance, if we consider the amplitude range in a score by using dynamics, morphological operators acting on functions can be leveraged. Also, we can change the base groups; for example if we want to analyze MIDI files, we may use $(\mathbb{Z}, +)$, and we may even apply these tools on signal operators such as the Short-Time Fourier Transform. In future works, we will also go deeper in the creation of an automatic analyzer that implements morphological tools based on the equivalence between chord structure and structuring element. We intend to provide a deterministic framework for the study of harmony at large scale.

References

1. Bilmes, J.A.: Timing is of the essence: perceptual and computational techniques for representing, learning, and reproducing expressive timing in percussive rhythm. Ph.D. Thesis, Massachusetts Institute of Technology (1993)
2. Heijmans, H.J.A.M., Ronse, C.: The algebraic basis of mathematical morphology I. Dilations and erosions. Comput. Vis. Graph. Image Process. **50**(3), 245–295 (1990)
3. Karvonen, M.: Using Mathematical Morphology for Geometric Music Retrieval. Ph.D. Thesis, University of Helsinky (2008)
4. Karvonen, M., Laitinen, M., Lemström, K., Vikman, J.: Error-tolerant content-based music-retrieval with mathematical morphology. In: Ystad, S., Aramaki, M., Kronland-Martinet, R., Jensen, K. (eds.) Exploring Music Contents, CMMR 2010. LNCS, vol. 6684, pp. 321–337. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23126-1_20
5. Lascabettes, P., Bloch, I., Agon, C.: Analyse de représentations spatiales de la musique par des opérateurs simples de morphologie mathématique. In: Journées d'Informatique Musicale. Strasbourg, France (2020)
6. Matheron, G.: Random Sets and Integral Geometry. Wiley, New-York (1975)
7. Romero-García, G.: Rhythm Transcription and Characterization for Performed Music and Arrhythmia Sequences. Master's Thesis, Sorbonne Université, Paris (2020)
8. Ronse, C., Heijmans, H.J.A.M.: The algebraic basis of mathematical morphology: II. Openings and closings. CVGIP: Image Underst. **54**(1), 74–97 (1991)
9. Serra, J.: Image Analysis and Mathematical Morphology. Academic Press, New-York (1982)

Computational Analysis of Musical Structures Based on Morphological Filters

Paul Lascabettes[1]([✉]), Carlos Agon[1], Moreno Andreatta[1,2], and Isabelle Bloch[3]

[1] STMS-UMR9912, Ircam, Sorbonne Université, CNRS, Ministère de la Culture, Paris, France
{lascabettes,agon,andreatta}@ircam.fr
[2] CNRS-IRMA, ITI CREAA, Université de Strasbourg, Strasbourg, France
andreatta@math.unistra.fr
[3] Sorbonne Université, CNRS, LIP6, Paris, France
isabelle.bloch@sorbonne-universite.fr

Abstract. This paper deals with the computational analysis of musical structures by focusing on the use of morphological filters. We first propose to generalize the notion of melodic contour to a chord sequence with the chord contour, representing some formal intervallic relations between two given chords. By defining a semi-metric, we compute the self-distance matrix of a chord contour sequence. This method allows generating a self-distance matrix for symbolic music representations. Self-distance matrices are used in the analysis of musical structures because blocks around the diagonal provide structural information on a musical piece. The main contribution of this paper comes from the analysis of these matrices based on mathematical morphology. Morphological filters are used to homogenize and detect regions in the self-distance matrices. Specifically, the opening operation has been successfully applied to reveal the blocks around the diagonal because it removes small details such as high local values and reduces all blocks around the diagonal to a zero value. Moreover, by varying the size of the morphological filter, it is possible to detect musical structures at different scales. A large opening filter identifies the main global parts of the piece, while a smaller one finds shorter musical sections. We discuss some examples that demonstrate the usefulness of this approach to detect the structures of a musical piece and its novelty within the field of symbolic music information research.

Keywords: Symbolic music information research · Music structure · Chord contour · Self-distance matrix · Mathematical morphology

1 Introduction

Mathematical morphology is an algebraic theory that analyzes shapes and is mostly used in image analysis and understanding. However, this theory is not

---

very common yet in the Mathematics and Music community. The fundamental idea of this theory is to modify the shape, the size or the topological properties of objects with non-linear and non-reversible transformations. Among the few existing applications of mathematical morphology to symbolic representations of music, automatic methods have been developed in Music Information Research community (MIR) to detect approximate occurrences of musical patterns in symbolic music databases [12,13]. In this case, mathematical morphology enables to match almost identical patterns. Moreover, mathematical morphology has also been used to analyze concept lattices based on musical intervals [2], and basic operators of mathematical morphology have been adapted to find a musical meaning, allowing for example extracting harmonic components or to obtain musical transformations [14].

The main contribution of this paper is to propose a novel method, based on mathematical morphology, to extract hierarchical musical structures from the self-distance matrix. This method can be applied to any type of similarity matrix and to any type of data. In our case, the self-distance matrix is computed from symbolic music representations, using a generalization of melodic contour to chord sequences. The purpose of this method is to homogenize the different regions of the self-distance matrix in order to identify the musical structures. Two basic morphological operations, the erosion and dilation, have already been successfully used to detect the repeating patterns longer than a minimum length into a time-lag matrix (a similar representation as the self-distance matrix) [15]. However, rather than identifying segments as in [15], we demonstrate the usefulness of the morphological opening operation in order to identify blocks in the self-distance matrix. This operation eliminates small details, while flatter and homogeneous regions are obtained. In addition, it reduces all the blocks around the diagonal, which correspond to musical sections, to a zero value. We discuss the form to choose when applying an opening filter to extract information from the self-distance matrix: a constant square-shaped filter. But the size can also be adjusted to detect different musical structures. A large opening will identify the global part of the piece while a smaller one will reveal shorter sections. This idea is illustrated by detecting different musical structures in Mozart's Piano Sonata *Alla Turca*.

This paper details the above ideas and is organized as follows. Section 2 proposes a method to generate a self-distance matrix from symbolic music representations. We introduce the concept of chord contour (Sect. 2.1) as a generalization of the usual melodic contour, and then define a distance to compute the self-distance matrix of a chord contour sequence (Sect. 2.2). Section 3 describes how to use morphological filters in order to extract musical structures from the self-distance matrix. After providing a short introduction to mathematical morphology (Sect. 3.1) we then demonstrate the relevant applications of the opening operation in order to identify the main blocks of the self-distance matrix (Sect. 3.2). Finally, in Sect. 4, we apply our proposed method and illustrate how morphological filters can be used to extract musical structures at multiple levels of granularity.

2   Creating a Self-distance Matrix from Symbolic Music

2.1   Converting Symbolic Music to Sequence Using Chord Contour

Whether for music perception [6,23], music analysis [1] or music theory [5], *melodic contour* has become a fundamental tool in the MIR community. This tool applies on monophonic structures, i.e., musical phrases or motives in which two notes never sound at once. It is defined by the set of the directions between consecutive pitches of a melody, $+1$ and -1 indicating respectively an ascending and a descending interval. Figure 1a illustrates this idea by representing each note of a melody by a circle in a time/pitch graph. Melodic contour summarizes intervallic information and can be used to compare and classify melodic patterns or to help understand their perception. Considering the importance of melodic contour, it is not surprising that multiple extensions have been proposed. For example, two other contours were defined in [3]: the *strong contour* (melodic contour of only the notes present on the beat) and the *weak contour* (strong contour with extra information if there is a contour variation within the beat). Moreover, it was proposed in [16,20] to observe the directions at longer range, i.e., all the directions between the ith and j^{th} pitches, not only between the ith and $(i+1)^{th}$ pitches as for the usual melodic contour. To this purpose, both works used a matrix representation: *Morris's comparison matrix (COM-matrix)* in [16], and *combinatorial contour matrix* in [20]. In the COM-matrix, the coefficient at position (i,j) is the pitch direction between notes i and j, and for the combinatorial contour matrix this coefficient is $+1$ if the jth note is higher in pitch than the ith note or 0 otherwise. However, these generalizations remain in the monophonic context, and they do not handle musical chords.

We propose a generalization of the melodic contour to chord sequences, i.e., not restricted to note sequences. In the proposed definition, the direction between the pitches of two given chords is no longer a number but a matrix, called *chord contour*. The coefficient (i,j) of the chord contour is the direction between the
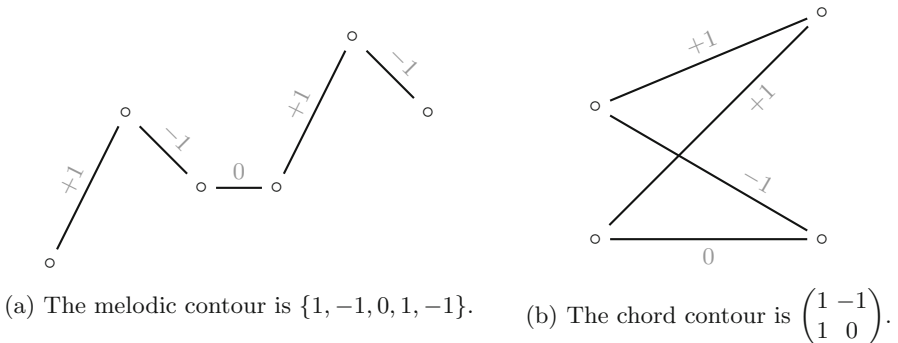

(a) The melodic contour is $\{1, -1, 0, 1, -1\}$.   (b) The chord contour is $\begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$.

Fig. 1. Illustration of the melodic contour and the chord contour.

i^{th} note of the first chord and the jth note of the second chord, where the notes of the chords are ordered in descending pitch order. Therefore, the chord contour from an n-note chord to an m-note chord is of size $n \times m$. Figure 1b illustrates the construction of the chord contour: in this example, the two chords have two notes, so the corresponding chord contour is a 2×2 matrix. The first row corresponds to the directions from the highest note of the first chord to the notes of the next chord, and so on. The chord contour sequence of the introduction of Edvard Grieg's *March of the Dwarfs* is graphically represented in Fig. 2. It will be analyzed in the next sections in order to find the main passages or blocks of this sequence.
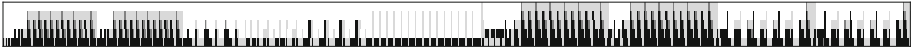
Fig. 2. Representation of the chord contour sequence of the introduction of *March of the Dwarfs*. Black, dark gray and light gray pixels map respectively to values of 1, 0 and −1.

2.2   Distance Matrix of a Chord Contour Sequence

In this section, we propose to define a distance between two chord contours. The main difficulty comes from the fact that chord contours, which are matrices, can have different sizes. First, we consider two chord contours with the same size. In this case the *Hamming distance* will be used. Let $A = (a_{i,j})$ and $B = (b_{i,j})$ be two chord contours of sizes $n \times m$, the Hamming distance $d(A, B)$ between the matrices A and B is defined as the number of coefficients which differ:

$$d(A, B) = |\{(i, j) \in [1...n] \times [1...m] \mid a_{i,j} \neq b_{i,j}\}|. \qquad (1)$$

If one of the two matrices has more rows (or columns) than the other matrix, one can reduce it by deleting rows (or columns) in order to get two matrices of the same size and use the previous formula, with the addition of the number of deleted rows (or columns). The rows (or columns) to be deleted are those that minimize the distance between the two matrices. Deleting a row (respectively a column) corresponds to omitting a note in the first chord (respectively the second chord). Thus, if A and B are two matrices of size $n_1 \times m_1$ and $n_2 \times m_2$, the distance $\mathcal{D}(A, B)$ between these two matrices is defined as:

$$\mathcal{D}(A, B) = \min_{A', B'} (d(A', B')) + |n_1 - n_2| + |m_1 - m_2|, \qquad (2)$$

where A' and B' are two matrices of size $\min(n_1, n_2) \times \min(m_1, m_2)$ such that A' (respectively B') is obtained by removing $n_1 - \min(n_1, n_2)$ rows and $m_1 - \min(m_1, m_2)$ columns from A (respectively B). From a mathematical point of view, the first distance d respects symmetry, identity of indiscernibles, non-negativity and triangular inequality. It is well defined as a metric on the space of

matrices with the same size. On the other hand, for the second distance $\mathcal{D}$, the triangular inequality is lost; hence, it is only a semi-metric in the mathematical sense. However, since we only make pairwise comparisons, without looking for a path from one matrix to another one, the triangular inequality is not essential.

In order to visualize the musical structures, the *self-similarity matrix* was proposed in [7], as a two-dimensional representation defined by computing the similarity between any two instants. As stated in [19], self-similarity matrices have become a major concept in the study of musical structures. In addition, the dual of self-similarity matrices are *self-distance matrices* where each coefficient describes the distance between two elements. Here we will focus on self-distances matrices, but the same logic can be transcribed on self-similarity matrices. Let c_k be the kth chord contour of the musical piece, i.e., from the kth chord to the $(k+1)^{th}$ chord. Then the coefficient of the line i and the column j of the self-distance matrix is defined by $\mathcal{D}(c_i, c_j)$. Figure 3 displays the self-distance matrix corresponding to the example of the introduction of *March of the Dwarfs* in Fig. 2. Since $\mathcal{D}$ is symmetric, the self-distance matrix is a symmetric matrix. The musical structures can be inferred from the information near the diagonal: the different blocks around the diagonal framed in red in Fig. 3 represent the musical sections. It is possible to understand the shape of the self-distance matrix in comparison to the chord contour sequence: blocks on the diagonal correspond to sections that are visually identifiable in Fig. 2.
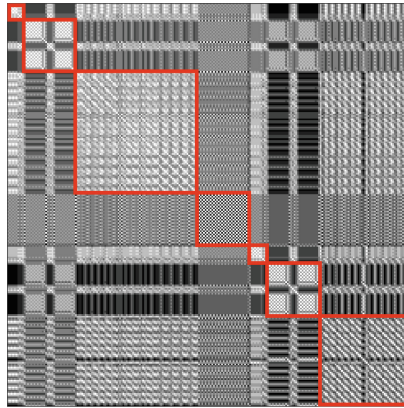

Fig. 3. Self-distance matrix of the introduction of *March of the Dwarfs* (white = 0, i.e. low distance and high similarity, black = high distance values and low similarity). (Color figure online)

3   Analysis of the Self-distance Matrix Using Morphological Operations and Filters

In this section, we propose an original method, based on mathematical morphology, to extract meaningful information from the self-distance matrix. This method can be applied to any type of self-distance or self-similarity matrix and to any type of data. The purpose of this method is to homogenize the different regions of the matrix in order to identify the musical structure.

3.1   A Short Introduction to Morphological Filters

Developed in the 1960s s by G. Matheron and J. Serra, *Mathematical Morphology* is, in its deterministic component, an algebraic theory developed initially to analyze shapes, and is widely used in image analysis. In this paper, we will rely on mathematical morphology defined on functions, typically used to analyze gray level images, making an analogy between self-distance matrices and images. Only the useful notions are recalled here, and more details can be found in [4,10,11, 18,21,22]. Let $(\mathcal{F}, \leq)$ be a lattice of functions (here we consider functions from $E = \mathbb{Z}^n$ into $\mathbb{R}^+$ to handle self-distance matrices, and the lattice is complete). A *dilation* is an operation that commutes with the supremum of the lattice, and an *erosion* an operation that commutes with the infimum. Concrete forms of these operations, which are often used, rely on the notion of *structuring element*, an element B of the lattice, which can be considered as a binary relation between elements of the underlying space E, or as a spatial neighborhood in our analogy with images, or more generally as a function with bounded support. *Dilation* $\oplus$ and *erosion* $\ominus$ in the complete lattice $(\mathcal{F}, \leq)$ are extensions of *Minkowski addition* [17] and *subtraction* [9] in the binary morphological case, and are defined for any $X \in \mathcal{F}$, any structuring element $B \in \mathcal{F}$ and any $x \in E$:

$$X \oplus B(x) = \sup_{t \in E} \left(X(t) + B(x - t) \right), \qquad X \ominus B(x) = \inf_{t \in E} \left(X(t) - B(t - x) \right). \quad (3)$$

Dilation extends bright zones and reduces dark ones, while erosion does the opposite. The other two fundamental operations result from the composition of these operators. Indeed, the *opening* ∘ is the composition of an erosion and a dilation and the *closing* • is a dilation followed by an erosion:

$$X \circ B = (X \ominus B) \oplus B, \qquad X \bullet B = (X \oplus B) \ominus B. \quad (4)$$

Opening and closing are increasing and idempotent operators, hence morphological filters. They can be used to eliminate small details (having higher values than their surrounding using opening, and smaller ones using closing) according to the size and shape of the structuring element. Therefore, by using these filters, some detailed information may be lost, while more flat and homogeneous regions are obtained. This property will be used to highlight homogeneous regions in the self-distance matrix, in order to exhibit the main musical structures, as detailed in the next section.

3.2   Application of Mathematical Morphology to the Self-distance Matrix

We propose to use morphological filters to identify the main blocks of the self-distance matrix. Blocks along the diagonal provide information on the musical structure of the piece since low distance values of the self-distance matrix correspond to passages with high similarity. In order to identify larger similar blocks, locally higher distance values should be removed. The opening operation is particularly well suited to this situation. To do this, the structuring element has to be constant and square-shaped in order to preserve the general organization of the matrix, which exhibits strong vertical and horizontal structures, as well as squared blocks. By using this operation, it is possible to homogenize the regions of the self-distance matrix and to reduce the blocks on the diagonal to a zero value (because the diagonal coefficients are equal to zero due to the identity of indiscernibles of the metric).
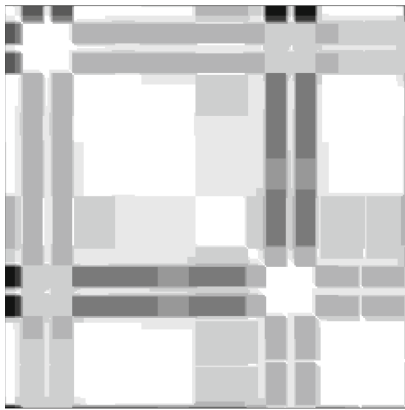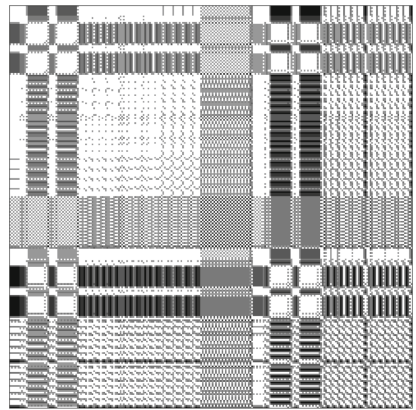
(a) Opening   (b) Threshold

Fig. 4. Filtering of the self-distance matrix using a morphological opening (a). As a comparison, a simple thresholding is shown in (b).

The result of this operation on the self-distance matrix of Fig. 3 with a square structuring element of size 12×12 is represented in Fig. 4a. Blocks on the diagonal appear in white, which is the minimal value (equal to zero), and we can easily detect them. To compare this method with simpler methods, thresholding is shown in Fig. 4b. Here, each coefficient below half of the maximum coefficient of the matrix is set to zero. However, this method does not detect the main blocks of the self-distance matrix. The threshold operation acts globally on the matrix, with the same threshold value applied everywhere. By contrast, opening is an operator that acts locally on the coefficients of the matrix, depending on local shape and size of the distance function, not on absolute values, which fits our filtering objective better.

The *self-similarity matrix*, introduced in [7], is also used in audio-based approaches to the analysis of musical structures. In this case, the values of the self-similarity matrix are inverted with respect to the self-distance matrix. The diagonal coefficients are the highest values (equal to one) and the goal is to remove locally lower values and to reduce the blocks around the diagonal to the highest value of the matrix. This change can also be handled with the morphological tools because dilation and erosion (respectively opening and closing) form pairs of dual operators [4]. This means concretely that applying a dilation (respectively an opening) on a self-distance matrix is equivalent to applying an erosion (respectively a closing) on a self-similarity matrix, and vice versa.

4   Changing the Shape of the Morphological Filter to Detect Different Musical Structures

The morphological operations provide new computational tools for the analysis and identification of the overall structure of a musical piece. Moreover, it is possible to detect musical structures at different scales, for example to refine the granularity of the analysis and identify the bars of the piece. This can be done by changing the size of the structuring element, in order to detect blocks of different sizes. With a smaller structuring element, it is possible to detect smaller blocks around the diagonal, representing for instance the bars of the piece, while a larger one will allow detecting the global musical structure at a bigger scale.

To illustrate the notion of filtering with different structuring elements, we consider the third movement of the Piano Sonata No.11 in A Major, composed by Wolfgang Amadeus Mozart and commonly known as *Alla Turca* or *Turkish Rondo*. The structures of the piece are represented in Fig. 5a, where each letter symbolizes 8 bars. This piece is divided into four main parts represented by red rectangles and linked with blue rectangles. There are two levels of structure: the 7 colored rectangles (global structure) or the 28 letters (detailed structure). As seen previously, the structuring element has to be constant and square-shaped, the only parameter to choose being the size. We applied an opening filter with a constant square-shaped structuring element of size 3×3 and 6×6 to the self-distance matrix (computed using the chord contour sequence). The result of these opening filters is displayed in Figs. 5b and 5c. For a clearer understanding, only the diagonal blocks (detected with the flood-fill algorithm) are shown in black in this figure, i.e., zero value coefficients connected to the diagonal of the matrix. We computed the *novelty score*, introduced in [8], of these two opening diagonals. The novelty score N is the correlation along the diagonal of a matrix M with the checkerboard kernel C:
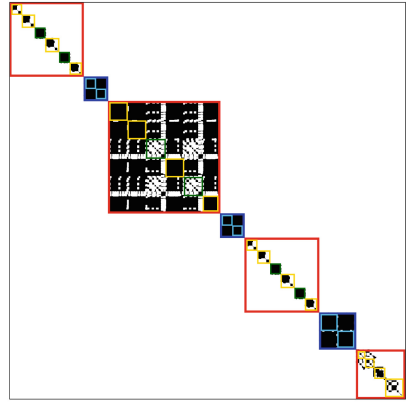
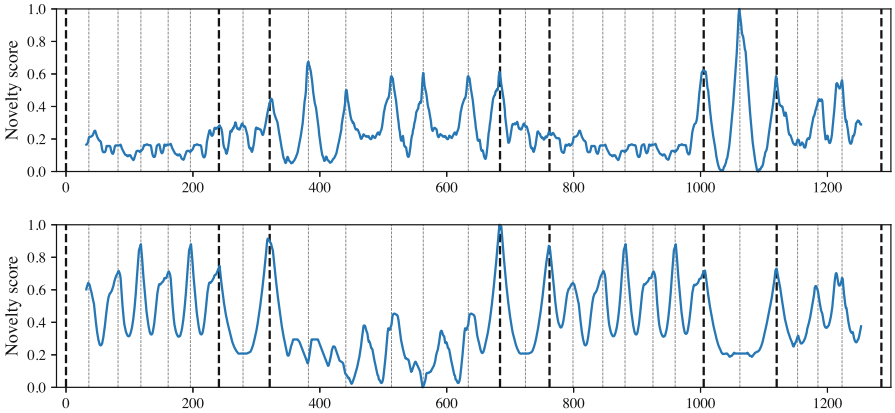$$N(t) = \sum_{i=-L/2}^{L/2} \sum_{j=-L/2}^{L/2} C(i,j)M(i+t, j+t), \tag{5}$$

(a) Musical structures of *Alla Turca* (W.A. Mozart).
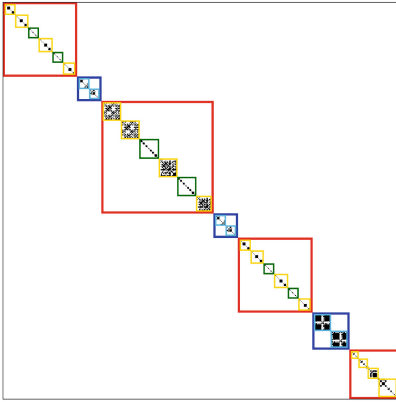
(b) Opening diagonal with a 3×3 constant square-shaped structuring element.

(c) Opening diagonal with a 6×6 constant square-shaped structuring element.

(d) Novelty score of the opening diagonal with a constant square-shaped 3×3 (top) and 6×6 (bottom) structuring element.

Fig. 5. Filtering of the self-distance matrix at different scales by the opening operation in order to obtain different musical structures. (Color figure online)

where C is the 64×64 symmetric matrix defined as

$$
C = \begin{pmatrix}
-1 & \cdots & -1 & 1 & \cdots & 1 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
-1 & \cdots & -1 & 1 & \cdots & 1 \\
1 & \cdots & 1 & -1 & \cdots & -1 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & \cdots & 1 & -1 & \cdots & -1
\end{pmatrix} .
\tag{6}
$$

Notice that we use the opposite of the checkerboard kernel presented in [8] because we have a self-distance matrix instead of a self-similarity matrix. This novelty score allows detecting changes, and therefore the limits of the blocks to identify. The novelty scores of the two opening diagonals are represented in Fig. 5d. We also add the boundaries of the musical structures shown in Fig. 5a with thick dotted lines (boundaries between rectangles) and thin dotted lines (boundaries between letters). The high value of the novelty score represents the boundaries of the piece. The novelty score of the opening diagonal with a 3×3 structuring element detects the boundaries between the D/D/E/D'/E/D' and C'/C' sections. While the novelty score of the opening diagonal with a 6×6 structuring element detects the boundaries between the rectangles and the A/A/B/A'/B/A' sections. With these two diagonals blocks, it is possible to detect two different structures of the piece.

Finally, we can adjust the size of the structuring element used to filter the self-distance matrix depending on the granularity that we want in the analysis of the musical structures (which enables for example to detect only few very long passages or a greater number of short passages). By varying the size of the structuring element, we can computationally grasp the segmentation process at multiple levels. In fact, every time we increase the size of the structuring element we force some segments to merge and become a new bigger segment, starting from few notes segments to the whole piece.

5   Conclusions

This article proposed a new method to visualize a piece of music and analyze its structures in an automatic way. By focusing on the pitch variations between the elements of a melodic line (notes) or harmonic progression (chords), we have proposed an original approach that generalizes the notion of melodic contour to a sequence of chords, called *chord contour*. In such a sequence, the pitch variation is described by matrices, instead of just a number as in the case of the traditional melodic contour. These matrices characterize a sequence of chords by using the direction of the pitch variation of the notes from one chord to the next one. We then introduced a proximity measure between two chord contours of any sizes with the semi-metric $\mathcal{D}$ in order to compute the self-distance matrix of a chord contour sequence. The self-distance matrix is used to analyze musical structures by leveraging the fact that the principal blocks around the diagonal

correspond to the main passages of the piece. The main idea of this paper is to use filters borrowed from the domain of mathematical morphology in order to identify these blocks. Mathematical morphology is a relatively new theory in the MIR community and we are convinced that it can be particularly useful in connection with the self-distance matrix. These morphological filters have been used to homogenize and identify well-defined regions of the self-distance matrix corresponding to musical entities. The opening operation has been successfully applied to the analysis of the musical structures of a piece because it locally removes the high values. With a constant square-shaped structuring element, it reveals the horizontal and vertical blocks of the self-distance matrix. In addition, the blocks around the diagonal, which correspond to a well-defined musical structure, all have a zero value. Moreover, by varying the size of the filter, it is possible to have different filtering levels in the automatic detection of the underlying structures of the musical piece. By filtering the self-distance matrix with a large opening, one is able to identify the main global parts of the piece, while using a smaller morphological filter reveals shorter musical sections. Some promising results of applying this new method in the field of music automatic segmentation have been obtained and discussed by presenting a computational analysis of an excerpt of Edvard Grieg's *March of Dwarfs* and of Mozart's Piano Sonata *Alla Turca*.

In this paper, we demonstrated the usefulness of morphological filters to homogenize musical sections to detect the musical structure. However, homogeneity is not the only criteria for music structure analysis, and the other main approach is based on repetition. Paulus et al. argue that a combined approach (based on homogeneity, novelty and repetition) provides promising results [19]. Our method does not handle repetition, because the goal of this paper is to show the application of mathematical morphology for music structures analysis. Due to the simplicity yet powerful utility of morphological filters, we strongly believe that this method can be reuse for future algorithms for the homogeneity step. Moreover, although we have applied this method on symbolic music representations with a chord contour sequence, this method can also be applied for audio-based analysis of musical structures. For future research, we plan to test this method on a large audio database with annotated structures in a hierarchical way to validate it experimentally.

References

1. Adams, C.R.: Melodic contour typology. Ethnomusicology **20**(2), 179–215 (1976)
2. Agon, C., Andreatta, M., Atif, J., Bloch, I., Mascarade, P.: Musical descriptions based on formal concept analysis and mathematical morphology. In: Chapman, P., Endres, D., Pernelle, N. (eds.) ICCS 2018. LNCS (LNAI), vol. 10872, pp. 105–119. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91379-7_9
3. Anagnostopoulou, C., Giraud, M., Poulakis, N.: Melodic contour representations in the analysis of children's songs. In: van Kranenburg, P., Anagnostopoulou, C., Volk, A. (eds.) 3rd International Workshop on Folk Music Analysis, pp. 40–43. Amsterdam, Netherlands (2013)

4. Bloch, I., Heijmans, H., Ronse, C.: Mathematical morphology. In: Aiello, M., Pratt-Hartmann, I., Van Benthem, J. (eds.) Handbook of Spatial Logics, pp. 857–944. Springer, Dordrecht (2007). https://doi.org/10.1007/978-1-4020-5587-4_14
5. Buteau, C., Mazzola, G.: Motivic analysis according to Rudolph Réti: formalization by a topological model. J. Math. Music **2**(3), 117–134 (2008)
6. Dowling, W.J.: Melodic contour in hearing and remembering melodies. In: Aiello, R., Sloboda, J.A. (eds.) Musical Perceptions, pp. 173–190. Oxford University Press, Oxford (1994)
7. Foote, J.: Visualizing music and audio using self-similarity. In: Seventh ACM International Conference on Multimedia (Part 1), pp. 77–80 (1999)
8. Foote, J.: Automatic audio segmentation using a measure of audio novelty. In: IEEE International Conference on Multimedia and Expo, ICME 2000. Latest Advances in the Fast Changing World of Multimedia, vol. 1, pp. 452–455 (2000)
9. Hadwiger, H.: Minkowskische Addition und Subtraktion beliebiger Punkt-mengen und die Theoreme von Erhard Schmidt. Mathematische Zeitschrift **53**, 210–218 (1950)
10. Heijmans, H.: Morphological Image Operators. Advances in Electronics and Electron Physics: Supplements. Academic Press (1994)
11. Heijmans, H., Ronse, C.: The algebraic basis of mathematical morphology I. dilations and erosions. Comput. Vis. Graph. Image Process. **50**(3), 245–295 (1990)
12. Karvonen, M., Laitinen, M., Lemström, K., Vikman, J.: Error-tolerant content-based music-retrieval with mathematical morphology. In: Ystad, S., Aramaki, M., Kronland-Martinet, R., Jensen, K. (eds.) CMMR 2010. LNCS, vol. 6684, pp. 321–337. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23126-1_20
13. Karvonen, M., Lemström, K.: Using mathematical morphology for geometric music information retrieval. In: International Workshop on Machine Learning and Music, MML 2008, Helsinki, Finland (2008)
14. Lascabettes, P., Bloch, I., Agon, C.: Analyse de représentations spatiales de la musique par des opérateurs simples de morphologie mathématique. Journées d'Informatique Musicale (2020)
15. Lu, L., Wang, M., Zhang, H.J.: Repeating pattern discovery and structure analysis from acoustic music data. In: 6th ACM SIGMM International Workshop on Multimedia Information Retrieval, pp. 275–282 (2004)
16. Marvin, E.W., Laprade, P.A.: Relating musical contours: extensions of a theory for contour. J. Music Theory **31**(2), 225–267 (1987)
17. Minkowski, H.: Volumen und Oberfläche. Mathematische Annalen **57**, 447–495 (1903)
18. Najman, L., Talbot, H.: Mathematical Morphology: From Theory to Applications. ISTE-Wiley (2010)
19. Paulus, J., Müller, M., Klapuri, A.: State of the art report: audio-based music structure analysis. In: ISMIR, Utrecht, pp. 625–636 (2010)
20. Quinn, I.: The combinatorial model of pitch contour. Music Percept.: Interdiscip. J. **16**(4), 439–456 (1999)
21. Ronse, C., Heijmans, H.: The algebraic basis of mathematical morphology: II. openings and closings. CVGIP: Image Underst. **54**(1), 74–97 (1991)
22. Serra, J.: Image Analysis and Mathematical Morphology. Academic Press Inc., London (1982)
23. Trehub, S.E., Bull, D., Thorpe, L.A.: Infants' perception of melodies: the role of melodic contour. Child Dev. **55**(3), 821–830 (1984)

Non-spectral Transposition-Invariant Information in Pitch-Class Sets and Distributions

Jason Yust[1(✉)] and Emmanuel Amiot[2]

[1] Boston University, Boston, MA 02215, USA
`jyust@bu.edu`
[2] Université de Perpignan, Perpignan, France
`manu.amiot@free.fr`

Abstract. The spectral information of a pitch-class set or distribution relates to its interval content and what Ian Quinn calls its harmonic qualities, the magnitudes of a discrete Fourier transform of a pitch-class vector. The spectrum is invariant with respect to transposition and inversion, but the existence of Z-related sets, which have equivalent spectra but are not related by transposition or inversion, means that the spectrum is not a complete description of a set class. We show how to isolate transposition-invariant phase information using products of Fourier coefficients. We describe some of the mathematical features of these coefficient products and show how they encode aspects of tonality, and can be useful for analyzing non-tonal music with an example from Takemitsu's "Air" for solo flute.

1 Pitch-Class Set Theory and Homometry

Allen Forte [6] originally defined set-class equivalence as equivalence of interval vectors, but subsequently reconsidered, using transpositional and inversional equivalence instead [7]. Forte's original definition is known in mathematics as *homometry*, and, as Amiot [3] has shown, can also be defined as *equivalence of spectra*. The spectrum is obtained by taking the characteristic function of a pitch class set and considering just the sizes of the coefficients of its discrete Fourier transform (DFT). Ian Quinn [8] refers to the spectrum as a point in *quality space*. Transpositions and inversions are homometric, but not vice versa. Therefore Forte's original definition of set class was stronger than his later one. The difference between them consists of what he calls "Z-related" sets, sets that are homometric but not related by transposition or inversion. With the exception of hexachords, the Z-relation is somewhat rare for ordinary pitch-class sets, but we can identify many more examples if we consider pitch-class multisets [9] or real-valued characteristic functions, in which case the set of all distributions homometric to a given one is a multi-dimensional torus, the orbit of the so-called *spectral units* group [3, chapter 4].

While the spectrum therefore provides much of the important information about a pitch-class set, it is not a complete description. Since the DFT is a lossless transformation, that means that there is transposition-invariant information in the *phases* of the DFT coefficients. In the first section we show that special *coefficient products* (specifically with coefficients whose indices sum to twelve) are transposition invariant, and therefore the phases of these include the desired non-spectral information. We then show the importance of these non-spectral set class properties for characterizing tonal sets, and offer an example analysis in a non-tonal context.

2   Products of DFT Coefficients

2.1   Definitions

Recall that any complex number z can be described by its magnitude $|z| \in \mathbf{R}_+$ and its phase $\arg(z) \in \mathbf{R}/2\pi\mathbf{Z}$:

$$\mathbf{C} \ni z = |z|e^{i\arg(z)}.$$

As mentioned in the preamble, it can be shown that homometry is exactly the equality of all Fourier coefficient magnitudes; Since these are invariant under transposition and inversion, it remains to consider the phases for non-homometry related information. Indeed, phase increases by a constant quantity under transposition and changes signum under inversion.

In the following, we normalize phase modulo 12 (or more generally, n, the cardinality of the chromatic aggregate) by setting

$$\varphi_k = \arg(\hat{a}_k) \qquad \Phi_k = \frac{12}{2\pi}\varphi_k = \frac{6}{\pi}\arg(\hat{a}_k)$$

where $\hat{a}_k = \sum_{x \in X} e^{-2i\pi kx/12}$ is the kth Fourier coefficient of pitch-class set X.

It was noticed in [14] that many pitch-class sets in tonal context satisfy an improbable equation:

$$\Phi_5 \approx \Phi_3 + \Phi_2. \qquad (\sharp)$$

This is an exact equality for diatonic scales, fifths, and several other prominent tonal collections. Notice however that the opposite equality ($\Phi_5 \approx -\Phi_3 - \Phi_2$) yields for a pentatonic scale (thus allowing a way, Fourier-wise, to tell pentatonic and diatonic scales apart, although the magnitude of all their Fourier coefficients except the 0*th* are equal). It yields, up to a small error, for major and minor triads.

This is an intriguing feature, since Φ, a complex logarithm, is anything but a linear map; also a comprehensive computation shows that for most pc-sets, equation ($\sharp$) is quite incorrect.[1]

---

[1]   For 80% of pc-sets, the error is larger than 10%.

For the sequel of this paper, we will rephrase it: since $\Phi_5 = -\Phi_7$ (a general feature of Fourier coefficients of characteristic, or in general real-valued, functions) we can state instead

$$\Phi_7 + \Phi_3 + \Phi_2 \approx 0 \pmod{12}$$

or by exponentiation

$$\exp\left(i(\Phi_7 + \Phi_3 + \Phi_2)\pi/6\right) = e^{i\Phi_7\pi/6}e^{i\Phi_3\pi/6}e^{i\Phi_2\pi/6} \approx e^0 = 1$$

or even better, multiplying by the magnitudes of the relevant Fourier coefficients to rebuild them anew, $\hat{a}_2\hat{a}_3\hat{a}_7 \approx |\hat{a}_2\hat{a}_3\hat{a}_7|$, meaning

$$\boxed{\hat{a}_2\hat{a}_3\hat{a}_7 \text{ is (almost) real positive.}}$$

We can then state a general definition, where $2 + 3 + 7 = 12$ is replaced by an integer partition:

Definition 1. *Let n be the cardinality of the chromatic aggregate and $k_1, k_2 \ldots k_r$ be an integer partition of n, i.e. the k_i are positive integers.*[2] *Then $\hat{a}_{k_1}\hat{a}_{k_2}\ldots\hat{a}_{k_r}$ is a (regular) coefficient product.*[3]
It is coherent *for a given pitch-class set if its value for that set is real positive, approximately coherent if it is close to real positive. For short, if the context is clear we will say that a pc-set is coherent if its (regular) product is coherent.*
As will be seen later on, the more general case of real-valued regular products (positive or negative) is notable. We will call such a product aligned.

Without spoilers, with partition $2 + 3 + 7 = 12$ we have coherent products for all single notes, dyads, diatonic scales, major sevenths; approximate coherence for major or minor triads; and aligned products for pentatonic scales.
A trivial but illuminating example of coherent product, for which we thank an anonymous reviewer, is any partition of the type $n = k + (n - k)$, since for any pc-set we get $a_k a_{n-k} = |a_k|^2 \geq 0$. In this case, Proposition 3 retrieves that transposed or inverted pc-sets are homometric.

2.2 Features

Proposition 1. *Singletons are coherent for all regular products.*

Proof. Let $a \in \mathbf{Z}_n$ be a pitch-class. Then we get for the kth Fourier coefficient of $A = \{a\}$ $\hat{a}_k = e^{-2ik\pi a/n}$ hence

$$\hat{a}_{k_1}\hat{a}_{k_2}\cdots = e^{-2ik_1\pi a/n}e^{-2ik_2\pi a/n}\cdots = e^{-2i(k_1 + k_2 + \ldots)\pi a/n} = e^{-2in\pi a/n} = 1.$$

---

[2] Not necessarily distinct.
[3] The qualification "regular" distinguishes these from an arbitrary product, but since non-regular products are of no evident interest, we will typically omit the qualifier.

As we will see this a special case of Proposition 3, A being a transposition of $\{0\}$ for which any product is trivially regular.

Lemma 1. *The argument of the sum (or mean) of two complex numbers with equal magnitude is the mean value of their arguments.*

Remark.

There is a catch here. Since arguments are defined modulo a whole circle ($2\pi, 12$, or n, depending on normalization), half-arguments are defined modulo half the circle ($\pi, 6$, a tritone).

However, out of these two opposite directions, the appropriate one is the mean value which directs the interior of the angles between the two complex numbers, see Fig. 1. In other words, both complex vectors and their sum/mean must lie in the same half-plane.

Proposition 2. *If A, B are disjoint and homometric, and coherent with respect to some coefficient product, then $A \cup B$ is aligned with respect to that product.*

Proof. Let $\hat{a}_k = |a_k| e^{i\varphi_k}$ be the kth Fourier coefficient for A and similarly $\hat{b}_k = |\hat{b}_k| e^{i\psi_k}$ for B. Then since A, B are homometric, $|\hat{a}_k| = |\hat{b}_k|$. Hence for $C = A \cup B$, one gets

$$\hat{c}_k = \hat{a}_k + \hat{b}_k = |\hat{a}_k|\left(e^{i\varphi_k} + e^{i\psi_k}\right) = |\hat{a}_k| \cos \frac{\varphi_k - \psi_k}{2} e^{i \frac{\varphi_k + \psi_k}{2}}$$

hence the sum of the phases (taken modulo 2π) is

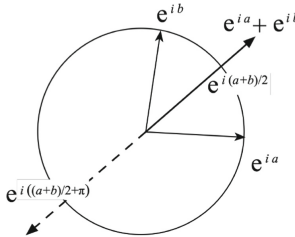$$\sum \frac{\varphi_k + \psi_k}{2} = \frac{1}{2}\left(\sum \varphi_k + \sum \psi_k\right) = 0 \mod \pi.$$

Fig. 1. Sum and phase of two complex numbers with the same length.

For instance for the diatonic partition $2+3+7 = 12$, reunions of homometric dyads are coherent (though most products are 0). Counter-examples would be for instance the chromatic dyad (01) for partition $3 + 4 + 5 = 12$: in this case

$$\hat{a}_3 \hat{a}_4 \hat{a}_5 = 1 - \sqrt{3} < 0.$$

and the product is aligned, but not coherent.

Both properties (coherent/aligned) are invariant by transposition and inversion. More precisely,

Proposition 3. *Coefficient products are transposition-invariant.*

Inversion negates the imaginary part of a coefficient product, while the real part is inversion-invariant.

Proof. For transposition let us have two pc-sets in $\mathbf{Z}_n$ such that $B = A + \tau$. Denoting their Fourier coefficients by $\hat{a}_k, \hat{b}_k$ we derive

$$\hat{b}_k = \hat{a}_k e^{-2i\pi k\tau/n}$$

and hence

$$\hat{b}_{k_1}\hat{b}_{k_2}\ldots\hat{b}_{k_r} = \hat{a}_{k_1}\hat{a}_{k_2}\ldots e^{-2i\pi k_1\tau/n} e^{-2i\pi k_2\tau/n}\ldots$$
$$= \hat{a}_{k_1}\hat{a}_{k_2}\ldots e^{-2i\pi(k_1+k_2+\ldots)\tau/n}$$
$$= \hat{a}_{k_1}\hat{a}_{k_2}\ldots e^{-2i\pi n\tau/n} = \hat{a}_{k_1}\hat{a}_{k_2}\ldots\hat{a}_{k_r}$$

whenever $k_1 + k_2 + \cdots = n$.

The inversion $A \mapsto -A$ just changes the signs of all phases, conjugating all Fourier coefficients, which leaves the real part invariant and inverts the imaginary part. For other inversions $A \mapsto \tau - A$, notice it is the previous inversion combined with a transposition.

It follows easily that any inversionally symmetric pc-set has aligned product. We can be more specific in the following case:

Proposition 4. *For generated scales a regular product is aligned, the sign depending on a product of sines.*

Proof. According to the last proposition we can assume that the generated scale begins on 0:

$$A = \{0, f, 2f, 3f, \ldots (d-1)f\} \quad \text{if the generator is } f \text{ and the cardinality } d.$$

Then we compute $\hat{a}_k = \sum_{j=0}^{d-1} e^{-2i\pi kjf}/n$, a geometric sum:

$$\hat{a}_k = \frac{e^{-2i\pi dkf/n}-1}{e^{-2i\pi kf/n}-1} = \frac{e^{-i\pi dkf/n}(e^{-i\pi dkf/n}-e^{+i\pi dkf/n})}{e^{-i\pi kf/n}(e^{-i\pi kf/n}-e^{+i\pi kf/n})} = e^{-i(d-1)kf\pi/n}\frac{\sin(dkf\pi/n)}{\sin(kf\pi/n)}.$$

Hence, depending on the sign of the sines quotient, the phase is either $\varphi_k = -(d-1)fk\pi/n$ or $\varphi_k = -(d-1)fk\pi/n + \pi$, or in normalized format

$$\Phi_k = -\frac{(d-1)fk}{2} \quad \text{or} \quad -\frac{(d-1)fk}{2} + \frac{n}{2}.$$

Then for any partition $n = k_1 + k_2 + \ldots$, the sum of the $-\dfrac{(d-1)fk_i}{2}$ is a multiple of $n/2$ and so is the sum of all phases, meaning that the coefficient product is real.

These two cases are exemplified by the diatonic and pentatonic scales in 12-tET, which are fifth- (or fourth-) generated: $n = 12, f = 5$ (or 7), and $d = 5$ or 7. The sine in the denominator is $\sin(kf\pi/n) = \sin(5k\pi/12)$, which is positive for $k = 2$, negative for $k = 3, 7$; and the numerator $\sin(dkf\pi/n) = \sin(5^2k\pi/12) = \sin(\pi/12)$ or $\sin(5 \times 7k\pi/12) = -\sin(\pi/12)$, hence the result.

NB: generally, albeit random pc-sets usually do not satisfy coherence, the previous propositions help us understand informally why man-made music may: it is not uncommon to compose using pc-sets built up from small units or bricks, like dyads, symmetric tetrachords, bits of generated scales, or pc-sets close to these, etc.

3   Example: Tonal Pitch-Class Distributions

Pitch-class distributions of tonal music have a number of regular features observable through the DFT, in particular high magnitudes of the fifth and third coefficients [15]. Phases of the fifth and third coefficients can be used to estimate the key of a passage [14]. Tonal distributions also have a clearly observable regularity in one of the coefficient products, $\hat{a}_2\hat{a}_3\hat{a}_7$. This means that the pitch-class watermarks of tonal music include not only the spectral features relating to intervallic content (diatonicity and triadicity) but also at least this one non-spectral feature, determined by the phases of $\hat{a}_2$, $\hat{a}_3$, and $\hat{a}_5$.

Figure 2 shows all of the coefficient products for a windowed analysis of Bach's 3-part inventions, excluding those with duplicated coefficients.[4] Only the inventions in 4/4 are included, and the distributions are taken over all four-beat windows in the piece. In addition to always being the largest coefficient product in all but one case (no. 3), the $\hat{a}_2\hat{a}_3\hat{a}_7$ values also are most consistent in phase, staying close to the positive real axis. Despite being small, the imaginary values also reliably distinguish mode, with the two minor mode pieces (numbers 4 and 9) having the only consistently positive imaginary values.

In Fig. 2 we observe that when the average coefficient product is reliably distinct from the origin, it is usually also approximately coherent, with the main exceptions being in one product, $\hat{a}_3\hat{a}_4\hat{a}_5$. This might be explained by the limited macroharmony tonal music, where *macroharmony* is Tymoczko's term for "the total collection of notes used over moderate spans of musical time" [10, p. 4]. Specifically, tonal music usually deals with a limited number of pitches in circulation at a time, so it is impossible to differentially weight all twelve pitch classes. Rather, the composer chooses a limited set of pitches (the macroharmony) and differentially weights these according to their status in the key and chord, and omits the rest. We therefore expect to see a pitch-class distribution with a "floor." We can model such distributions by imagining starting with the

---

[4] For example, $\hat{a}_5\hat{a}_5\hat{a}_2$ or $\hat{a}_4\hat{a}_4\hat{a}_4$. These are also regular coefficient products and can have interesting applications, but we focus instead on products of three unique coefficients here.
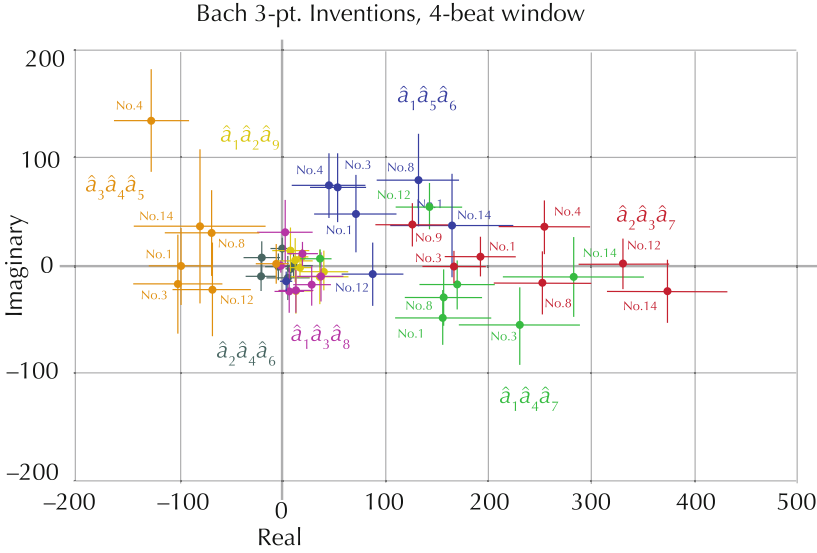
Fig. 2. Average values of coefficient products from common-time Bach three-part inventions. We take averages over all four-beat windows. Standard errors, shown with bars, are corrected for overlap.

full space of distributions with values balanced around zero and normalized to lie between -1 and 1, then applying a clipping filter to eliminate negative values:

$$x \mapsto \begin{cases} x & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}.$$

This is the product of identity by a step function, and can also be expressed as $(x + |x|)/2$.

In the space of continuous maps on (say) $[-1, 1]$, with hermitian norm $f \mapsto \sqrt{\int_{-1}^{1} f^2}$, we get a decent quadratic approximation to this map with $x \mapsto (3 + 16x + 15x^2)/32$, as can be seen in Fig. 3.

Given some distribution in the full space, then, the corresponding clipped distribution will differ primarily by the addition of a positive quadratic term. The DFT of this quadratic term will consist of the products of coefficients in the original distribution, and adding these to the coefficients of the original distribution will push all coefficient products in the direction of coherence.

For instance, if we take the sum of pitch classes from a large number of major-key pieces transposed to C major we get a pitch-class distribution like the one in Fig. 4 (here we use the distribution obtained in [1], but very similar ones could be taken from many other studies). We can approximately resynthesize this from just its two largest Fourier coefficients, $\hat{a}_3$ and $\hat{a}_5$, by taking a sum of these and applying the clipping filter. The result is similar to the original distribution, in
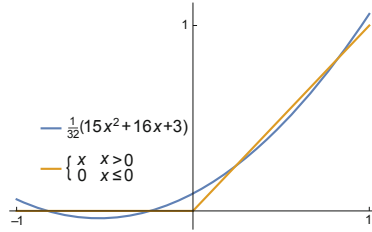
Fig. 3. The clipping function and its best quadratic approximation.

particular recovering an $\hat{a}_2$ similar to the one in the original distribution. The main difference is that the derived distribution has an $\hat{a}_4$ which is suppressed in the original distribution. These $\hat{a}_2$ and $\hat{a}_4$ components of the derived distribution are attributable to the squared term in the quadratic approximation of the clipping function.[5]

We might interpret the data in Fig. 2, then, with the claim that the entire pitch-class distribution is determined roughly by $\hat{a}_3$, $\hat{a}_4$, and $\hat{a}_5$, plus the assumption of limited macroharmony. The limited macroharmony (clipping) filter accounts for the observed values of $\hat{a}_1$ and $\hat{a}_2$, which make coherent products with the other coefficients ($\hat{a}_1\hat{a}_4\hat{a}_7$ and $\hat{a}_2\hat{a}_3\hat{a}_7$).
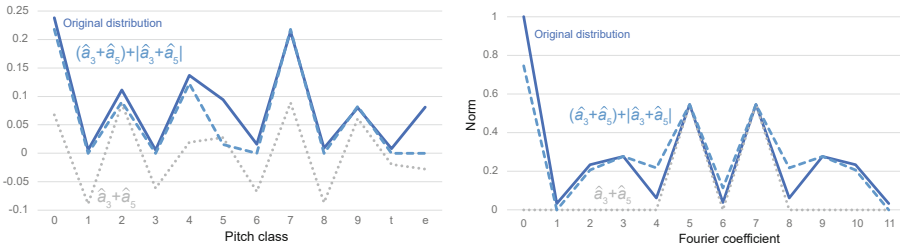
Fig. 4. On the left, a major-key pitch-class distribution from [1], the sum of its 3rd and 5th Fourier coefficients, and the clipping filter applied to this. On the right, the spectra of these.

The coefficient products $\hat{a}_3\hat{a}_4\hat{a}_5$ and $\hat{a}_2\hat{a}_3\hat{a}_7$ both involve the third and fifth coefficients. We might contrast coherence/incoherence in these two products by considering how intervals that are farther apart in $\hat{a}_3$ and $\hat{a}_5$, tritones and semitones, appear in sets that are otherwise relatively concentrated in these dimensions. Figure 5 shows the $\hat{a}_3/\hat{a}_5$ *phase space*, a toroidal space where the coordinates are phases of different DFT coefficients [2,11]. There are two relatively

---

[5] A complication here is that there are two contributors to $\hat{a}_2$ in the quadratic term, $\hat{a}_3\hat{a}_7$ and $\hat{a}_5\hat{a}_5$, and the latter is larger in the distribution derived using the clipping filter. In the original distribution, the phase of $\hat{a}_2$ is closer to that of $\hat{a}_3\hat{a}_7$ than $\hat{a}_5\hat{a}_5$.

parsimonious ways to connect semitones and tritones, along SW-NE or NW-SE diagonals. The former is associated with diatonic semitones and tritones, the latter with chromatic semitones and tritones along the minor-thirds axis. ([11] refers to these as "intervallic axes.") We suggest the term "blues tritone" for NW-SE orientation because it could result from adding blue notes to a pentatonic scale. If pitch classes tend to cluster around a diatonic SW-NE diagonal, it will have a positive $\hat{a}_2\hat{a}_3\hat{a}_7$ and negative $\hat{a}_3\hat{a}_4\hat{a}_5$, which is what we observe in the Bach inventions. If, on the other hand, they cluster around a chromatic/octatonic NW-SE diagonal, we will see the opposite pattern, a negative $\hat{a}_2\hat{a}_3\hat{a}_7$ and positive $\hat{a}_3\hat{a}_4\hat{a}_5$. This pattern, although it can involve equally large values of $|\hat{a}_3|$ and $|\hat{a}_5|$, would be rarely observed in eighteenth-century tonal pitch-class distributions. We will use the term "diatonic" to refer to the first type of set, and "anti-diatonic" the latter type.
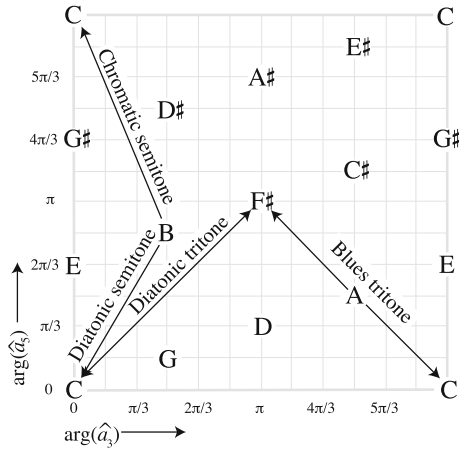

Fig. 5. Different kinds of semitones and tritone in $\hat{a}_3/\hat{a}_5$ phase space associated with positive real $\hat{a}_2\hat{a}_3\hat{a}_7$ (diatonic semitone and tritone) and positive real $\hat{a}_3\hat{a}_4\hat{a}_5$ (chromatic semitone and blues tritone). (Color figure online)

The following example shows that these non-spectral distinctions between diatonic and anti-diatonic material remain salient for twentieth composers in non-tonal contexts.

4   Example: All-Interval Tetrachords and Takemitsu

The all-interval tetrachords (AITs), set classes (0146) and (0137), are a unique instance of small-cardinality Z-related sets, and as such are of particular interest for musically exploring non-spectral properties of set types. The four AIT set types occupy unique locations in $\hat{a}_1\hat{a}_2\hat{a}_9$ and $\hat{a}_2\hat{a}_3\hat{a}_7$ spaces. These are products

that do not involve $\hat{a}_4$ or $\hat{a}_8$, which means that the eight-note chromatic complement of a tetrachord is equal to its four-note octatonic complement (because the octatonic is nil on all coefficients). Therefore the AIT pairs (0137)-(0256) and (0146)-(0467), which are octatonic complements, behave like ordinary complements in these spaces, with equal magnitude and opposite phases. This special relationship between AITs and the octatonic relates to the CUP property explored by Childs [5] and Capuzzo's Q-operations [4].

The differences in $\hat{a}_2\hat{a}_3\hat{a}_7$ explain how the AITs differ in quality despite their equivalent intervallic content. The imaginary dimension is associated with the major/minor contrast. Inversion reverses the sign of the imaginary part, so (0137) and (0467) have the same real part in $\hat{a}_2\hat{a}_3\hat{a}_7$, but opposite imaginary part. The "major" (0467) and (0256) have negative imaginary values. The real dimension distinguishes whether the thirds and fifth are arranged to imply a diatonic semitone and tritone (positive) or a chromatic semitone and blues tritone (negative). The (0137) tetrachords are distinguished by their triadic subset and the (0146)s by the distinctive non-diatonic subset, (014).

Takemitsu's solo flute piece, "Air," uses an all-interval tetrachord as its principal motive, and the major/minor and diatonic/anti-diatonic contrasts of $\hat{a}_2\hat{a}_3\hat{a}_7$ space are important to the harmonic language of the piece. Figure 6 shows a parsing of the first 14 measures, and Fig. 7 plots these in $\hat{a}_2\hat{a}_3\hat{a}_7$ space. The central thematic role of (0467) is immediately apparent in its prominent statement in the opening and in m. 6. The opening gesture also defines a larger set, (014578), which is similar to (0467) in $\hat{a}_2\hat{a}_3\hat{a}_7$: it is close in phase, and slightly farther from the origin. This larger set returns in m. 9. Altogether, this establishes a departure-return script in which the principal motive alternates with harmonically contrasting material.

Octatonic and whole-tone material are essential to Takemitsu's harmonic methods even though complete octatonic and whole-tone collections never appear. These collections are special in that they have $\hat{a}_2 = \hat{a}_3 = \hat{a}_5 = 0$. This makes them useful to create a kind of negative space: while adding a complete octatonic or whole-tone collection has no effect on $\hat{a}_2\hat{a}_3\hat{a}_7$, adding an incomplete collection has the effect of negating the missing pitch class(es), which may be understood as a kind of partial complementation [12]. The set (024689), for example, is a whole-tone collection plus an "anti-semitone" (the added note and omitted note are a semitone apart): (02468t)\t ∪ 9. The set type (023468), directly following it, similarly, is a whole-tone collection plus "anti-fifth": (02468t)\t ∪ 3. The next set, (0346), is a diminished seventh plus an anti-fifth, and is therefore has the same $\hat{a}_2\hat{a}_3\hat{a}_7$ value. Whereas dyads are always on the real axis in a coefficient product space, anti-dyads are always on the imaginary axis.

Takemitsu's first contrast juxtaposes the principal motive with these two whole-tone-plus-anti-dyad collections, which neutralize the diatonic element, and highlight major/minor contrasts on the imaginary axis. In particular, the large minor value of (024689) contrasts with the large major value of the principal motive.
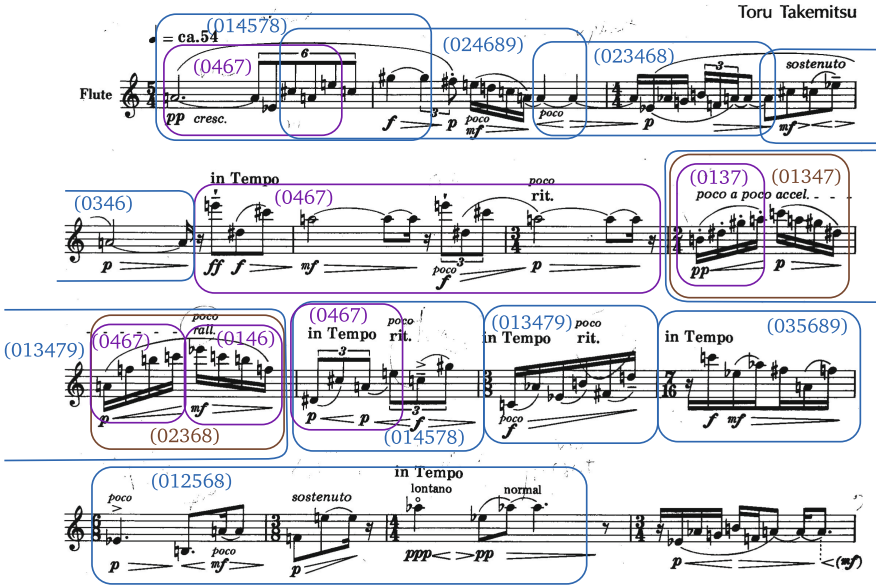
Fig. 6. Meas. 1–14 of Takemitsu's *Air* for solo flute, with important 4–6 note sets identified.
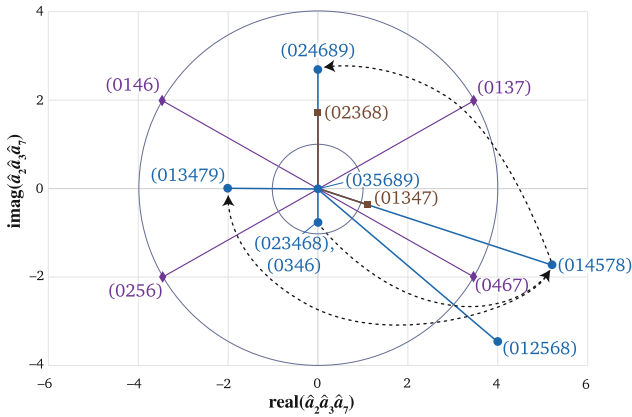

Fig. 7. Pitch-class sets from *Air* in $\hat{a}_2\hat{a}_3\hat{a}_7$ space. Diamonds show the AITs. Dots and squares show set types that appear in the passage.

The second contrast takes us into the anti-diatonic region through the use of octatonic collections. This is the first place where Takemitsu uses the contrasting AITs, as well as a transposition of the initial (0467). These combine into larger octatonic sets, the octatonic complement of (034) in m. 7 and the octatonic complement of (046) in m. 8. The entire two measures constitute a single hexachord, which is the octatonic complement of (04). The two pentachords are

equivalent in $\hat{a}_2\hat{a}_3\hat{a}_7$ phase to two sets that we have already heard (the principal motivic hexachord in m. 1, and the first contrasting hexachord in m. 2). Their combination, however, introduces a new anti-diatonic element. As if to underscore the point, Takemitsu restates this anti-diatonic octatonic hexachord more compactly in m. 10, immediately after repeating the head motive in m. 9.

The last gesture uses a pitch-class set that is harder to easily characterize, yet Takemitsu communicates a sense of return with the rhythmic broadening, the clear phrase break, and the return to the high G♯/A♭ that marked the registral goal of the basic idea in mm. 1 and 9. The $\hat{a}_2\hat{a}_3\hat{a}_7$ value is consistent with this: it is large and close in phase to (0467) and (014578).

5   Conclusion

Coefficient products were first discovered empirically in [13,14], in the form of coherent $\hat{a}_2\hat{a}_3\hat{a}_7$ in tonal distributions, which were at first difficult to explain. The present study reveals some of the general properties of coefficient products and why we might observe coherent products in distributions from real music. While Fourier coefficients are mathematically independent in principle, the constraints on real distributions mean that they are not always independent in practice. In tonal distributions, the presence of significant $\hat{a}_2$ may actually be a mathematical artifact of $\hat{a}_5$, $\hat{a}_3$, and limited-macroharmony constraints. At the same time, a significant $\hat{a}_4$ coefficient may actually be concealed by similar artifacts.

We have also revealed a potential wealth of other applications of coefficient products, including analysis of Z-related sets, post-tonal music, and distinguishing anti-diatonic sets, particularly those like the pentatonic that share a large $|\hat{a}_5|$ with diatonic sets.

References

1. Albrecht, J., Shanahan, D.: The use of large corpora to train a new type of key-finding algorithm. Music Percept. **31**, 59–67 (2013)
2. Amiot, E.: The torii of phases. In: Yust, J., Wild, J., Burgoyne, J.A. (eds.) MCM 2013. LNCS (LNAI), vol. 7937, pp. 1–18. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39357-0_1
3. Amiot, E.: Music in Fourier Space. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-45581-5
4. Capuzzo, G.: Lewin's Q operations in Carter's Scrivo in Vento. Theory Pract. **27**, 85–98 (2002)
5. Childs, A. P.: Structural and transformational properties of all-interval tetrachords. Music Theory Online **12**(4) (2006)
6. Forte, A.: A theory of set-complexes for music. J. Music Theory **8**(2), 136–183 (1964)
7. Forte, A.: The Structure of Atonal Music. Yale University Press, New Haven (1973)
8. Quinn, I.: General equal-tempered harmony: parts two and three. Perspect. New Music **45**(1), 4–63 (2006)
9. Robinson, T.: Pitch-class multisets. Ph.D Diss., City University of New York (2009)

10. Tymoczko, D.: A Geometry of Music: Tonality and Counterpoint in the Extended Common Practice. Oxford University Press, Oxford (2011)
11. Yust, J.: Schubert's harmonic language and Fourier phase space. J. Music Theory **59**(1), 121–81 (2015)
12. Yust, J.: Harmonic qualities in Debussy's "Les sons et les parfums tournent dans l'air du soir". J. Math. Music **11**(2–3), 155–173 (2017)
13. Yust, J.: A three dimensional model of tonality. Paper Presented to the American Mathematical Society, Athens, GA, 3 May 2016
14. Yust, J.: Probing questions about keys: tonal distributions through the DFT. In: Agustín-Aquino, O.A., Lluis-Puebla, E., Montiel, M. (eds.) MCM 2017. LNCS (LNAI), vol. 10527, pp. 167–179. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-71827-9_13
15. Yust, J.: Stylistic information in pitch-class distributions. J. New Music Res. **48**(3), 217–231 (2019)

Tetrachordal Folding Operations

Jason Yust[✉]

Boston University, Boston, MA 02215, USA
`jyust@bu.edu`

Abstract. Jonathan Bernard's trichordal folding operations relate trichords with a maximum of shared interval content. This paper generalizes this to any cardinality of chord, focusing on the case of tetrachordal folding. A tetrachordal folding holds one trichordal subset fixed and inverts another around a shared dyad, so that the two tetrachords share five interval classes and two trichordal subsets. These operations generalize naturally from pitch space to pitch-class space and to set classes. The last section of the paper demonstrates the analytical application of tetrachordal folding networks on Morton Feldman's "For Stephan Wolpe."

Keywords: Pitch-class set theory · Folding · Interval content · Morton Feldman

1 Trichordal Folding

In his work on Edgar Varése, Jonathan Bernard [2] defines "infolding" and "unfolding" operations that relate trichords of different types. He uses a successive-interval notation which I will adopt here, generalizing over transposition. The following definition is essentially Bernard's, with some new notation.

Definition 1. *Let a pitch-space trichord A be given by successive intervals (a_1, a_2), for $a_1, a_2 \in \mathbf{Z}$, then there are four unfolding operations.*

$$
\begin{aligned}
\mathrm{unf}_{1a}(A) &= (a_1, a_1 + a_2) \\
\mathrm{unf}_{1b}(A) &= (a_1 + a_2, a_1) \\
\mathrm{unf}_{2a}(A) &= (a_1 + a_2, a_2) \\
\mathrm{unf}_{2b}(A) &= (a_2, a_1 + a_2)
\end{aligned}
\tag{1}
$$

There are also two infolding operations.

$$
\begin{aligned}
\mathrm{inf}_a(A) &= \begin{cases} (a_1, a_2 - a_1), & \text{if } a_1 \le a_2 \\ (a_2, a_1 - a_2), & \text{if } a_2 \le a_1 \end{cases} \\
\mathrm{inf}_b(A) &= \begin{cases} (a_2 - a_1, a_1), & \text{if } a_1 \le a_2 \\ (a_1 - a_2, a_2), & \text{if } a_2 \le a_1 \end{cases}
\end{aligned}
\tag{2}
$$

Altogether these are the complete set of folding operations.

Figure 1 shows an example of these operations, which we can interpret as taking any one note from the trichord and inverting it around one of the other two notes (hence the six possibilities). The a and b versions of each operation are clearly always going to be related by inversion, so we can immediately simplify these operations by generalizing over inversions, as Bernard [2] does. Therefore if we consider unfolding and infolding as a relation, *folding*, we immediately have:

Proposition 1. *Folding is a symmetrical relation.*

Proposition 2. *If $A = (a_1, a_2)$ is in the folding relation with B, then so is the inversion of A, (a_2, a_1).*

| | A | $\mathrm{unf}_{1a}(A)$ | $\mathrm{unf}_{1b}(A)$ | $\mathrm{unf}_{2a}(A)$ | $\mathrm{unf}_{2b}(A)$ | $\mathrm{inf}_a(A)$ | $\mathrm{inf}_b(A)$ |
| $(1, 4)$ | $(1, 5)$ | $(5, 1)$ | $(5, 4)$ | $(4, 5)$ | $(1, 3)$ | $(3, 1)$ |

Fig. 1. Examples of unfoldings and infoldings of a chord.

This study pursues generalizations and applications of Bernard's idea taking the key feature to be that folding operations always preserve all the intervals of the set except at most one. These operations therefore express the minimum possible change in interval content (as represented for pitch-class sets, e.g., by Forte's interval vector [4]). While Definition 1 follows Bernard in defining foldings in pitch space (using intervals in $\mathbf{Z}$), I will be primarily interested in the generalization to pitch-class sets defined below. Most of the results in this section and the next nonetheless apply in both domains.

The next section will generalize folding operations to tetrachords and higher cardinalities with these priorities in mind, and the third section applies tetrachordal folding networks to an analysis of Morton Feldman's "For Stephan Wolpe."

For a trichord, (a_1, a_2), there are three possible ways to exchange one interval with a new one specified in Definition 1. Either we exchange a_2 for $2a_1 + a_2$ (unf_1), a_1 for $a_1 + 2a_2$ (unf_2), or $a_1 + a_2$ for $\pm(a_2 - a_1)$ (inf).

Notice that sets will be generalized over transposition throughout this study (hence the use of interval strings to define them) but not necessarily over inversion or octave equivalence. However, it is possible to transfer all of the definitions of foldings to pitch or pitch-class sets proper (not generalized over transposition) by fixing the transposition of the fixed dyad (or fixed trichord, in the case of tetrachordal foldings below), as the illustration in Fig. 1 does.

In Bernard's applications of the folding operation the most important feature is that they are defined on interval strings, so he can interpret them as operating on trichords in pitch space. However they easily generalize to operations of set-classes. To do so, we reconceive the operations as acting on ordered sets as in [7],

relaxing the assumption that order is conferred by registral position and allowing a_1 and a_2 to take negative values. Bernard's distinction between "unfolding" and "infolding" then becomes less meaningful, and we rename the operations simply as *foldings* indexed by the order position of the moving note followed by the order position of the note it is inverted around. Hence:

Definition 2. *There are six trichordal* folding *operations defined on interval strings, (a_1, a_2), with $a_1, a_2 \in \mathbf{Z_{12}}$. These are defined as in Definition 1, but with sums taken modulo 12, as* $\text{fold}_{21} = \text{unf}_{1a}$, $\text{fold}_{31} = \text{unf}_{1b}$, $\text{fold}_{23} = \text{unf}_{2a}$, $\text{fold}_{13} = \text{unf}_{2b}$, $\text{fold}_{12} = \text{inf}_a$, $\text{fold}_{32} = \text{inf}_b$,

Definition 3. *Two trichordal set classes, A and B, relate by* folding, *or $A \asymp B$ if any pitch-space representatives of A and B relate by one of the folding operations from Definition 1.*

Proposition 3. *If two set classes, A and B, are related by folding, $A \asymp B$, then for any pitch-space representative of A, there is a pitch-space representative of B that relates to it by folding.*

Proof. It suffices to show that applying transposition, inversion, octave shift, or permutation to A results in the same operation applied to B (and possibly a change of the exact unfolding or infolding relation). The first two are straightforward. Octave shift refers to adding or subtracting multiples of 12 to/from individual pitches. For instance, replacing (a_1, a_2) with $(a_1, a_2 + 12)$ (last note moves up by octave). This clearly will simply induce some octave shift in B. Finally, given a permutation on A we can apply the same permutation to the indices of the folding operation and leave B unchanged.   □

Figure 2 displays the network of trichordal set classes relating by $\asymp$ using a 2-dimensional parameterization of the interval vector. I chose the parameterization arbitrarily with the goal of disambiguating all of the set classes and avoiding crossing edges. In addition this parameterization shows the symmetry of the network under the M5 automorphism of $\mathbb{Z}_{12}$ [8] by having the horizontal dimension dependent only on #ic1 and #ic5.

Note that we could extend the network in Fig. 2 to include doubled ic1 and ic5 dyads, (001) and (005). However, other doubled dyads, whole tone chords, and diminished triads cannot exist in the same network. More generally, the sets in a given network have to have the same minimal embedding equal temperament. A whole-tone chord has 6-tET as a minimal embedding ET, and a doubled minor third, (003), has 4-tET as a minimal ET. This is because the interval of a semitone (or fourth or fifth, etc.) can never be produced by sums and differences of intervals in a smaller minimal embedding universe. This applies in pitch space as well as pitch-class space. For instance, the pitch-space chord (5,5) does not interact with the chords that have 12-tET as their minimal universe, even though its pitch-class equivalent (2,5) does.
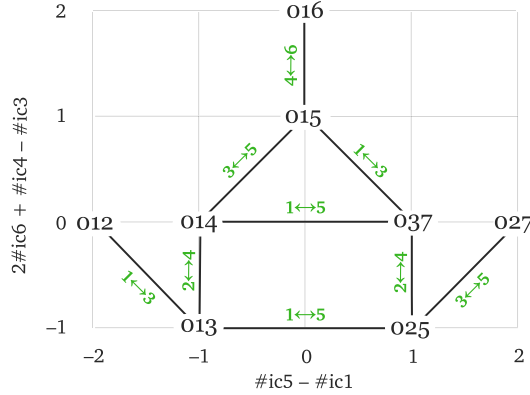
Fig. 2. Trichordal folding network.

2   Tetrachordal Folding

Bernard [1,2] explores the possibility of extending the folding operations to tetrachords, but his definition is far too loose, resulting in an unwieldy number of relations. If our goal is to preserve the property of maximizing similarity of interval content, however, an effective generalization is ready at hand. In the trichordal foldings, we hold one dyad constant while moving the third note so as to preserve one of the intervals it makes with the other two notes, by inverting it. This process guarantees that two of the three intervals will remain the same. This immediately generalizes to tetrachords: hold three of the notes constant, and choose another trichordal subset containing the fourth note. Invert this trichordal subset around the dyad it shares with the first trichord. Then exactly one note moves, preserving two of the trichordal subsets (up to inversion), and by extension, five of the six intervals.

This definition of tetrachordal folding makes a useful and manageable relation. The possibilities are listed in Table 1 and Fig. 3 provides an example on a set in pitch space. The choice of two out of four trichordal subsets leads to twelve possible folding relations, which can be reduced to six by inversional equivalence. We can immediately see that the properties of trichordal folding described in the previous section generalize to tetrachordal folding, including Propositions 1, 2, and 3.

A similar generalization to larger sets is immediately evident, but there is an important caveat. For a set of size n, when we choose two $(n-1)$ subsets, their intersection (of size $n-2$) must be inversionally symmetrical for the operation to be well defined. For $n-2 = 1, 2$ this is guaranteed, but for $n \geq 5$ it becomes a significant restriction, more so for larger n. Interestingly for $n = 5$ in 12-tET all set classes have at least one inversionally symmetrical trichordal subset, so pentachordal folding operations are worth exploring, but I will pursue this no further at present.

Table 1. Tetrachordal folding operations on an intervallically defined set (a_1, a_2, a_3), permuted to keep all the intervals positive for $a_3 > a_1$ and $a_1 > a_3$ respectively.

| Operation	Definition	Operation	Definition	Interval change
$\text{fold}_{2(13)}$	(a_2, a_1, a_3)	$\text{fold}_{4(13)}$	(a_3, a_1, a_2)	$(a_2 + a_3) \rightarrow (a_1 + a_3)$
$\text{fold}_{3(24)}$	(a_1, a_3, a_2)	$\text{fold}_{1(24)}$	(a_2, a_3, a_1)	$(a_1 + a_2) \rightarrow (a_1 + a_3)$
$\text{fold}_{1(23)}$	$(a_2, a_1, a_3 - a_1)$	$\text{fold}_{4(23)}$	$(a_3 - a_1, a_1, a_2)$	$(a_1 + a_2 + a_3) \rightarrow |a_3 - a_1|$
	or $(a_2, a_3, a_1 - a_3)$		or $(a_1 - a_3, a_3, a_2)$	
$\text{fold}_{1(34)}$	$(a_2, a_3, a_1 + a_2)$	$\text{fold}_{2(34)}$	$(a_1 + a_2, a_3, a_2)$	$(a_1) \rightarrow (a_1 + 2a_2 + a_3)$
$\text{fold}_{3(12)}$	$(a_2, a_1, a_2 + a_3)$	$\text{fold}_{4(12)}$	$(a_2 + a_3, a_1, a_2)$	$(a_3) \rightarrow (a_1 + 2a_2 + a_3)$
$\text{fold}_{3(14)}$	$(a_1, a_3 - a_1, a_1 + a_2)$	$\text{fold}_{2(14)}$	$(a_1 + a_2, a_3 - a_1, a_2)$	$(a_2) \rightarrow |a_3 - a_1|$
	or $(a_3, a_1 - a_3, a_2 + a_3)$		or $(a_2 + a_3, a_1 - a_3, a_3)$	
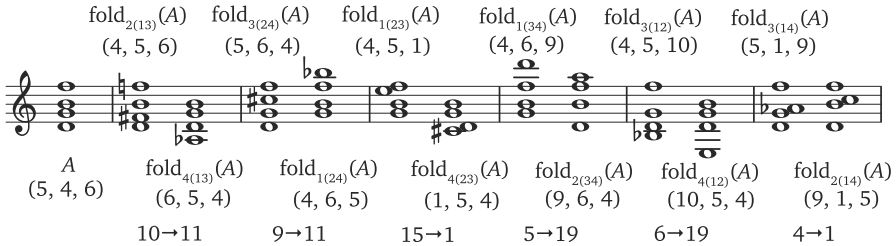
Fig. 3. An example of tetrachordal folding operations on an open-position dominant seventh chord in pitch space.

Figure 4 shows the network of tetrachordal folding operations on non-degenerate tetrachords without doublings. Again, I choose an arbitrary parameterization that disambiguates all of the set classes, avoids crossing edges, and shows the M5 automorphism as a mirror symmetry around a vertical axis. Note that the network is not planar, so it is impossible to eliminate all crossing edges. It is also impossible to disambiguate all of the set classes based on the interval vector alone, because of the all-interval tetrachords (0146) and (0137), which have the same interval vector. Therefore, I also include (in the horizontal dimension) a count of two trichord types, the major/minor triad (037) and its M5 partner (014). This means that edges representing the same change of interval classes are not always exactly the same distance in the horizontal dimension.

The tetrachordal folding operations always preserve two out of four trichordal subsets and five out of six intervals by design. A natural question is whether they are the only such operations. In pitch space, this is in fact the case, but not in pitch-class space.

Proposition 4. *Two tetrachordal pitch sets A and B, not related by transposition or inversion, are related by a folding operation if and only if they share two trichord types (generalized over inversion) as subsets and five interval types.*

Proof. The forward implication (only if) is true by construction. It is only necessary to prove the converse.
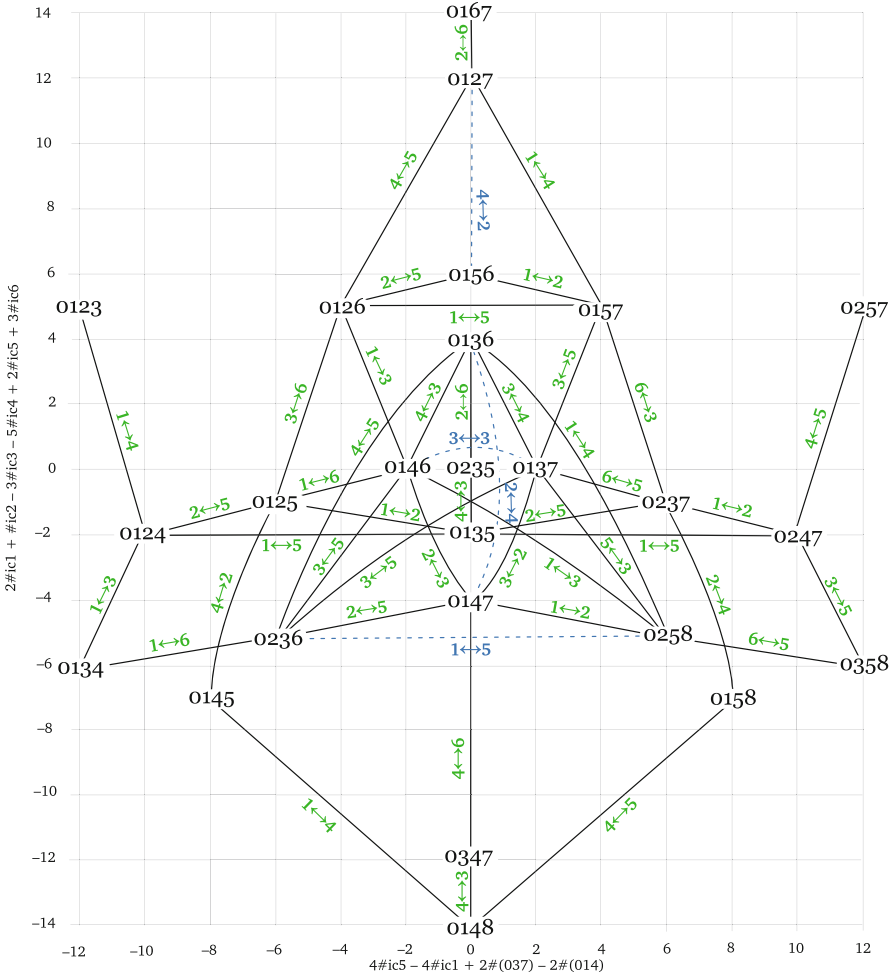
Fig. 4. Tetrachordal folding network.

Assume then that A has trichordal subsets, α and β, and B has subsets of the same types, and the two sets share five interval types. Let B' be either B itself or an inversion of B that contains α, and let β' be the subset of B' related to β by transposition and/or inversion. Either $\beta' = \mathrm{T}_\tau(\beta)$ for some interval τ or $\beta' = \mathrm{I}_\sigma(\beta)$ for some inversional index σ. For both cases, let α have the interval series (a_1, a_2) and A be (a_1, a_2, a_3) such that β is (a_2, a_3). Note that there is no loss of generality through free choice of permutation (allowing, e.g., that a_1, a_2, a_3 can take negative values).

By construction, β' must share a dyad with α; let this be an interval b. The first possibility is that β' shares the same dyad with α as β, and $b = a_2$, which means that the five intervals shared by A and B, $a_1, a_2, a_3, a_1 + a_2$, and

$a_2 + a_3$, and the one belonging only to A, $a_1 + a_2 + a_3$, are all potentially unique. Otherwise, there is a redundancy in the interval β' shares with α. Either $b = (a_1$ or $a_1 + a_2)$ (as a subset of α) $= (a_3$ or $a_2 + a_3)$ (as a subset of β'), or $a_1 = a_2$ or $a_1 + a_2 = a_2$.

First consider the last two possibilities. If $a_1 = a_2$, then α is symmetrical. Transpose or invert B' to map β' onto β. α and its transposition or inversion will still have a common dyad, so the sets will relate by a flip of α. If $b = a_1 + a_2 = a_2$, then $a_1 = 0$ and α has a doubled note. Freely choosing between the doubled notes, this situation then coincides with the regular $b = a_2$ scenario.

For the remaining possibilities first assume that $\beta' = \mathrm{T}_\tau(\beta)$. If $b = a_2 \neq a_1$ then $B' = A$, contradicting the premise. For all of the other possibilities, we can find an additional inversion of β that shares a dyad with β', or an inversion of α sharing a dyad with α. (I leave it to the reader to work out the details.) Therefore $A \asymp B$.

Finally assume that $B' = \mathrm{I}_\sigma(\beta)$. For $b = a_2 \neq a_1$ we have $A \asymp B$. For $a_1 = a_3$ or $a_1 + a_2 = a_2 + a_3$, we have that α is an inversion of β, and hence also $A \asymp B$. In the remaining two cases we violate the interval-sharing premise. Consider $b = a_1 + a_2 = a_3$; B then has two copies of a_2 whereas A has two copies of $a_1 + a_2$, in addition to the other distinct intervals $2a_2 \in B$ and $2(a_1 + a_2) \in A$. The remaining case $b = a_1 = a_2 + a_3$ is essentially the same swapping the roles of a_3 and a_1. □

An interesting consequence of this proof is in the last condition: it is possible for two tetrachords to share two trichordal subsets but not five intervals, in which case they are not directly related by folding, specifically when there is some duplicated interval in the one set and a different duplicated interval in the other. An example would be (0135) and (0136), which each have (013) and (025) subsets, but the first has an (024) subset with two copies of ic2 and an ic4, and the other an (036) subset with two copies of ic3 and an ic6.

This proof only holds in pitch space. Most of it transfers to pitch-class space, except one conclusion: it is possible for β' to overlap α in the interval that makes a_2 and be a non-trivial transposition of β if a_2 is a tritone. Therefore, when dealing with set classes, there is one operation that is not a folding but has the same properties of preserving five intervals and two trichord types. Specifically, for a tetrachord containing a tritone, transpose one of the notes not belonging to it by tritone. This preserves both intervals with the tritone, and just changes one, the interval between the two non-tritone notes.

We can also generalize this to other cardinalities, as *T-shift*. Specifically:

Definition 4. *Let pitch-class set A of size n have a T_x-symmetrical subset of size $n-2$. A T-shift of A fixes a size $n-1$ subset that includes the T_x-symmetrical subset, and moves the remaining note by some multiple of x.*

Figure 4 shows the T-shift operations on tetrachords not equivalent to foldings with dashed lines.

This motivates the following, which I leave as conjectures.

Conjecture 1. Let A and B be pitch sets of cardinality n. Then A and B share two subsets of cardinality $n-1$ and $\binom{n}{2}-1$ intervals if and only if they relate by a folding operation.

Conjecture 2. Let A and B be pitch-class sets of cardinality n. Then A and B share two subsets of cardinality $n-1$ and $\binom{n}{2}-1$ intervals if and only if they relate by a folding operation or a T-shift operation.

3   Application to Morton Feldman's "For Stephan Wolpe"

The first 12 minutes of Morton Feldman's work for chorus and vibraphones, "For Stephan Wolpe," repeat a single progression of ten chords in four-part chorus with small variations, primarily in rhythm, transposition, and voicing (changing the registral ordering and octaves of the notes without changing the set type), reflecting his concept of "crippled symmetry" [3,5]. The set classes that Feldman uses exist in a relatively compact region of the folding network, as shown in Fig. 5. Successive chords are usually two or three steps apart, the only exception being both progressions involving the exceptional (0057) chord. If we skip over this chord then the entire progression is in steps of 2 or 3.

The special property of these progressions is that exactly one trichord is shared between successive chords in all cases except the last progression, (0136)-(0135), which share two. In almost all cases the shared trichord is (015) – here again the exception is the penultimate (0136) chord. This shares an (016) with the preceding (0156) and (013) and (025) with the following (0135). Figure 5 shows regions defined by shared trichords between the chord types that Feldman uses.

The first version of the progression, given in Fig. 6 is representative, containing the voicings used most frequently in all subsequent versions of the progression. Although the trichordal subsets are never voiced in exactly the same way from one chord to the next, they usually involve a single octave adjustment so that individual intervals are preserved in their exact pitch distance. In particular, the characteristic 4-semitone interval of (015) is almost always present, usually as E-G♯ or B-D♯. The 7-semitone interval ties together instances of (027), (016), and (015), and the 14-semitone interval ties together (027) and (025). The final chord, which changes the voicing and transposition of the initial (0135) but preserves its bass note, serves as a kind of summary of the whole progression, including all of these intervals.

The second half of the piece regularizes the rhythm and leaves behind the ten-chord progression in favor of a series of repeated three-chord progressions. These are organized into eight phrases by means of the punctuating vibraphone passages. The first six phrases explore small regions of the tetrachord network, as shown in Fig. 7. Adjacent chords in this section are often related directly by folding and never by more than two links in the network (with the exception of the last chord of phrase 6, an isolated whole-tone chord which does not occur in the network). The first four phrases rely exclusively on chords with (014) subsets.
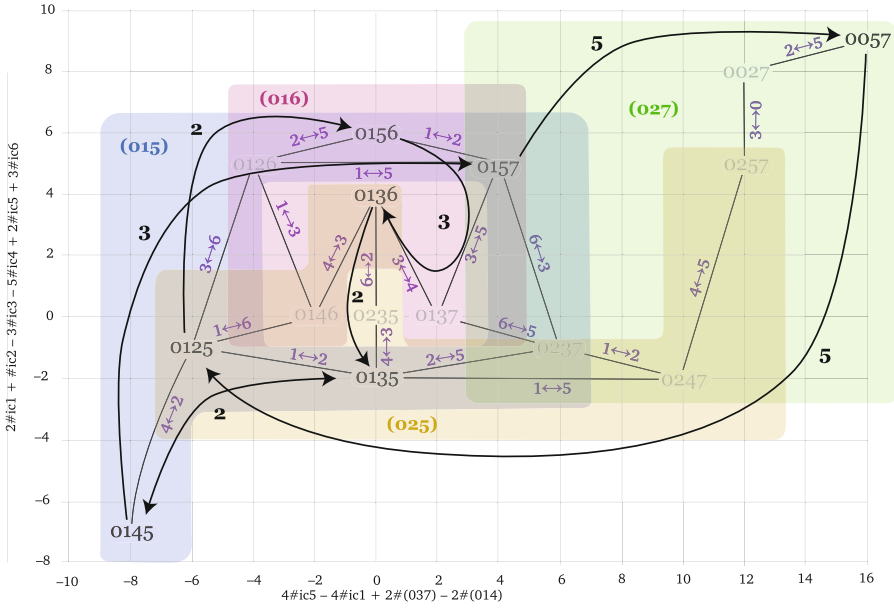
Fig. 5. Tetrachordal folding network for the first part of Feldman's "For Stephan Wolpe." Shaded regions show significant shared trichords.
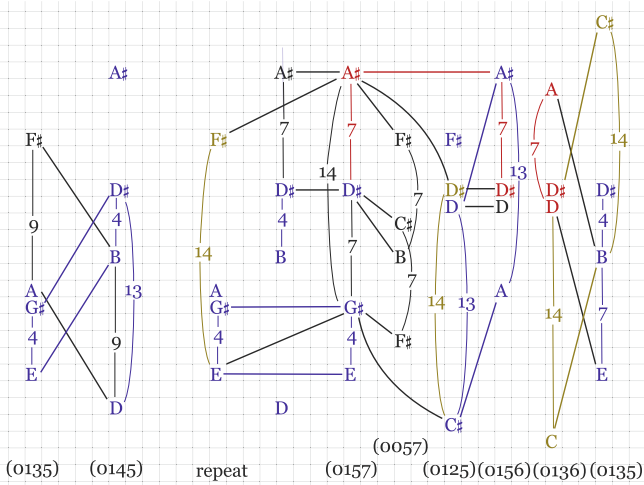

Fig. 6. The first (and representative) version of the chord progression that defines the first half of Feldman's "For Stephan Wolpe" in graph notation. Pitch-space intervals shared between successive chords are highlighted. Colors selectively indicate membership in different trichordal subsets, with blue for (015), red for (016), gold for (025). (Color figure online)
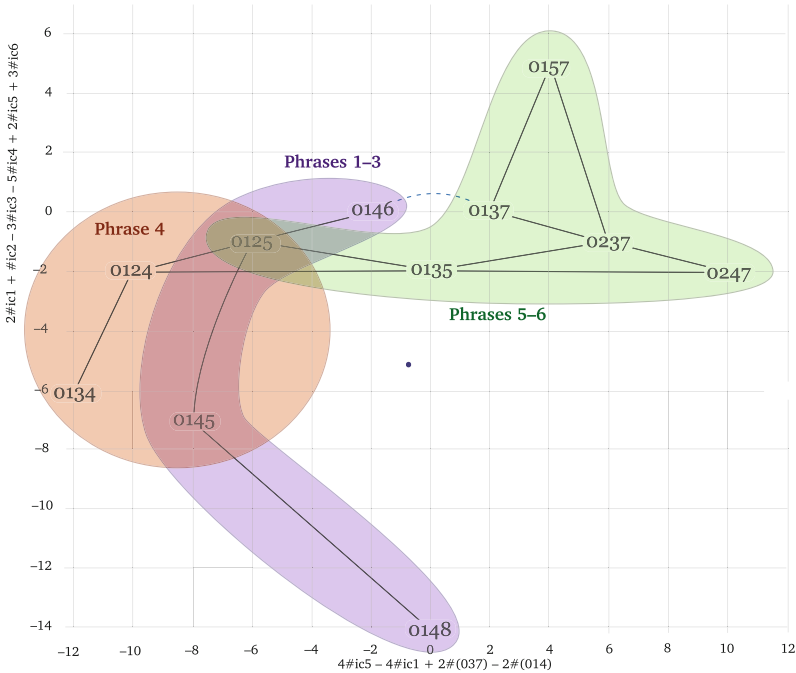
Fig. 7. Tetrachordal folding network for the first six phrases of the second part of Feldman's "For Stephan Wolpe."

A distinct shift happens in the second-to-last phrase, which is considerably longer than any other, at 27 measures with repeats. Now adjacent chords are 2–4 steps apart except for two instances where (0135) and (0235) are adjacent. The progression includes isolated extreme chords, (0123) and (0127), along with (0257), which is more connected with other chords in the passage though it is never directly adjacent to (0247) in the progression. Figure 8 shows these chords in the tetrachordal folding network along with those of the last phrase. The last phrase continues to focus on progressions between chords that are 2–4 steps apart, but overall relies upon a more connected set of chords, excluding (0123) and (0257). The central harmony of phrases 1–6, (0125), returns in this last phrase after being absent for all of the penultimate phrase.

The tetrachordal folding networks thus help us circumnavigate some of the usual problems of analysis and form in Feldman's late music. Hanninen [6], for example, points out that pervasive repetition and lack of textural changes in these pieces inhibits segmentation. While this description appears to characterize the second half of "For Stephan Wolpe" well, a closer look at the use of chord types and their arrangement in the folding network reveals a more definite plan in distinct stages. A relatively limited subnetwork first expands, and then moves back towards a region familiar from the first part of the piece. At the last stage, Feldman also returns to progression types familiar from the first part, charac-
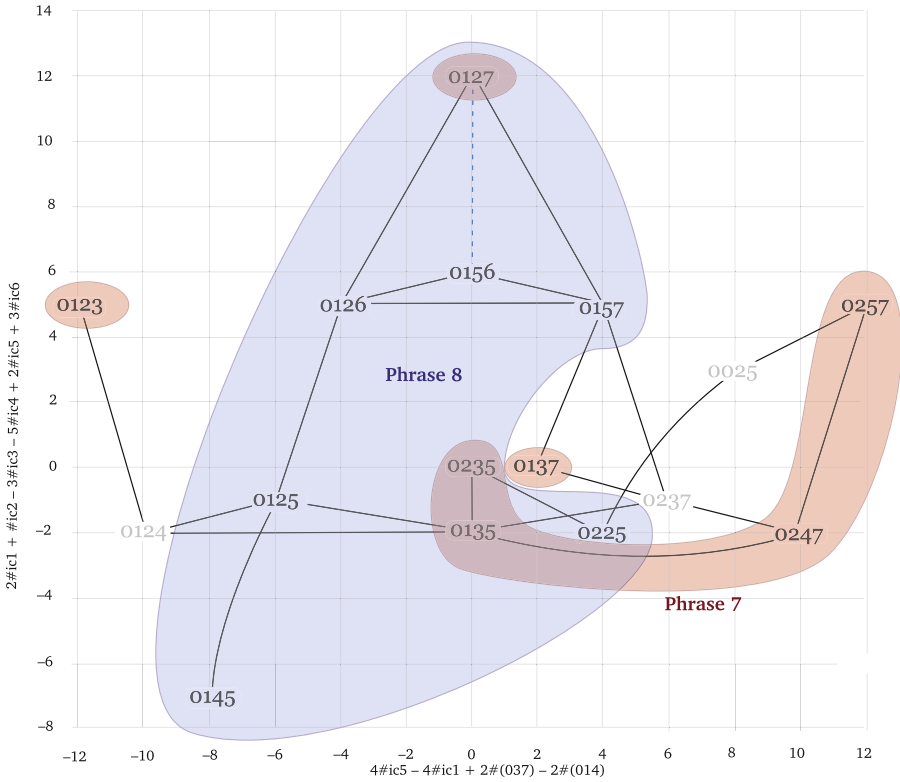
Fig. 8. Tetrachordal folding network for the last two phrases of the second part of Feldman's "For Stephan Wolpe."

terized by larger distances in the network. At the same time, the chord palette expands to the far reaches of the network. The piece thus exhibits a cogent form that combines principles of return and terminal expansion.

4   Conclusion

Folding operations, a generalization of Bernard's [2] unfolding and infolding operations, can be defined on sets of any size, and can operate in pitch-space or pitch-class space, with or without generalizations over inversion. Networks of these operations are useful for mapping out distances between chord-types based on interval and subset content. The application to tetrachords in Feldman's "For Stephan Wolpe" shows that distances in these networks are musically meaningful and that chord types limited to connected subnetworks can be compositionally useful. Feldman's manipulation of distances between adjacent chords and connectedness and location of his subnetworks illustrate his sensitivity to these properties, and reveals a shape to this piece that is not immediately apparent on its rather simple and static surface.

References

1. Bernard, J.: A theory of pitch and register for the music of Edgar Varése. PhD diss., Yale University (1977)
2. Bernard, J.: The Music of Edgar Varése. Yale University Press, New Haven (1987)
3. Feldman, M.: Give My Regards to Eighth Street: Collected Writings of Morton Feldman. In: Friedman, B.H. (eds.). Exact Change Press, Cambridge (2000)
4. Forte, A.: The Structure of Atonal Music. Yale University Press, New Haven (1973)
5. Gelleny, S.A.: Variation techniques and other theoretical issues in Morton Feldman's "carpet" compositions. Ph.D. diss, SUNY Buffalo (2000)
6. Hanninen, D.A.: Feldman, analysis, experience. Twentieth-Century Music **1**(2), 225–251 (2004)
7. Lewin, D.: Generalized Musical Intervals and Transformations, 2nd edn. Oxford University Press, Oxford (2011)
8. Starr, D.: Sets, invariants and partitions. J. Music Theory **22**(1), 1–42 (1978)

Mathematical Techniques and Microtonality

Continuous Chromagrams and Pseudometric Spaces of Sound Spectra

Jordan Lenchitz[1(✉)] and Anthony Coniglio[2]

[1] Florida State University, Tallahassee, FL 32306, USA
jlenchitz@fsu.edu
[2] Columbia University, New York, NY 10027, USA
coniglio@math.columbia.edu

Abstract. In this paper we extend the ubiquitous music information retrieval technology of the quantized chromagram (or chroma feature) to propose a continuous chromagram, an octave-reduced spectrogram that logarithmically reduces the frequencies of a sound spectrum onto a half-open interval $[a, 2a)$ we call a chroma octave. We prove that for any real number $r > 1$, any two logarithmically reduced spectrograms onto intervals of reduction $[a_1, ra_1)$ and $[a_2, ra_2)$ with $a_1 \neq a_2$ are equivalent up to logarithmic scaling and rotation. In the case $r = 2$ this proof shows why all chroma octaves bounded by both the upper and lower frequencies of the sound spectrum in question yield essentially the same continuous chromagram. We then propose a family of pseudometrics on sound spectra and discuss potential applications to analysis and composition.

Keywords: Continuous chromagram · Generalized octave reduction · Pseudometric spaces

1 Introduction

1.1 Pitch and Chroma

Although nowhere near as slippery a notion as timbre, pitch-relative to its centrality in theoretical discourse-can easily be taken for granted definitionally. Depending on who was asked and when, pitch has been variously defined as "that attribute of auditory sensation by which sounds are ordered on the scale used for melody in music" [1], "that attribute of sensation whose variation is associated with musical melodies" [2], "[that which] can be reliably matched by adjusting the frequency of a pure tone of arbitrary amplitude" [3], or "a subjective measure [of] relative highness or lowness" [4].

Whatever definition one prefers, pitch in the most abstract clearly merits a continuous rather than a discrete understanding. Shepard's [5] foundational results on the perceptual representation of pitch suggested that the pitch p of a sound can be decomposed into values of chroma c and tone height h: $p = 2^c \cdot 2^h$, where c is any real number $[0, 1)$ and h is any integer in order

for the decomposition to be unique. Patterson [6] later generalized Shepard's two-dimensional representation of pitch into a two-dimensional representation of frequency: $f = 2^c \cdot 2^h$, where once again c is any real number $[0, 1)$ and h is any integer in order for the decomposition to be unique. And recent neuroscience [4] suggests that such two-dimensional representations are compatible with the tonotopic organization of pitch in human auditory cortex.

Wakefield [7] presented an ad hoc method for producing chroma-time representations of sounds in the context of analyzing the resonance of an individual human singing voice before subsequently [8] presenting limited formalization of his ad hoc method. Building on the models of Shepard [5] and Patterson [6], in this paper we provide a non-ad hoc formalization of a continuous chromagram, an octave-reduced spectrogram that sums all amplitudes found in a sound spectrum onto a half-open interval $[a, 2a)$ we call a chroma octave. Working from any sound spectrum, our formalization generalizes the notion of octave reduction to any interval of logarithmic reduction and fosters the formulation of pseudo-metric spaces of sound spectra for measuring distances between musical sounds. After presenting illustrative technical aspects of our theoretical apparatus, we conclude with brief discussions of potential applications of the present work to analysis and composition as well as future directions for continued exploration of chroma perception.

1.2   Motivations

There are a number of reasons for the consideration of continuous chromagrams. Firstly, the 12-bin equal quantization of frequency underlying chromagrams (also known as chroma features) in most music information retrieval applications does a disservice to the reality of musical diversity. As musical inquiry more broadly begins to come to terms with how to expand its purview, continuous models of chroma provide a natural basis for consideration of musical traditions with different underlying pitch structures than twelve-tone equal partitions of 2:1 octaves.

Secondly, chroma-time representations naturally erase boundaries between timbre and pitch in the consideration of the aggregate quality of a sound. Computational models of chord spacing and chord quality may embrace the possibility of pitch percepts of spectral components other than those of fundamentals. Continuous chromagrams privilege only those octave equivalence classes with more amplitude than others. In light of pitch's longstanding status as a fertile field of inquiry and more recent explorations of timbre using multidimensional models, considering them holistically is attractive from the perspective of perceptually informed theorizing.

Thirdly, chroma-time representations represent a possibility for new and broader directions in the theoretical exploration of parsimony. Fundamental frequencies are obvious places to listen for pitch percepts but they are by no means the only frequencies that a listener might perceive as pitches, especially in harmonically rich polyphonic timbres. To this end, the continuous chromagram facilitates thinking about chromatic parsimony in the very literal sense of chroma and can provide new directions for analysis and composition of musical sounds.

2   Continuous Chromagrams
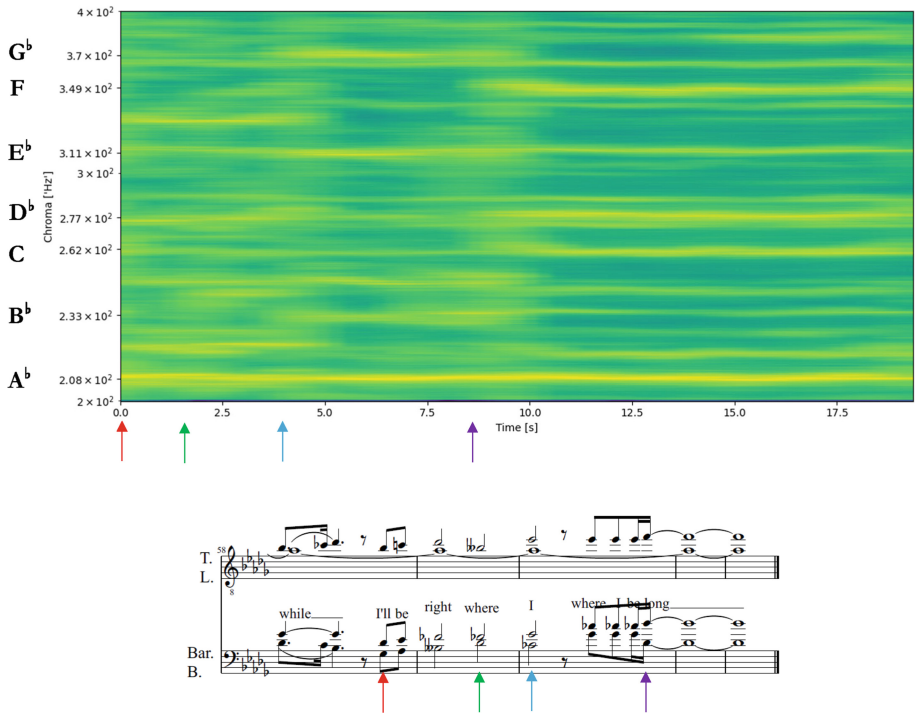
2.1   Illustration and Definition

Fig. 1. The ending of the tag of Vocal Spectrum's arrangement of "Go the Distance" as represented through a continuous chromagram and via staff notation. A recording may be accessed at https://youtu.be/DNmW4t22v0s?t=201. (Color figure online)

Figure 1 presents a representative example of a continuous chromagram contextualized by staff notation as well as the recording of the excerpt in question. This continuous chromagram presents the ending of the tag to "Go the Distance" from the 1997 Disney film *Hercules* as arranged by the international champion barbershop quartet Vocal Spectrum and is discussed further in Sect. 4.1. In the meantime, it will suffice to present foundational notions for defining and understanding continuous chromagrams in general.

Logarithmically Reduced Spectra. Let $n, N \in \mathbb{R}$ with $0 < n < N$ and let $f : [n, N] \to \mathbb{R}$ represent a sound spectrum. Choose a *reduction parameter* $r \in \mathbb{R}$ with $r > 1$ and suppose $a \in \mathbb{R}$ satisfies $n \leq a < ra \leq N$. We define the *logarithmically reduced spectrum* $g : [a, ra) \to \mathbb{R}$ by

$$g(x) := \sum_{\substack{\{y \in [n,N] \,:\, \exists k \in \mathbb{Z} \\ \text{s.t. } x = r^k y\}}} f(y), \qquad \text{for } x \in [a, ra), \tag{1}$$

where $[a, ra)$ is the *interval of reduction*. In particular, $[a, 2a)$ yields a spectrum reduced onto a *chroma octave*, while $[a, ra)$ for $r \approx 2$ reduces onto a *chroma pseudo-octave*. For $r = 2$ the logarithmically reduced spectrum may be said to be octave reduced in the conventional musical sense. If the chosen $r \neq 2$ we define the process of producing the associated logarithmically reduced spectrum as *generalized octave reduction*.

Logarithmically Reduced Spectrograms and Continuous Chromagrams. Implicit in the above formulation of logarithmically reduced spectra is the elimination of time variance. In practice, this would tend to stem from either only looking at one minuscule temporal window or taking the average of the spectrum over some time interval that ideally does not feature too many transients. In computational work on *a cappella* singing, for instance, one might well compute continuous chromagrams in the middle of sustained chords where vowels and formants are most stable. Incorporating time into our formulation of logarithmically reduced spectra is nonetheless worthwhile because of the number of applications which motivate working with transients and/or other phenomenon where time variance matters strongly.

Suppose $[0, T] \subset \mathbb{R}$ for some $T \in \mathbb{R}$ represents time. Let $f : [n, N] \times [0, T] \to \mathbb{R}$ represent a time variant sound spectrum. For a chosen reduction parameter r and interval of reduction $[a, ra)$ as defined above, we define the *logarithmically reduced spectrogram* $g : [a, ra) \times [0, T] \to \mathbb{R}$ by

$$g(x, t) := \sum_{\substack{\{y \in [n,N] \,:\, \exists k \in \mathbb{Z} \\ \text{s.t. } x = r^k y\}}} f(y, t), \qquad \text{for } x \in [a, ra), \quad t \in [0, T]. \tag{2}$$

For $r = 2$, g is a *continuous chromagram* providing a chroma-time representation of a sound analogous to the frequency-time representation of a conventional spectrogram but with the amplitudes of all frequencies sharing a chroma value summed onto a single representative in the chroma octave. Continuous in this context is therefore used to contrast with the conventional chromagram (or chroma feature), which can most accurately be described as a quantized chromagram since it reduces amplitudes into twelve logarthmically equal bins corresponding to the twelve chroma of twelve-tone equal temperament.

2.2   Invariance to Interval of Reduction

We next present a proof justifying why one's choice of interval of reduction in producing a logarithmically reduced spectrum does not essentially change the result. For quantized chromagrams the bins are almost always determined in relation to a fixed reference frequency such as A4 = 440 Hz using the convenience of multiplying by powers of $\sqrt[12]{2}$. In the case of logarithmically reduced spectra, on the other hand, there are no structural *a priori* assumptions that would motivate the choice of any particular interval of reduction absent any more specific context. One may nonetheless choose any appropriately bounded interval of reduction for a given sound spectrum and produce the same logarithmically reduced spectrum up to logarithmic scaling and rotation.

Theorem 1. *For a fixed reduction parameter $r > 1$, let g and h be logarithmically reduced spectra with intervals of reduction $[a_1, ra_1)$ and $[a_2, ra_2)$, respectively and with $a_1 \neq a_2$. Then g and h are equivalent up to logarithmic scaling and rotation.*

Proof. Without loss of generality, we will assume that $a_1 < a_2$. (Otherwise, swap the roles of g and h in what follows.)

We will consider the following two cases separately and prove the Theorem separately for each case:

1. Case 1: there exists some $m \in \mathbb{N}$ such that $\frac{a_2}{a_1} = r^m$; and
2. Case 2: there does *not* exist any $m \in \mathbb{N}$ such that $\frac{a_2}{a_1} = r^m$.

<u>Proof of the Theorem for Case 1</u>: Suppose that $\frac{a_2}{a_1} = r^m$ for some $m \in \mathbb{N}$. We claim that h can be written in terms of g as

$$h(x) = g(r^{-m}x) \quad \text{for } x \in [a_2, ra_2). \tag{3}$$

Before verifying Eq. 3, we need to check that the expression on the right-hand side of Eq. 3 is well-defined; that is, we need to check that $r^{-m}x \in [a_1, ra_1)$ if $x \in [a_2, ra_2)$, otherwise it would not make sense to apply g to $r^{-m}x$ as shown on the right-hand side of Eq. 3 (since g is defined only on $[a_1, ra_1)$ a priori).

Indeed, if $x \in [a_2, ra_2)$ then $a_2 \leq x$ and thus, using the equality $\frac{a_2}{a_1} = r^m$, we conclude $a_1 = r^{-m}a_2 \leq r^{-m}x$. Furthermore, the inclusion $x \in [a_2, ra_2)$ implies $x < ra_2$ and thus, using the equality $\frac{a_2}{a_1} = r^m$, we conclude $r^{-m}x < r^{-m}(ra_2) = r \cdot r^{-m}a_2 = ra_1$. Combining these results, we see that $a_1 \leq r^{-m}x < ra_1$. Since our choice of $x \in [a_2, ra_2)$ was arbitrary, we conclude that $r^{-m}x \in [a_1, ra_1)$ for all $x \in [a_2, ra_2)$ and hence that the right-hand side of Eq. 3 is indeed well-defined.

So, now we proceed to verify the equality in Eq. 3. We compute, for $x \in [a_2, ra_2)$, that

$$g(r^{-m}x) = \sum_{\substack{\{y \in [n,N]\,:\,\exists k \in \mathbb{Z} \\ \text{s.t. } r^{-m}x = r^k y\}}} f(y) = \sum_{\substack{\{y \in [n,N]\,:\,\exists k \in \mathbb{Z} \\ \text{s.t. } x = r^{k+m}y\}}} f(y) = \sum_{\substack{\{y \in [n,N]\,:\,\exists l \in \mathbb{Z} \\ \text{s.t. } x = r^l y\}}} f(y) = h(x).$$

Therefore, Eq. 3 indeed holds. Equation 3 states, in other words, that h is obtained from g by a logarithmic scaling (by r^{-k}). In this case, no rotation is necessary in order to obtain h from g.

<u>Proof of the Theorem for Case 2:</u> Suppose that there does *not* exist any $m \in \mathbb{N}$ with $\frac{a_2}{a_1} = r^m$. Let $m_0 := \min \left\{ m \in \mathbb{Z} : r^m > \frac{a_2}{a_1} \right\}$. That is, m_0 is the unique integer such that $r^{m_0} a_1 \in [a_2, ra_2)$. So, $a_2 \leq r^{m_0} a_1 < ra_2$. We claim that h can be written in terms of g as

$$h(x) = \begin{cases} g(r^{1-m_0} x) & \text{if } x \in [a_2, r^{m_0} a_1) \\ g(r^{-m_0} x) & \text{if } x \in [r^{m_0} a_1, ra_2). \end{cases} \tag{4}$$

Before verifying Eq. 4, we need to check that the expressions on the right-hand side of Eq. 4 are well-defined; that is, we need to check that $r^{1-m_0} x \in [a_1, ra_1)$ if $x \in [a_2, r^{m_0} a_1)$ and that $r^{-m_0} x \in [a_1, ra_1)$ if $x \in [r^{m_0} a_1, ra_2)$, otherwise it would not make sense to apply g to those values as shown on the right-hand side of Eq. 4 (since g is defined *a priori* only on $[a_1, ra_1)$):

- First suppose that $x \in [a_2, r^{m_0} a_1)$. Then $a_2 \leq x$ and thus, using the inequality $r^{m_0} a_1 < ra_2$, we conclude $a_1 = r^{-m_0}(r^{m_0} a_1) < r^{-m_0} ra_2 = r^{1-m_0} a_2 \leq r^{1-m_0} x$. Furthermore, the inclusion $x \in [a_2, r^{m_0} a_1)$ implies $x < r^{m_0} a_1$ and thus $r^{1-m_0} x < r^{1-m_0} r^{m_0} a_1 = ra_1$. Combining these results, we conclude that $a_1 \leq r^{1-m_0} x < ra_1$, i.e., $r^{1-m_0} x \in [a_1, ra_1)$.
- Next, suppose that $x \in [r^{m_0} a_1, ra_2)$. Then $r^{m_0} a_1 \leq x$ and thus $a_1 = r^{-m_0}(r^{m_0} a_1) \leq r^{-m_0} x$. Furthermore, the inclusion $x \in [r^{m_0} a_1, ra_2)$ implies $x < ra_2$ and thus, using the inequality $a_2 \leq r^{m_0} a_1$, we conclude $r^{-m_0} x < r^{-m_0} ra_2 = r^{1-m_0} a_2 \leq r^{1-m_0}(r^{m_0} a_1) = ra_1$. Combining these results, we conclude that $a_1 \leq r^{-m_0} x < ra_1$, i.e., $r^{-m_0} x \in [a_1, ra_1)$.

Now that we have shown that the right-hand side of Eq. 4 is well-defined, we now proceed to verify the equality. We first verify it for $x \in [a_2, r^{m_0} a_1)$ and then verify it for $x \in [r^{m_0} a_1, ra_2)$.

If $x \in [a_2, r^{m_0} a_1)$, then

$$g(r^{1-m_0} x) = \sum_{\substack{\{y \in [n,N] : \exists k \in \mathbb{Z} \\ \text{s.t. } r^{1-m_0} x = r^k y\}}} f(y) = \sum_{\substack{\{y \in [n,N] : \exists k \in \mathbb{Z} \\ \text{s.t. } x = r^{k+m_0-1} y\}}} f(y) = \sum_{\substack{\{y \in [n,N] : \exists l \in \mathbb{Z} \\ \text{s.t. } x = r^l y\}}} f(y) = h(x),$$

and so Eq. 4 holds for $x \in [a_2, r^{m_0} a_1)$. On the other hand, if $x \in [r^{m_0} a_1, ra_2)$, then

$$g(r^{-m_0} x) = \sum_{\substack{\{y \in [n,N] : \exists k \in \mathbb{Z} \\ \text{s.t. } r^{-m_0} x = r^k y\}}} f(y) = \sum_{\substack{\{y \in [n,N] : \exists k \in \mathbb{Z} \\ \text{s.t. } x = r^{k+m_0} y\}}} f(y) = \sum_{\substack{\{y \in [n,N] : \exists l \in \mathbb{Z} \\ \text{s.t. } x = r^l y\}}} f(y) = h(x),$$

and so Eq. 4 holds for $x \in [r^{m_0} a_1, ra_2)$.

Therefore, Eq. 4 indeed holds for all $x \in [a_2, ra_2)$. Equation 4 states, in other words, that h is obtained from g by a logarithmic scaling of, separately, $[a_2, r^{m_0} a_1)$ by r^{1-m_0} and of $[r^{m_0} a_1, ra_2)$ by r^{-m_0}. This can be equivalently

thought of as a rotation of $[r^{m_0}a_1, ra_2) \subset [a_2, ra_2)$ to be immediately to the left of $[a_2, r^{m_0}a_1)$ while scaled logarithmically followed by a logarithmic scaling of the new full interval down until it precisely overlays $[a_1, ra_1)$.   □

For the reduction parameter $r = 2$, this proof justifies why any two chroma octaves bounded by both the upper and lower frequencies of the sound spectrum in question yield essentially the same continuous chromagram. Additionally, the proof's independence of reduction parameter r demonstrates how our notion of a continuous chromagram can be extended to intervals other than the octave in the process we defined as generalized octave reduction. Consider for instance a recording of monophonic music composed using the Bohlen-Pierce scale (BP scale) which repeats at the tritave (the frequency ratio 3:1) rather than the 2:1 octave and performed by a timbre such as the clarinet whose spectrum consists overwhelmingly of odd harmonics outside of higher registers. A continuous chromagram of such a sound signal onto a chroma octave such as $[200, 400)$ Hz could be arguably unhelpful at best in understanding the music relative to its compositional structure. Choosing a chroma tritave of $[200, 600)$ Hz instead would make for a more parsimonious mapping between the representation and the compositional structure. Logarithmically reduced spectrograms additionally offer new possibilities for investigating musical sounds with inharmonic timbres. Indeed, generalized octave reduction of sound spectra onto intervals of reduction $[a, ra)$ with $r \neq 2$ would represent a logical choice for exploring sounds such as chords of tones with stretched or compressed partials produced by Slaymaker [9] and Mathews [10] as well as the consonant nonoctaves used as illustrations by Sethares [11]. As the above proof has shown, such generalized octave reductions may also pick any interval of reduction within the bounds of the sound spectrum in question and obtain the same result up to logarithmic scaling and rotation.

2.3   Computational Implementations

A tension between the terms continuous and chromagram is inevitable in signal processing applications which necessarily entail sampling. Continuous may thus be understood in computational implementations as referring to the highest resolution possible via one's spectral analysis of a given signal. In a context of either continuous chromagrams or generalized octave reduction, one may observe two limitations of the fast Fourier transform (FFT) for spectral analysis. Because its linear frequency bins result in logarithmic increases in resolution in higher octaves, the logarithmic reduction of fibers to their image in the chroma octave (or other interval) sums amplitudes unevenly. The presence of a zero-frequency/DC term in the spectrum violates one of the hypotheses of our previous theorem (namely that the spectrum is defined on $[n, N]$ for $n > 0$). A Python implementation of a continuous chromagram generated via the FFT (and thus bearing the two aforementioned limitations) is freely available on Github [12]. For future work, the constant-Q transform-oft-maligned for its lack of total invertibility although extendible to least-squares invertibility [13]-may address both of these limitations at once.

3   Pseudometric Spaces of Sound Spectra

One application of continuous chromagrams and generalized octave reduction is to measuring the similarity or dissimilarity of musical sounds. If comparisons based on octave reduction in the usual sense measure distances between chords in a notes-on-a-page sense, then comparisons based on continuous chromagrams instead measure distances along the logarithmically reduced chroma helix [4]. We present as an illustration a family of pseudometrics arising from the theory of L^p, metric, and pseudometric spaces [14,15] on sets of sound spectra and then observe how it can be used to derive metric spaces of sound spectra. This family is by no means the only potentially fruitful source of pseudometrics for comparing musical sounds. Among myriad options, one could for instance imagine adapting earth mover's distance-based similarity analysis (EMDSA) [16] to pseudometric comparisons of sound spectra.

3.1 L^p Pseudometrics

Suppose $n, N \in \mathbb{R}$ with $0 < n < N$. Let $r \in \mathbb{R}$ satisfy $r > 1$, and suppose $a \in \mathbb{R}$ satisfies $n \leq a < ra \leq N$. For a function $f : [n, N] \to \mathbb{R}$, define its logarithmically reduced function $\hat{f} : [a, ra) \to \mathbb{R}$ by

$$\hat{f}(x) := \sum_{\substack{\{y \in [n,N] : \exists k \in \mathbb{Z} \\ \text{s.t. } x = r^k y\}}} f(y), \qquad \text{for } x \in [a, ra).$$

L^p pseudometrics with $p \in [1, \infty)$. Let $p \in [1, \infty)$.
The L^p norm $|| \cdot ||_{L^p([a,ra))}$ on the vector space

$$L^p([a, ra)) := \left\{ g : [a, ra) \to \mathbb{R} \,|\, g \text{ is Lebesgue-measurable and } \int_a^{ra} |g(x)|^p \, dx < \infty \right\},$$

where two functions which are equal almost everywhere are considered to define the same element of $L^p([a, ra))$, is defined by

$$||g||_{L^p([a,ra))} := \left(\int_a^{ra} |g(x)|^p \, dx \right)^{\frac{1}{p}}. \tag{5}$$

The metric $d_{L^p([a,ra))}$ on $L^p([a, ra))$ induced from $|| \cdot ||_{L^p([a,ra))}$ is

$$d_{L^p([a,ra))}(g_1, g_2) := ||g_1 - g_2||_{L^p([a,ra))} \tag{6}$$

for $g_1, g_2 \in L^p([a, ra))$. From Eq. 6 we may define on the set $\widetilde{S}^p_{[a,ra)}([n, N]) := \left\{ f : [n, N] \to \mathbb{R} \,|\, \hat{f} \in L^p([a, ra)) \right\}$ an associated pseudometric $\tilde{d}_{\widetilde{S}^p_{[a,ra)}([n,N])}$ by

$$\tilde{d}_{\widetilde{S}^p_{[a,ra)}([n,N])}(f_1, f_2) := d_{L^p([a,ra))}(\hat{f}_1, \hat{f}_2) \tag{7}$$

for $f_1, f_2 \in \widetilde{S}^p_{[a,ra)}\big([n, N]\big)$; we refer to $\tilde{d}_{\widetilde{S}^p_{[a,ra)}([n,N])}$ as an L^p pseudometric.

L^p pseudometric with $p = \infty$. One may obtain for $p = \infty$ a pseudometric $\tilde{d}_{\widetilde{S}^\infty_{[a,ra)}([n,N])}$ in the same manner as for the above L^p pseudometrics by replacing the instances of the integral $\int_a^{ra} |g(x)|^p dx$ with the essential supremum

$$\operatorname{ess\,sup}_{x \in [a,ra)} |g(x)|$$

and leaving the rest of the formulation essentially unchanged [14, 15, 17].

3.2   Metrics Induced from Pseudometrics

If $\tilde{X}$ is any set and $\tilde{d}$ is a pseudometric on $\tilde{X}$, one can canonically obtain a corresponding metric space as follows [15, 18]. Put the equivalence relation $\sim$ on $\tilde{X}$ given by $x \sim y$ if $\tilde{d}(x, y) = 0$. If one lets X be the set of equivalence classes of $\tilde{X}$ under this equivalence relation and defines $d([x], [y]) := \tilde{d}(x, y)$ for the equivalence classes $[x]$ and $[y]$ of x and y respectively, then d is a well-defined metric on X, and thus (X, d) is a metric space. This may be applied, for each $p \in [1, \infty]$, to the pseudometric space $\Big(\widetilde{S}^p_{[a,ra)}\big([n, N]\big), \tilde{d}_{\widetilde{S}^p_{[a,ra)}([n,N])}\Big)$, with the aforementioned equivalence relation $\sim$, to produce a metric on the set

$$S^p_{[a,ra)}\big([n, N]\big) := \left(\Big\{ f : [n, N] \to \mathbb{R} \,|\, \hat{f} \in L^p\big([a, ra)\big) \Big\} \Big/ \sim \right),$$

i.e., on the set of spectra being reduced into $L^p\big([a, ra)\big)$ wherein two spectra are considered to be equivalent if their logarithmically reduced spectra are equal almost everywhere on $[a, ra)$.

3.3   Comparing Time Variant Spectra

Pseudometric measures of similarity or dissimilarity can be extended analogously to our generalized octave reduction apparatus to account for time variance as follows. If we consider logarithmically reduced spectrograms of the form $f : [n, N] \times [0, T] \to \mathbb{R}$, then analogously to above, for $p \in [1, \infty)$ one may yield metrics and pseudometrics by replacing the instances of the integral $\int_a^{ra} |g(x)|^p dx$ with

$$\int_0^T \int_a^{ra} |g(x, t)|^p \, dx \, dt$$

and leaving the rest of the formulation essentially unchanged.

4   Applications

4.1   Visualizations of Musical Sounds

Returning to Fig. 1, the color-coded arrows provide a temporal reference while the equal-tempered letter names are for reference within the chroma octave

[200, 400) 'Hz'. The ordinate is labeled as 'Hz' because these are not frequencies corresponding to single pitches but rather representatives of their chroma within that chroma octave. This continuous chromagram visualization highlights the salience of several prominent spectral components beyond the four sung fundamental frequencies. These components stem from the combination of formant spread and formant tuning [19] compounded with precise vertical intonation to maximize the production of ring [20].

This illustration demonstrates how continuous chromagrams can facilitate new framings of parsimonious voice-leading. 'Above' the lead's sustained A♭, for instance, we can observe a veritable thicket of chroma across which it would not be unreasonable to trace smooth voice-leading trajectories. The perceptual implications of such trajectories are not to be overstated until they might be validated by ecologically valid experiments on listening. In the meantime, if we listen again we can find that this chroma-time representation can not only represent such phenomena but also (and perhaps even more importantly) facilitate new ways of listening.

4.2 Creative and Theoretical Possibilities

Continuous chromagrams, generalized octave reduction, and the various pseudometrics we have presented also have further, more speculative applications. Among these are creative sonic practices and possibilities for future theorizing. One elegant formulation of a compositional constraint could easily stem from the fact that there are infinitely many sounds in the fiber of any logarithmically reduced spectrum.

Remark 1. Suppose that we are given a function $g : [a, ra) \to \mathbb{R}$. Let us construct a function $f : [n, N] \to \mathbb{R}$ such that g is the logarithmically reduced version of f, i.e.,

$$\sum_{\substack{\{y \in [n,N] : \exists k \in \mathbb{Z} \\ \text{s.t. } x = r^k y\}}} f(y) = g(x), \qquad \text{for } x \in [a, ra). \tag{8}$$

For each $x \in [n, N]$, let $\tilde{x}$ denote the unique number in $[a, ra)$ such that $x = 2^k \tilde{x}$ for some $k \in \mathbb{Z}$. Then,

$$f(x) := \frac{g(\tilde{x})}{|\{y \in [n, N] : \exists k \in \mathbb{Z} \text{ s.t. } x = r^k y\}|} \tag{9}$$

is one such function, where the vertical bars denote cardinality.

Although it is by construction impossible to regain the information lost through the process of logarithmically reducing a sound spectrum, this does not mean that one could not generate sounds from a logarithmically reduced spectrogram. From a musical perspective, a sort of chromatic (in the sense of chroma) minimalism could stem from treating none of pitch, timbre or rhythm alone as a basis for structural repetition and instead repeating the entire image

of fibers within the chroma octave of a continuous chromagram. Another possible compositional constraint derived from logarithmically reduced spectra would be to limit oneself to sounds all equidistant under a given metric or pseudometric, giving at once many more possibilities than similar constraints on dots-on-a-page representations of musical similarities while also posing substantial limitations on the timbral possibilities for the sounds in question. In entirely acoustic contexts such constraints are probably more readily posited than realized. Nonetheless, through the use of digital tools speculative applications of chromatic (again in the sense of chroma) thinking might well prove musically rewarding.

One additional compositional approach motivated by our exploration of generalized octave reduction is in following with Slaymaker [9], Mathews [10], and Sethares [11] in designing timbres based on non-octave intervals. If we, for instance, sought to produce a given number of discrete amplitudes in an $\left[a, \frac{7}{4}a\right)$ interval of reduction through purely additive synthesis in just intonation, we would necessarily want to think differently both in terms of relationships between spectral components and in terms of where they lie to begin with than if we were reducing to the chroma octave $[a, 2a)$. In a similar vein, composers interested in historical keyboard temperaments might analogously consider limitations imposed on an $\left[a, \sqrt{5}a\right)$ interval of reduction if they wished to partake in contemporary meantone composition.

Finally, another theoretical possibility would be to use logarithmically reduced spectra to quantitatively characterize spectral aggregates. Recent work by Kahrs [21] on the music of Gubaidulina explored "dissonance" in relationship to distance from white noise using the relatively common music information retrieval measures of spectral centroid and spectral flatness. One can imagine how entropy-related or otherwise probabilistic measurements of sound signals derived from logarithmically reduced spectra could extend these notions of distance from noise as possibilities for both analysis and composition. It is our hope that these illustrations of the possibilities afforded by logarithmically reduced spectra and logarithmically reduced spectrograms will begin a conversation that sheds light on further future prospects.

References

1. Pitch. https://asastandards.org/terms/pitch/. Accessed 8 Jan 2022
2. Plack, C.J., Oxenham, A.J.: Overview: the present and future of pitch. In: Plack, C.J., Fay, R.R., Oxenham, A.J., Popper, A.N. (eds.) Pitch. SHAR, vol. 24, pp. 1–6. Springer, New York (2005). https://doi.org/10.1007/0-387-28958-5_1
3. Hartmann, W.: Signals, Sound, and Sensation. AIP Press, Woodbury (1997)
4. Langner, G.: The Neural Code of Pitch and Harmony. Cambridge University Press, Cambridge (2015). https://doi.org/10.1017/CBO9781139050852
5. Shepard, R.: Circularity in judgements of relative pitch. J. Acoust. Soc. Am **36**(12), 2346–2353 (1964). https://doi.org/10.1121/1.1919362
6. Patterson, R.: Spiral detection of periodicity and the spiral form of musical scales. Psychol. Music **14**, 44–61 (1986). https://doi.org/10.1177/0305735686141004

7. Wakefield, G.: Chromagram visualization of the singing voice. In: Proceedings of the International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications, MAVEBA 1999, pp. 24–29 (1999)
8. Wakefield, G.: Mathematical representation of joint chroma-time distributions. In: Proceedings of the 9th SPIE Conference on Advanced Signal Processing Algorithms, Architectures, and Implementations, pp. 637–645 (1999)
9. Slaymaker, F.H.: Chords from tones having stretched partials. J. Acoust. Soc. Am. **47**(2), 1569–1571 (1970). https://doi.org/10.1121/1.1912089
10. Mathews, M., Roads, C.: Interview with Max Mathews. Comput. Music J. **4**(4), 15–22 (1980). https://doi.org/10.2307/3679463
11. Sethares, W.: Tuning, Timbre, Spectrum, Scale. Springer, London (2005). https://doi.org/10.1007/b138848
12. Lenchitz, J.: Continuous Chromagram. https://github.com/jordan-lenchitz/continuous-chromagram. Accessed 14 Mar 2022
13. Ingle, A., Sethares, W.: The least-squares invertible constant-Q spectrogram and its application to phase vocoding. J. Acoust. Soc. Am. **132**(2), 894–903 (2012). https://doi.org/10.1121/1.4731466
14. Folland, G.: Real Analysis, 2nd edn. Wiley, New York (1999)
15. Howes, N.R.: Modern Analysis and Topology. Springer, New York (1995). https://doi.org/10.1007/978-1-4612-0833-4
16. Ruzon, M., Tomsasi, C.: Edge, junction, and corner detection using color distributions. IEEE Trans. Pattern Anal. Mach. Intell. **23**(11), 1281–1295 (2001). https://doi.org/10.1109/34.969118
17. Evans, L.: Partial Differential Equations, 2nd edn. American Mathematical Society, Providence (2010)
18. Simon, B.: A Comprehensive Course in Analysis. American Mathematical Society, Providence (2015)
19. Kalin, G.: Formant frequency adjustment in barbershop quartet singing. Master's thesis, KTH Royal Institute of Technology, Stockholm, Sweden (2005)
20. Averill, G.: Bell tones and ringing chords: sense and sensation in barbershop harmony. World Music **41**(1), 37–51 (1999)
21. Kahrs, N.: Consonance, dissonance, and formal proportions in two works by Sofia Gubaidulina. Music Theory Online **26**(2) (2020). https://doi.org/10.30535/mto.26.2.7

N2D3P9

Dave Keenan[1][✉] and Douglas Blumeyer[2][✉]

[1] Brisbane, Australia
d.keenan7@gmail.com
[2] San Francisco, USA
douglas.blumeyer@gmail.com
https://dkeenan.com, https://douglasblumeyer.com

Abstract. N2D3P9 is a mathematical function which was developed to help in designing the Sagittal microtonal music notation. Given a rational number $\frac{n}{d}$ representing a pitch (relative to some tonic note), N2D3P9 estimates its rank in popularity among all rational pitches in musical use. A low value of N2D3P9 indicates that the ratio is used often, and so should have a simple accidental symbol, while a high value indicates that the ratio is used rarely and so can have a more complex symbol if necessary. It may also be useful in designing rational scales or tunings.

Keywords: Microtonal · Just intonation · Sagittal notation

1 Background on Sagittal Notation

Outside the hegemony of standard musical tuning, a huge frontier of alternative tuning systems is rife for exploration. For however many musicians one may find who are interested in microtonality, one may find as many (if not more) alternative tunings, which boast a wide variety of strengths, weaknesses, and special characteristics. Some of these tunings deviate from standard tuning in diametrically opposed manners. And while each tuning offers its own unique wealth of potential for new musical expression, the flip side of this coin is that each tuning also poses a unique challenge when it comes to notating it.

Historically, the microtonal community has been defined by fracturing, on multiple levels, from top to bottom, from ethos to practice. A family of deviants deviates even from itself. This effect had held true with regards to problems of notation, where we've seen a proliferation of tuning-specific notations. Sometimes a more popular tuning may even boast multiple of these custom-tailored notations, locked in to separate particular ways of interpreting the tuning. This sort of fracturing poses a problem for the breed of microtonalist who is interested in writing and/or performing music that samples from across the full spectrum of possibilities. Learning microtonal concepts, assembling microtonal instruments, and mastering microtonal ear-training is trouble enough, without also having to learn many different notation systems.

This problem is not easily solved however. It wasn't until the early 2000's that the two-person team of George Secor and Dave Keenan, with input from

many others on the Yahoo Tuning group, began chipping away at the myriad of math and design challenges required to produce a single, unified notation system that was both powerful and flexible enough to work well for any microtonal tuning. This collaboration blossomed into what we now know as the **Sagittal microtonal notation system** [1].

Sagittal is the world's first pitch notation system to be designed from the ground-up to extend conventional notation in a way that works well for all types of microtonal music, and in particular the two most popular types: **just intonation (JI)**, and **equal divisions of the octave (EDOs)**. Of course, the latter category includes equal divisions of the octave other than the standard 12. Sagittal introduces a cohesive set of new symbols, each representing a differently sized pitch alteration, and thereby extends the conventional pitch notation beyond what can be accomplished using only the Pythagorean nominals A, B, C, D, E, F, and G, plus the sharp and flat symbols.

Sagittal's central design begins with the assumption that musical pitch is interpreted as ratios of frequency. In the case in JI, these ratios are exact, and in EDOs and other tunings, they are approximate, but Sagittal works just as well either way.

Sagittal then returns to the historical roots of conventional notation by treating the sequence of Pythagorean nominals plus sharps and flats as the notation for a chain of perfect fifths, i.e. the frequency ratio $\frac{3}{2}$ (or its approximation). And octave shifts work as normal. This conventionally-notated chain of fifths thereby forms the backbone for Sagittal notation, allowing for the notation of any 3-limit pitch, or in other words, any pitch whose ratio contains only prime factors of 2 and 3, such as $\frac{4}{3}$, $\frac{9}{8}$, $\frac{27}{16}$, $\frac{128}{81}$, etc. For any other pitch, that is, one with prime factors above 3, such as 5, 7, or 11, a Sagittal symbol is available which represents the pitch alteration necessary to reach from one of these 3-limit pitches to a nearby higher-limit pitch.

For example, there is a Sagittal symbol which enables the notation of pitches with a single 5 in their prime factorization; using it, the pitch $\frac{10}{9}$ is notated as a deviation from $\frac{9}{8}$, or $\frac{5}{3}$ from $\frac{27}{16}$, or $\frac{9}{5}$ from $\frac{16}{9}$. In all three of these cases, the pitch alteration that the Sagittal symbol represents is the same small frequency ratio of $\frac{81}{80}$ (please check this for yourself, if you are unfamiliar with these relationships). Small ratios of this sort are known in the microtonal community as **commas**.

The $\frac{81}{80}$ comma used in the previous examples to notate pitches with single 5's in their ratios is quite important, and is well-known—even outside the microtonal community—as the syntonic comma. There is another Sagittal symbol, then, that enables of notation of pitches with a single 7 in their ratio, e.g. allowing for the notation of $\frac{8}{7}$ as deviation from $\frac{9}{8}$, etc. This comma's ratio is $\frac{64}{63}$, and this one is also reasonably well-known. The reason these two commas are well-known is due to their importance: they allow for the notation of the most popular pitches beyond the 3-limit backbone of conventional notation.

While there are yet other Sagittal symbols that allow for the notation of various combinations of 5's and 7's, such as $\frac{10}{7}$, $\frac{25}{24}$, or $\frac{35}{32}$, and there are Sagittal symbols that allow for notating pitches with 11's in them, or various combinations of 11's with 5's and 7's, as the prime numbers themselves get higher and

higher, and the count of these higher primes increases, the interest in musical pitches decreases, and with it goes the demand for symbols to notate them. In other words, almost every microtonalist needs to be able to notate pitches like $\frac{5}{4}$ and $\frac{7}{6}$, but almost no one writes or performs music with ratios like $\frac{37}{23}$ or $\frac{729}{625}$. So, Sagittal offers symbols that notate the former pitches exactly, while the latter can be only be notated using (very close) approximations.

Any microtonal pitch system like Sagittal, in order to be well-crafted, should offer an economical set of symbols to its users: the minimum set able to work cohesively to notate the most desired pitches in microtonal music. In the early days of Sagittal's design, the choice of which symbols to include and which not to include was informed by a combination of usage data, mathematical formulae, and first-hand knowledge from years of experience in the microtonal community. But it was not until the design for the final tier of Sagittal symbols was having its commas assigned, i.e. providing Sagittal symbols to satisfy the most obscure and microtonal of the microtonal community's needs, that it was decided that the best way forward was to consolidate the data, formulae, and knowledge into a proprietary semi-objective metric. This metric would compare every pitch ratio, classified by its prime factorization with the factors of 2 and 3 removed—those two simplest primes already being covered by conventional notation, as discussed above—and it would rank them by popularity, to determine which were the most important to provide notational symbols for. This metric is what became **N2D3P9**.

2   A Preliminary Pop Culture Reference

N2D3P9, or Entoo-Deethree-Peenine, is a fictional character in the Star Wars franchise. In an alternative timeline, the young Anakin Skywalker assembles the droid N2D3P9 from the parts of three other droids: R**2D**2, C**3P**0 and **N**R-N9**9**.

We're only joking, but we hope this helps with remembering and pronouncing the name. In actuality, the name "N2D3P9" is an abbreviation of key components of its formula, as described below.

3   Formula

Before describing how to calculate N2D3P9, we define three simpler terms that are used in its formula:

1. **2, 3-free** ratios, which are also known as "5-rough" ratios. Because factors of 2 and 3 in pitch ratios are already notated by changing octaves or moving along the chain of fifths (... B𝄫 F♭ C♭ G♭ D♭ A♭ E♭ B♭ F C G D A E B F♯ C♯ G♯ D♯ A♯ E♯ B♯ F𝄪 ...), N2D3P9 only operates on ratios that have had their factors of 2 and 3 removed. For example, there are various numbers of factors of 2 and 3 in the following ratios: $\frac{16}{15}, \frac{10}{9}, \frac{6}{5}, \frac{5}{4}, \frac{27}{20}, \frac{45}{32}, \frac{64}{45}, \frac{40}{27}, \frac{8}{5}, \frac{5}{3}, \frac{9}{5}, \frac{15}{8}$, but when their factors of 2 and 3 are removed, they all reduce to $\frac{1}{5}$ or $\frac{5}{1}$, and so they can all be notated using the same microtonal accidental, pointing

either up or down, combined with different letters and sharps or flats. We say that $\frac{1}{5}$ or $\frac{5}{1}$ is the **2, 3-removed** or **2, 3-free** form of these pitch ratios, and because $\frac{1}{5}$ and $\frac{5}{1}$ use the same accidental pointing either up or down, and because N2D3P9 only operates on ratios whose numerator is larger than their denominator (superunison ratios), $\frac{5}{1}$ can represent this entire **2, 3-equivalent pitch ratio class** or **2, 3-equivalence-class** for the purpose of notation design.

2. The **copfr** function, which stands for "**co**unt **of p**rime **f**actors with **r**epeats". It applies to any positive integer. For example 175 has the prime factorization $5 \times 5 \times 7$, which has three factors including the repeat of 5, so $\mathrm{copfr}(175) = 3$. Conventionally, $\mathrm{copfr}(1) = 0$. copfr is also called the "prime omega" or "big omega" function, Ω.

3. The **prime-limit** function, which is also known as **gpf**, which stands for "**g**reatest **p**rime **f**actor". $\mathrm{prime\text{-}limit}(175) = 7$. Some authors leave $\mathrm{prime\text{-}limit}(1)$ undefined; we avoid the question because we define $\mathrm{N2D3P9}(\frac{1}{1}) \equiv \mathrm{N2D3P9}(\frac{3}{1}) = 1$. This is because the ratios in the equivalence class represented by the 2, 3-removed $\frac{1}{1}$ actually have a prime limit of 3.

Now we can give the formula for N2D3P9$(\frac{n}{d})$ as:

$$\mathrm{N2D3P9}(\frac{n}{d}) = \frac{n}{2^{\mathrm{copfr}(n)}} \times \frac{d}{3^{\mathrm{copfr}(d)}} \times \frac{\mathrm{prime\text{-}limit}(nd)}{9} \tag{1}$$

$$\mathrm{N2D3P9}(\frac{1}{1}) = 1. \tag{2}$$

Note that where

$$n = 5^{n_5} \times 7^{n_7} \times 11^{n_{11}} \times ... \tag{3}$$

we have

$$2^{\mathrm{copfr}(n)} = 2^{n_5} \times 2^{n_7} \times 2^{n_{11}} \times ... \tag{4}$$

and so

$$\frac{n}{2^{\mathrm{copfr}(n)}} = (\tfrac{5}{2})^{n_5} \times (\tfrac{7}{2})^{n_7} \times (\tfrac{11}{2})^{n_{11}} \times ... \tag{5}$$

and similarly

$$\frac{d}{3^{\mathrm{copfr}(d)}} = (\tfrac{5}{3})^{d_5} \times (\tfrac{7}{3})^{d_7} \times (\tfrac{11}{3})^{d_{11}} \times ... \tag{6}$$

These can be described respectively as "product of half prime factors of the numerator (with repeats)" and "product of one-third prime factors of the denominator (with repeats)". So we can describe the procedure for calculating N2D3P9$(\frac{n}{d})$ as:

Take the prime factorization of the numerator and divide all the primes by 2, then multiply it out again. Do the same with the denominator but divide the primes by 3 instead of 2. Multiply these two results together then multiply by the prime limit of the ratio and divide by 9.

N2D3P9 can also be written more succinctly as:

$$\text{N2D3P9}(\frac{n}{d}) = \frac{nd \cdot \text{gpf}(nd)}{2^{\Omega(n)}3^{\Omega(d)+2}} \tag{7}$$

where nd is established in music theory as a ratio's "product complexity" or Benedetti height.

The division by 9 does not affect the ranking, but it has the convenient effect that N2D3P9 values are almost the same as the ranks they produce when applied to all 2, 3-free superunison ratios. Putting it another way, there are approximately N 2, 3-free pitch ratios with N2D3P9 $\leq N$.

For example, N2D3P9$(\frac{77}{5}) = \frac{7}{2} \times \frac{11}{2} \times \frac{5}{3} \times \frac{11}{9} \approx 39$, suggesting there are approximately 38 other 2, 3-free pitch ratios more popular than $\frac{77}{5}$. There are actually about 4% fewer than that on average. In this case there are 36.

4   Justification

Why should we believe that N2D3P9 accurately ranks the popularity of 2, 3-equivalent pitch classes?

N2D3P9 was developed (or discovered) rather late in the development of Sagittal notation. The Sagittal designers previously relied on actual ratio usage data from the Huygens-Fokker Foundation's scale archive [2], kindly provided by Manuel Op de Coul.

All scales in the archive were treated equally, as there was no information about their relative importance. Each occurrence of a pitch ratio in a scale was counted as one vote for that ratio. Then the ratios were grouped into 2, 3-equivalent pitch classes and a single figure obtained for each 2, 3-free superunison ratio (representing the class). There were 29,403 votes, allocated to 820 2, 3-free ratios.

Like the frequency of use of letters in an alphabet, when sorted in order of decreasing popularity, the ratios obeyed an approximate Zipf's law distribution, with the Nth most popular ratio having votes proportional to approximately $N^{-1.37}$. This meant that about half the ratios had only one vote each, and three quarters of them had 3 votes or less. Such low numbers of votes meant that the data on the less popular ratios was vulnerable to "historical noise". In other words, the position of such a ratio in the list might not be a good predictor of its relative frequency of use in the future.

In the early stages of Sagittal design, when allocating symbols for the most popular ratios, the designers could rely on the Scala archive data, but when they moved on to less popular ratios they needed some "less noisy" way to rank them.

Keenan and Blumeyer found that N2D3P9 is a psychoacoustically plausible function of a ratio's prime factorization that:

a. ranks 10 of the 11 most popular ratio classes in exactly the same way as the archive data, and
b. ranks all 820 ratio classes in a way that has a low sum of squared errors in their ranks, relative to the archive data, and

c. is sufficiently simple, having only two parameters, that it cannot be overfitting the data, and should therefore serve to average out the historical noise in the ranking of the less popular ratios, including ratios that do not occur in the archive at all.

However, their approach was not able to consider "all" possible psychoacoustic reasons for a ratio's popularity. For example, N2D3P9 does not evaluate whether some member of a 2, 3-equivalence-class might be very close in pitch to some member of another 2, 3-equivalence-class, such as $\frac{65}{64}$ being very close to $\frac{1}{1}$ (Figs. 1 and 2).
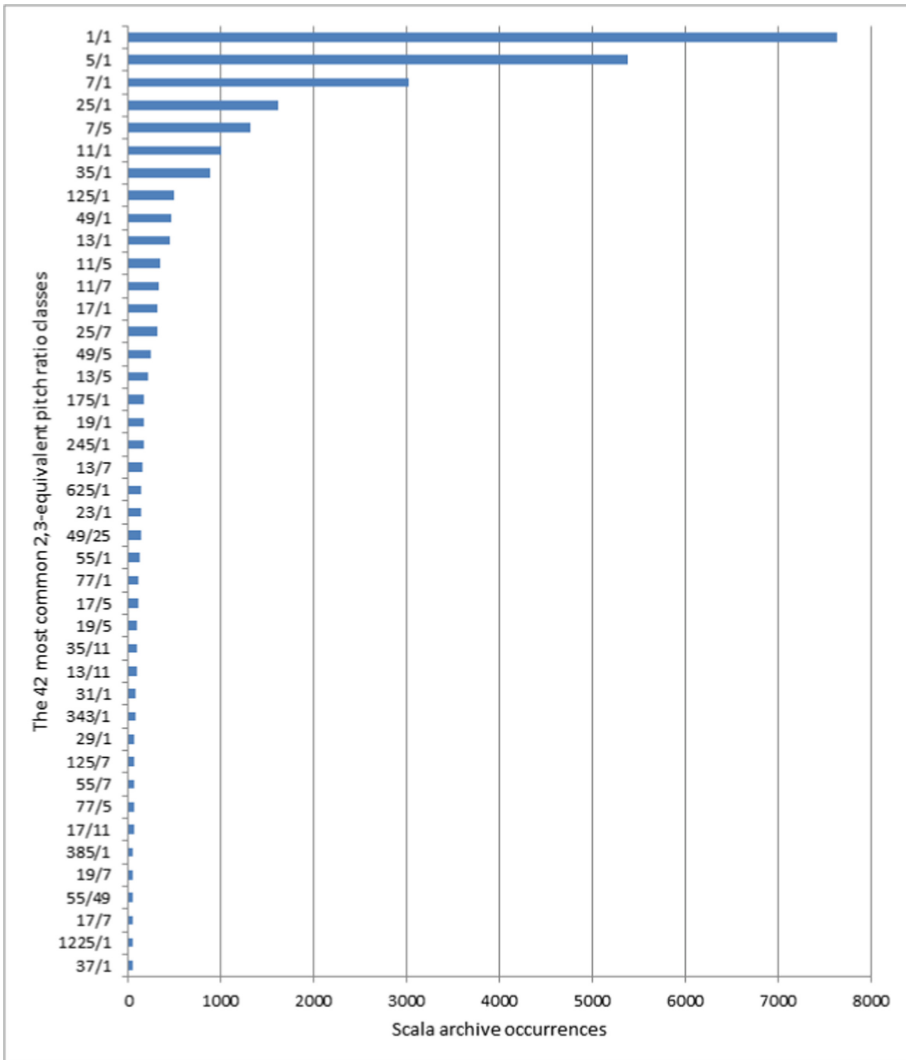
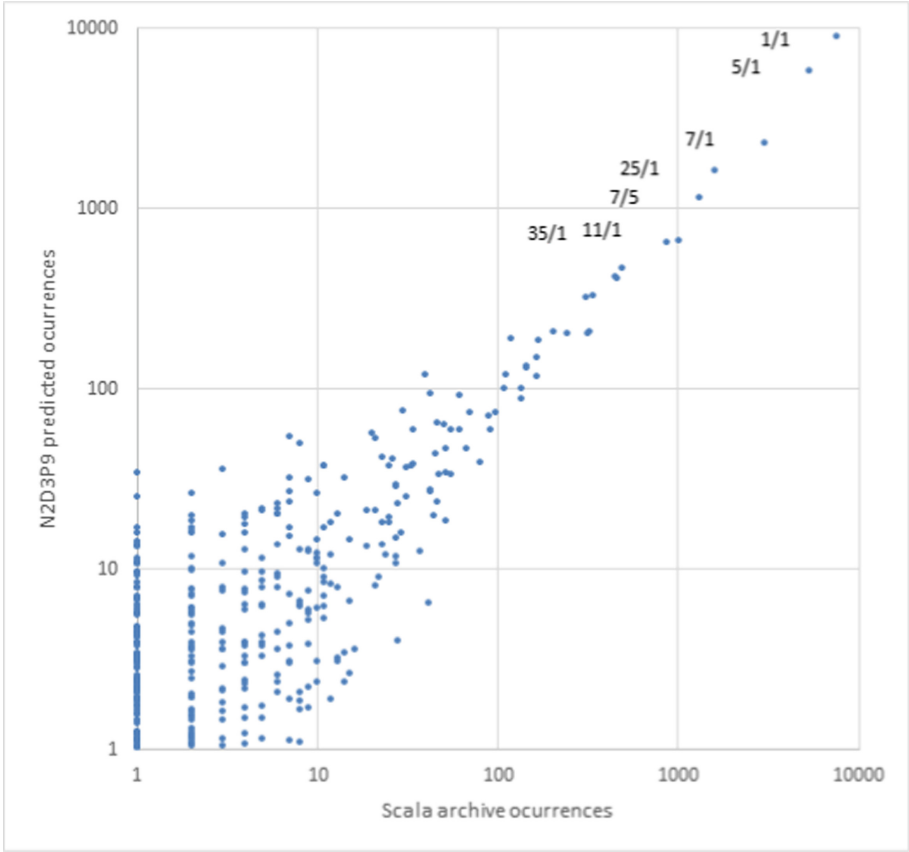**Fig. 1.** Most common 2, 3-equivalent pitch classes by Scala archive occurrences.

Fig. 2. N2D3P9 predicted occurrences of 2, 3-equivalent pitch classes, versus actual Scala archive occurrences.

5   Development/Discovery

From May to August 2020, a collaborative effort [3] to find this function was carried out by members of the Sagittal forum, led by Sagittal co-creator Dave Keenan and Douglas Blumeyer. Many functions besides N2D3P9 were considered before selecting it as the best function for its purpose.

Estimation of pitch ratio popularity is possible because it correlates with numeric simplicity. N2D3P9 is most useful when comparing ranks of more complex ratios, because usage data about such ratios is sparse. By fitting a function to the statistical usage data which is available for simpler ratios, N2D3P9 enables the extension of the patterns found in these simpler ratios.

Rather than attempt to fit functions to the exact counts of votes for each ratio, the functions were fit to the rank indices of each ratio; in other words, a function only needed to sort ratios the same as the actual data, and within each rank position it was unimportant how close its estimate of votes was. In

technical parlance, the goal was to maximize the Spearman's rank coefficient between the estimated ranks and the actual ranks. For purposes of comparing competing functions, maximizing Spearman's rank coefficient could be simplified to minimizing the sum of squared differences between the ranks. But because fitting to the simpler ratios which had more votes is more important, a Zipf's-law weighting was applied to the ranks by taking their reciprocals before calculating their squared differences. A fractional ranking strategy was used to ensure that stretches of the data with tied vote counts did not distort the measurement.

The overall strategy, then, was to minimize this weighted rank error, while also minimizing the complexity of the function, to avoid overfitting. An earlier notational popularity ranking function for 2, 3-removed-ratios, that had been used by the creators of Sagittal was **sopfr** (**s**um **o**f **p**rime **f**actors with **r**epetition). It does a remarkably good job of estimating the rank of pitch ratios given how simple it is. However the weighted sum of squared errors that sopfr gives for the Scala stats is about 0.026, while N2D3P9 reduces that to about 0.010. Functions giving sums of squares as low as 0.008 were found, however, these functions were so complex that they probably were fitting to noise in the Scala stats instead of to the true nature of musical pitch. An informal "chunk" metric was devised to compare function complexity in terms of ability to fit to the data, with considered functions ranging from one chunk sopfr to eight chunks; the winning function N2D3P9 has five chunks.

Several techniques were used to find and decide on N2D3P9 as the best 2, 3-removed-ratio notational-popularity rank-estimation function. Initial observations about shortcomings of sopfr, such as its failure to differentiate balanced ratios from their imbalanced equivalents—such as $\frac{11}{5}$ versus $\frac{55}{1}$—or those with different prime limits such as $\frac{13}{5}$ and $\frac{11}{7}$, despite those pairs of ratios exhibiting remarkably different actual ranks in the Scala stats, formed the basis of the investigation. Psychoacoustic plausibility of functions was used as a top-down guide for experimentation. Optimization tools such as Excel's Evolutionary Solver [4] were used to navigate toward ideal values for each parameter. The approach that was finally successful was a brute-force approach implemented by Douglas Blumeyer, whereby nearly 2 billion functions combined out of constituent "submetrics" were checked automatically. In the end, one of the functions on the short-list generated from the brute-force checker was recognized as being re-writable in a much simpler form with parameter values rounded to whole numbers without doing much damage to its sum-of-squares, and thus N2D3P9 was born.

After deciding upon N2D3P9, the Sagittal forum members checked the ratios for the existing Sagittal symbols against it, to see how well they'd been served by the Scala archive stats and the earlier sopfr metric. Each symbol in Sagittal's JI notations has a default value, or primary comma, which for our purposes is a rational number between 1 and $\sqrt{\frac{3^{19}}{2^{30}}} \approx 1.040404$ exclusive, that by multiplication, takes us from some 3-smooth ratio (notated using standard music notation) to a nearby 5-rough ratio. This allows the symbol to exactly notate all the ratios in the corresponding 2, 3-equivalence-class. For example, the most

important such comma is $\frac{81}{80}$ (the "1/5-comma"), whose symbol is an upward pointing left-half-arrow "/|", which takes us from $\frac{2^4}{3^4}$ (A♭) to $\frac{1}{5}$ (A♭/|). Based on N2D3P9, it was found that only a couple of these commas should be changed (these were among the rarest-used symbols in Sagittal). This was as expected; N2D3P9 was developed primarily in order to add new accent marks to Sagittal, to enable it to exactly notate even rarer JI pitches than it already does (Table 1).

Table 1. Top 100 2, 3-equivalent pitch ratio classes by N2D3P9.

2, 3-equivalent pitch class ratio	N2D3P9	N2D3P9 rank	Scala archive rank	Scala archive occurrences
$\{1\}_{2,3}$	1	1	1	7624
$\{5\}_{2,3}$	1.39	2	2	5371
$\{7\}_{2,3}$	2.72	3	3	3016
$\{25\}_{2,3}$	3.47	4	4	1610
$\{7/5\}_{2,3}$	4.54	5	5	1318
$\{11\}_{2,3}$	6.72	6	6	1002
$\{35\}_{2,3}$	6.81	7	7	875
$\{125\}_{2,3}$	8.68	8	8	492
$\{13\}_{2,3}$	9.39	9	10	447
$\{49\}_{2,3}$	9.53	10	9	463
$\{11/5\}_{2,3}$	11.2	11	11	339
$\{25/7\}_{2,3}$	11.34	12	14	312
$\{13/5\}_{2,3}$	15.65	13	16	205
$\{11/7\}_{2,3}$	15.69	14	12	324
$\{49/5\}_{2,3}$	15.88	15	15	246
$\{17\}_{2,3}$	16.06	16	13	318
$\{55\}_{2,3}$	16.81	17	24	119
$\{175\}_{2,3}$	17.01	18	17	168
$\{19\}_{2,3}$	20.06	19	18	166
$\{625\}_{2,3}$	21.7	20	21	143
$\{13/7\}_{2,3}$	21.91	21	20	145
$\{65\}_{2,3}$	23.47	22	50	40
$\{77\}_{2,3}$	23.53	23	25	111
$\{245\}_{2,3}$	23.82	24	19	165
$\{49/25\}_{2,3}$	26.47	25	23	134
$\{17/5\}_{2,3}$	26.76	26	26	108

(*continued*)

Table 1. (*continued*)

2, 3-equivalent pitch class ratio	N2D3P9	N2D3P9 rank	Scala archive rank	Scala archive occurrences
$\{25/11\}_{2,3}$	28.01	27	47	42
$\{125/7\}_{2,3}$	28.36	28	33	62
$\{23\}_{2,3}$	29.39	29	22	136
$\{91\}_{2,3}$	32.86	30	57	30
$\{343\}_{2,3}$	33.35	31	31	70
$\{19/5\}_{2,3}$	33.43	32	27	97
$\{13/11\}_{2,3}$	34.43	33	29	89
$\{121\}_{2,3}$	36.97	34	42.5	46
$\{17/7\}_{2,3}$	37.46	35	40	50
$\{25/13\}_{2,3}$	39.12	36	52.5	34
$\{77/5\}_{2,3}$	39.21	38	28	92
$\{55/7\}_{2,3}$	39.21	38	34	61
$\{35/11\}_{2,3}$	39.21	38	35.5	55
$\{85\}_{2,3}$	40.14	40	78	20
$\{275\}_{2,3}$	42.01	41	147	7
$\{875\}_{2,3}$	42.53	42	76	21
$\{29\}_{2,3}$	46.72	43	32	67
$\{19/7\}_{2,3}$	46.8	44	37.5	52
$\{23/5\}_{2,3}$	48.98	45	44	45
$\{95\}_{2,3}$	50.14	46	72	23
$\{143\}_{2,3}$	51.64	47	66	26
$\{31\}_{2,3}$	53.39	48	30	80
$\{3125\}_{2,3}$	54.25	49	52.5	34
$\{91/5\}_{2,3}$	54.77	51	68	25
$\{65/7\}_{2,3}$	54.77	51	102.5	11
$\{35/13\}_{2,3}$	54.77	51	102.5	11
$\{49/11\}_{2,3}$	54.9	53	54	33
$\{343/5\}_{2,3}$	55.58	54	55.5	31
$\{119\}_{2,3}$	56.19	55	252.5	3
$\{325\}_{2,3}$	58.68	56	604.5	1
$\{385\}_{2,3}$	58.82	57	37.5	52
$\{17/11\}_{2,3}$	58.87	58	35.5	55
$\{1225\}_{2,3}$	59.55	59	41	47

Table 1. (*continued*)

2, 3-equivalent pitch class ratio	N2D3P9	N2D3P9 rank	Scala archive rank	Scala archive occurrences
$\{169\}_{2,3}$	61.03	60	86	14
$\{121/5\}_{2,3}$	61.62	61	147	7
$\{77/25\}_{2,3}$	65.35	62	63	27
$\{125/49\}_{2,3}$	66.17	63	63	27
$\{25/17\}_{2,3}$	66.9	64	134.5	8
$\{23/7\}_{2,3}$	68.57	65	47	42
$\{17/13\}_{2,3}$	69.57	66	47	42
$\{125/11\}_{2,3}$	70.02	67	147	7
$\{133\}_{2,3}$	70.19	68	329	2
$\{625/7\}_{2,3}$	70.89	69	113	10
$\{115\}_{2,3}$	73.47	70	604.5	1
$\{19/11\}_{2,3}$	73.54	71	55.5	31
$\{37\}_{2,3}$	76.06	72	42.5	46
$\{49/13\}_{2,3}$	76.68	73	147	7
$\{29/5\}_{2,3}$	77.87	74	59.5	28
$\{455\}_{2,3}$	82.15	75	186.5	5
$\{539\}_{2,3}$	82.35	76	186.5	5
$\{1715\}_{2,3}$	83.37	77	76	21
$\{25/19\}_{2,3}$	83.56	78	79.5	19
$\{143/5\}_{2,3}$	86.06	80	217.5	4
$\{65/11\}_{2,3}$	86.06	80	164	6
$\{55/13\}_{2,3}$	86.06	80	90	13
$\{121/7\}_{2,3}$	86.27	82	164	6
$\{19/13\}_{2,3}$	86.91	83	45	44
$\{187\}_{2,3}$	88.31	84	217.5	4
$\{31/5\}_{2,3}$	88.98	85	68	25
$\{91/25\}_{2,3}$	91.28	86	329	2
$\{55/49\}_{2,3}$	91.5	87	39	51
$\{605\}_{2,3}$	92.43	88	94.5	12
$\{343/25\}_{2,3}$	92.63	89	68	25
$\{41\}_{2,3}$	93.39	90	72	23
$\{119/5\}_{2,3}$	93.66	92	123.5	9
$\{85/7\}_{2,3}$	93.66	92	–	0
$\{35/17\}_{2,3}$	93.66	92	217.5	4

330 D. Keenan and D. Blumeyer

<div align="center">Table 1. (<i>continued</i>)</div>

2, 3-equivalent pitch class ratio	N2D3P9	N2D3P9 rank	Scala archive rank	Scala archive occurrences
$\{125/13\}_{2,3}$	97.8	94	604.5	1
$\{275/7\}_{2,3}$	98.03	95.5	102.5	11
$\{175/11\}_{2,3}$	98.03	95.5	147	7
$\{425\}_{2,3}$	100.35	97	329	2
$\{169/5\}_{2,3}$	101.71	98	329	2
$\{121/25\}_{2,3}$	102.7	99	217.5	4
$\{43\}_{2,3}$	102.72	100	58	29

References

1. Secor, G.D., Keenan, D.C.: Sagittal: A Microtonal Notation System. Xenharmonikôn, vol. 18 (2006). https://sagittal.org/sagittal.pdf. Accessed Mar 2022
2. The Huygens-Fokker Foundation's scale archive. https://www.huygens-fokker.org/microtonality/scales.html. Accessed Mar 2022
3. Sagittal forum thread "developing a notational comma popularity metric". https://forum.sagittal.org/viewtopic.php?f=4&t=493. Accessed Mar 2022
4. New and Improved Solver - Microsoft 365 Blog. https://www.microsoft.com/en-us/microsoft-365/blog/2009/09/21/new-and-improved-solver/. Accessed Mar 2022

Performing Easley Blackwood's *Twelve Microtonal Etudes*: An Open-Source Software Development Project

Richard Leinecker and William R. Ayers(✉)

University of Central Florida, Orlando, FL 32816, USA
{richard.leinecker,william.ayers}@ucf.edu

Abstract. This paper outlines an open-source development project dedicated to the performance of microtonal music, specifically Easley Blackwood's *Twelve Microtonal Etudes* (1980). Despite the piece's fixed format, Blackwood stated a specific desire to have the work performed live. This project incorporates multiple elements to this effect, including a standalone software synthesizer and a web application. The paper details prior efforts for microtonal performance on traditional keyboard instruments by Joel Mandelbum and Robert Hasegawa, describes the emergent issues inherent in such endeavors to translate between standard performance practices and microtonal tunings, and proposes methods that will allow for the accurate and easily replicable performance of Blackwood's etudes. The software synthesizer, proposed as a pedagogical design project for computer science students at the University of Central Florida, is currently under active development. The web application, available at the project website, includes source code and documentation for the use of future development teams.

Keywords: Microtonal · Equal temperament · Easley Blackwood · Performance · Keyboard · Gesture

1 Introduction

Easley Blackwood's *Twelve Microtonal Etudes* (1980) [1] are a unique case in the world of electronic music. The pieces, each in a different equal-tempered tuning, were created as a fixed-format collection, performed and recorded by Blackwood on the Polyfusion Series 2000 Synthesizer between March 1979 and July 1980, but unlike most works in this genre, the pieces also subsist in a printed score published in 1982 [2], offering analysts an opportunity to investigate with relative confidence the novel harmonic structures that Blackwood composed. Upon the rerelease of the recordings for a 1994 album, Blackwood noted his desire to hear new interpretations of the work by other performers [3]; his prefatory notes to the score indicate the same sentiment but also warn of the difficulties of the task since the pieces are not performable on standard instruments. While occasional arrangements of the collection have appeared over the years, the full work has seemingly never received a live performance despite having a complete score dedicated

to this task. Blackwood provides a roadmap for the performance of these works, but he does not include a detailed set of directions or instructions that can lead us to a satisfying destination.

This situation creates a remarkable circumstance for Blackwood's work since it is, at once, a static entity in its original fixed form and a potential space for dynamic interpretation. The problems with such an enterprise are well known. The musical score itself is not easily performable due to a lack of available microtonal instruments; neither is the music simple to arrange with its unorthodox set of accidentals (see Fig. 1). Additionally, finding an adequate number of performers who are willing and able to learn such complex notation schemes is a significant undertaking. The current project sets out to resolve the performance issues of this work through a multifaceted strategy involving musical analysis, gestural interpretation, and software development. The aim is to create a procedure for the performance of these works that is not only functional but also able to be easily replicated. To this goal, the project is completely open-source from the outset.

Fig. 1. Easley Blackwood, *Twelve Microtonal Etudes*, "21 Notes," mm. 1–6

This paper offers a set of first steps toward this goal. We will first outline some other microtonal performance practices that align with Blackwood's vision. These methods can serve as inspiration for the project's efforts. We will then detail the parts of the current project that are required to achieve a full performance of Blackwood's etudes, focusing on two applications: a software synthesizer that provides native support for equal-tempered microtonal playback and MIDI keyboard mapping and a web application that replicates some of this functionality for a general audience. Through these development elements, we intend to produce a full performance of Blackwood's etudes, living up to his stated aspirations for the work.

2   Examining Some Keyboard Mappings for Microtonal Music

While keyboards with hexagonal layouts (such as the MicroZone from Starr Labs or the novel Lumatone devised by Siemen Terpstra) have received a great deal of support in

recent years due to their accommodation of microtonal scales and tunings, the cost and unconventionality of these instruments leaves little opportunity for their widespread use among performing keyboardists. Notably, the Polyfusion synthesizer used for Blackwood's etude recordings is not a dedicated instrument for microtonal performance—it uses a four-octave keyboard with a traditional twelve-tone organization—but its size, complicated design, and rarity make it an uncommon instrument for most keyboard players to have on hand. The traditional keyboard allowed Blackwood to use standard gestures that would have been natural to him as a concert pianist, but it also required a translation of the non-twelve-tone tunings into the common twelve-tone layout. The etudes, composed in equal temperaments using between thirteen and twenty-four notes per octave, necessitate a keyboard mapping that either (a) denies gestural octave equivalence by using a one-to-one correspondence of notes to keys or (b) leaves out certain notes of a tuning to maintain a sense of commonality with standard twelve-tone orientations. The choice is between having an unfamiliar note layout (and therefore unfamiliar gestures) or not having all the notes immediately available without changing the mapping. While the first of these options may seem to be a reasonable compromise since it would allow performance of fully "chromatic" music in any of the equal temperaments under consideration, the latter option, leaving out some notes to preserve gestural practices (and some basic intervallic relationships) has a rich tradition in the performance practice of microtonal music. To examine this practice more closely, I'd like to consider two pieces written for traditional keyboards tuned in nineteen-tone equal temperament.

Arranging nineteen tones with roughly even distribution within a single octave has been practiced at least since Guillaume Costeley's experiments in the sixteenth century [4]. Figure 2 provides the most common notation scheme for this tuning, represented as a circle of perfect fifths and as a nineteen-tone chromatic pitch-class circle. Note that the perfect fifth in this tuning is roughly 695 cents and the chromatic step interval roughly 63 cents. These constructive intervals cause a number of notes to lie "between the cracks" of pitch perception when attempting to perform in a given key or a traditional scale. Additionally, as discussed by Julian Hook [5], among others, the notation of nineteen-tone equal temperament does not follow the canonical enharmonic equivalence relation of twelve-tone equal temperament. For instance, the note names "C-sharp" and "D-flat" refer to different pitch classes in this system while still having functions similar to their counterparts in twelve-tone tuning. Table 1 provides basic tuning information (given in approximate cent values rounded to the nearest whole number) for a single octave of nineteen-tone equal temperament.

A number of modern compositions make use of this notation scheme, adapting the tuning for use on traditional keyboard layouts. Joel Mandelbaum's 1961 collection of *Nine Preludes for Two Pianos in 19-Tone Equal Temperament* [6] uses retuned twelve-note keyboards that share the set of frequencies assigned to the black keys, the F-sharp major pentatonic collection, but the pianos utilize different tunings for the white keys. Figure 3 provides Mandelbaum's mapping and notation scheme for two pianos in nineteen-tone equal temperament, displaying the notes on the nineteen-tone pitch-class circle and how they are oriented on the two keyboards. Notably, both pianos use a maximally even distribution of twelve notes in the nineteen-tone space, ensuring the closest possible approximation to twelve-tone equal temperament. This circumstance

(a)   (b)

Fig. 2. The circle of fifths and the chromatic pitch-class circle for nineteen-tone tuning

Table 1. Notation and tuning of nineteen-tone equal temperament

Pitch class	Note name	Cents from C (approximate)
0	C-natural	0
1	C-sharp/D-double-flat	63
2	C-double-sharp/D-flat	126
3	D-natural	189
4	D-sharp/E-double-flat	253
5	D-double-sharp/E-flat	316
6	E-natural	379
7	E-sharp/F-flat	442
8	F-natural	505
9	F-sharp/G-double-flat	568
10	F-double-sharp/G-flat	632
11	G-natural	695
12	G-sharp/A-double-flat	758
13	G-double-sharp/A-flat	821
14	A-natural	884
15	A-sharp/B-double-flat	947
16	A-double-sharp/B-flat	1011
17	B-natural	1074
18	B-sharp/C-flat	1137

suits Mandelbaum's compositional choices quite well, especially for his sixth prelude which explores twelve-tone serialism in the nineteen-tone tuning. However, the unusual note names for the white keys of Piano II may cause some unwanted performance issues as a pianist would have to retrain certain associations when reading the music. The gestural motions of the pianist using this mapping are likely to be similar to those of traditional performance practice due to the maximally even distribution, but the process of learning the piece with such an unfamiliar notation scheme could prove laborious (or could at least cause an unnecessary difficulty).

Piano I	Piano II
B♯/C♭ C C♯ B D♭ B♭ D A♯ D♯ A E♭ A♭ E G♯ E♯/F♭ G F G♭ F♯	B♯/C♭ C C♯ B D♭ B♭ D A♯ D♯ A E♭ A♭ E G♯ E♯/F♭ G F G♭ F♯
C♯ D♯ F♯ G♯ A♯ C D E F G A B	C♯ D♯ F♯ G♯ A♯ C♭ D♭ E♭ F♭ G♭ A♭ B♭
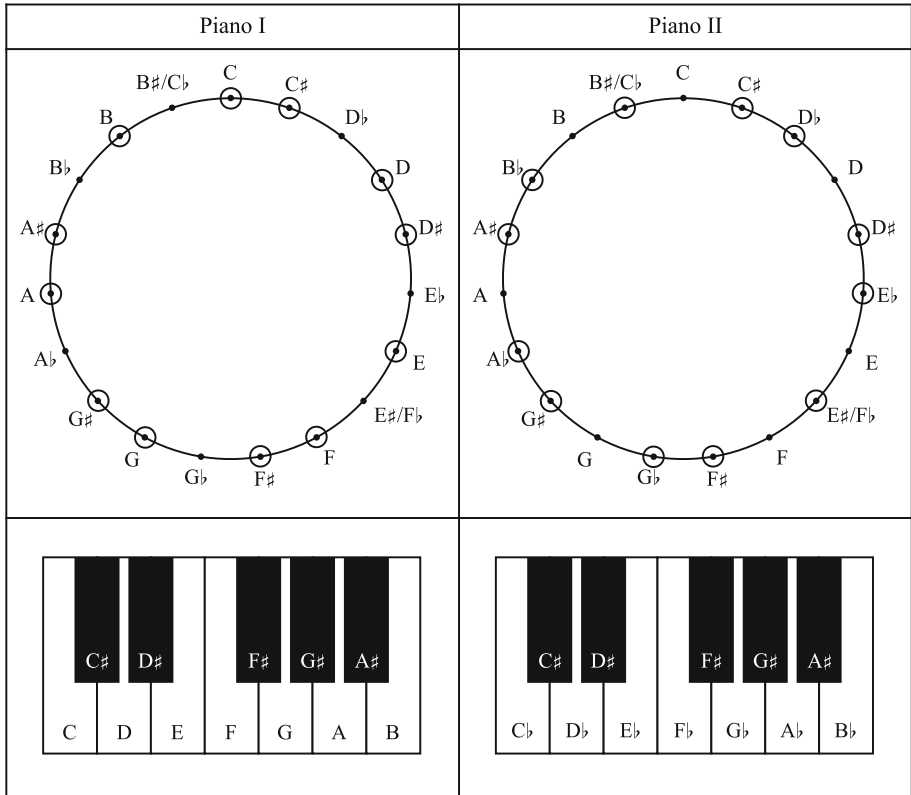
Fig. 3. Mandelbaum's mapping for two pianos in nineteen-tone equal temperament

The mapping used in Robert Hasegawa's *Due Corde* [7], while quite similar, rectifies the atypical note-name-to-key associations from Mandelbaum. Figure 4 provides Hasegawa's mapping for two pianos in nineteen-tone equal temperament, introduced to him by Jon Wild. This piece not only uses a maximally even distribution of notes for both pianos but also maintains traditional associations between notes and keys for both pianos. The notes of Piano I are the same chosen by Mandelbaum, but Piano II aligns the flat note names with their traditionally corresponding keys. The pianos share the notes of the F major pentatonic. This mapping allows for both easy readability and familiar

gestures when offering the work to a performer, alleviating many of the performance issues that arise when dealing with microtonal music.
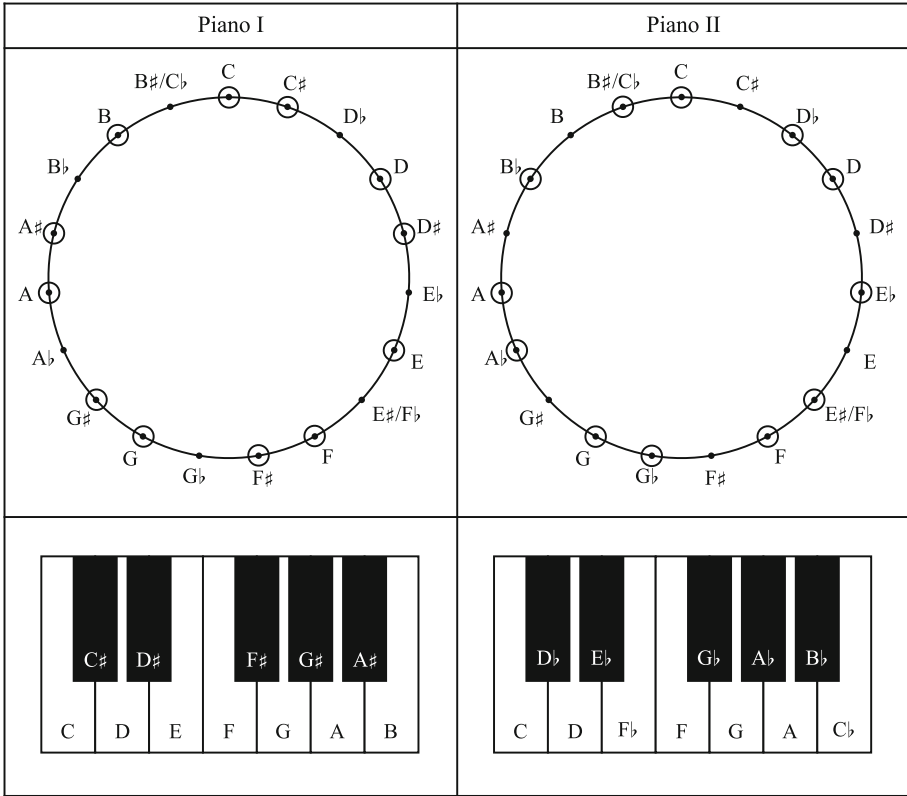

Fig. 4. Hasegawa's (and Wild's) mapping for two pianos in nineteen-tone equal temperament

From this brief examination of tuning practices for nineteen-tone microtonal music, it seems that a successful mapping requires both readability and gestural familiarity to promote easy performance. Applying this concept to Blackwood's etudes is slightly more involved due to the number of different tunings, the complications of his notation, and the sheer scope of the work, but these mappings by Mandelbaum and Hasegawa highlight some of the steps that need to be taken in order to achieve this feat.

3   An Open-Source Project for Microtonal Music Software

The current development project takes a multifaceted approach to resolving these issues of performing Blackwood's works. The project will ultimately include three novel elements: (1) an open-source, cross-platform softsynth that allows for front-end remapping of MIDI keyboard input, (2) a score that translates chosen mappings to traditional notation for ease of reading and reproduction, and (3) an online application for practice

and demonstration during the rehearsal and development process. This section briefly outlines the first two elements before detailing completed work on the final element.

The first of these elements, a software synthesizer capable of microtonal output and MIDI keyboard remapping, will be the first point of contact for performers, so the main concerns for its development are ease of use and functionality. The program itself was proposed as a pedagogical design project for senior Computer Science majors at the University of Central Florida. The program (see Fig. 5) requests a base frequency (in Hz) and the cardinality of the tuning (the number of notes per octave) to generate a single octave of frequencies, which can then be mapped to a keyboard representation. Mappings and parameters are fully adjustable in a popup window. The program includes multiple oscillators to generate instrument sounds that replicate Blackwood's timbres (which are also fully adjustable for fine tuning), allowing the program to bypass compatibility issues with virtual studio technologies and digital audio workstations. The key mappings can be customized through nodes and lines drawn between frequencies and keys and can also be saved for future use. While these remapping, tuning, and generation abilities exist in prior efforts, they are often split between programs. This open-source endeavor combines and simplifies these functionalities for an easily replicable performance practice.
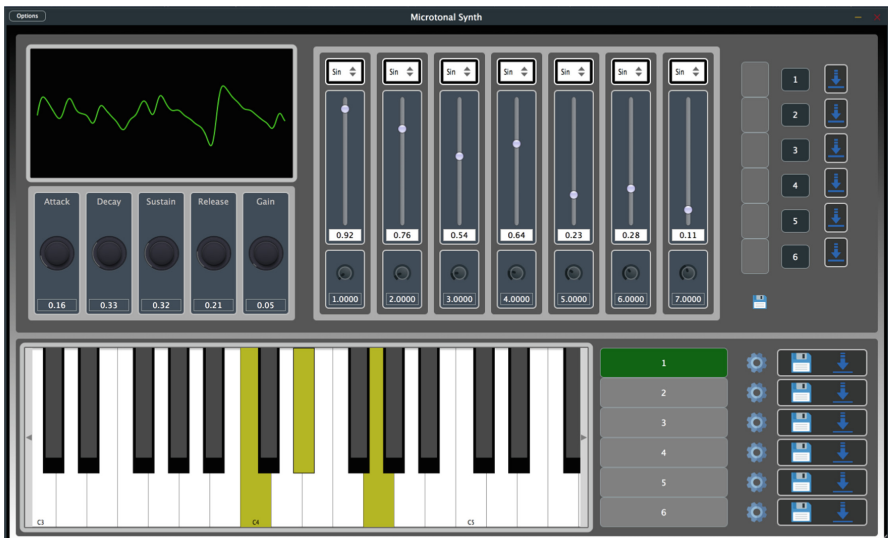

Fig. 5. User interface for a software synthesizer capable of microtonal output and MIDI keyboard remapping

The second element of this project, a reworked score that provides traditional notation and instructions for performers, introduces some interesting difficulties that are worth mentioning here. The purpose of this element is to offer an easy method for trained pianists to recreate the musical output while avoiding the difficulties of learning a completely new system of musical notation. Since Blackwood's general attitude toward composition with microtonal tunings is to incorporate elements of tonal structure [8], his score allows for preset mappings of the synthesizer program that utilize maximally

even distributions of notes in most cases, much like Mandelbaum's and Hasegawa's mappings, which would provide (at least nominally) a sense of gestural simplicity or familiarity. Therefore, the only choices to be made are how to organize Blackwood's voices, only a single musical line each, into performable keyboard parts that contain associated material fitting into a single mapping. There are moments, as in the second theme of the etude "19 Notes," where distributions are far more chromatic in nature, as mentioned by Hook [5] in his examination of the piece. From the density of notes in the score, the piece is conceivably performable by only four keyboard players, but moments of dense chromaticism problematize the organization of players and potentially require adjustments to this number.

The final element of this project is crucial to early stage demonstration and for the purposes of executing and testing mapping possibilities and gestural ideas for score and part production. The novel online application for this project replicates basic design elements of the softsynth, including microtonal mapping (see Fig. 6), but allows anyone with a web browser to test out microtonal keyboard mappings and their effect on gestural performance. The browser application allows users to set a reference frequency, the number of notes within one octave of the chosen tuning, and the basic wave type (sine, square, triangle, or some other synthesized options). While the application removes some levels of flexibility due to the web format, specifically the adjustable instrument sounds, it provides plug-and-play ability with a MIDI keyboard through the Web MIDI API [9], meaning that anyone who loads the page with a keyboard plugged into their computer can control the output exactly as the standalone program would. Without a MIDI device, users can still type on the computer keyboard or click on the screen to activate notes and test sound combinations. The ultimate goal of this application is to test arrangement options for Blackwood's piece and to offer an introduction to interested performers and organizers. Further information about the application, including all source code and documentation, is available at the project website [10].
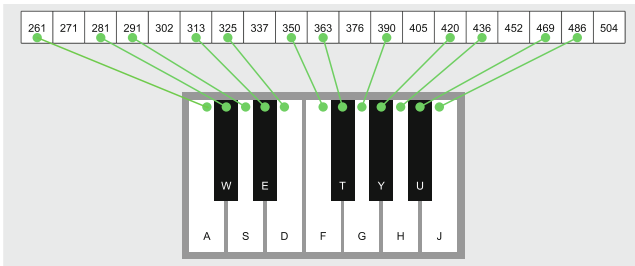

Fig. 6. Portion of the user interface for a web application capable of microtonal output and MIDI keyboard remapping

4   Conclusion

This project is in the first year of a three-year development schedule, including time for score/part production, fine tuning/debugging, and rehearsal/implementation. Progress

thus far has produced a functional standalone program, a web application, and extensive documentation for potential adaptations by future teams. Notably, many sections of Blackwood's score (including the opening of the etude in nineteen-tone equal temperament) can be reproduced reliably by the web application in its current state, allowing effective demonstrations of the project's potential. Our goal is to produce a full performance of the work that can be reproduced with minimal effort, using a single cross-platform application with an associated score that requires little in terms of preparation beyond rehearsal. The hope is to live up to Blackwood's stated desire to replicate his music in a live format.

References

1. Blackwood, E.: Twelve Microtonal Etudes for Electronic Music Media [Album]. Easley Blackwood, E-639 (1980)
2. Blackwood, E.: Twelve Microtonal Etudes for Electronic Music Media. G. Schirmer, New York (1982)
3. Blackwood, E.: Easley Blackwood: Microtonal [Album]. Cedille Records, CDR 90000 018 (1994)
4. Wibberley, R.: Syntonic tuning: creating a model for accurate electronic playback. Music Theory Online **10**(1) (2004). https://mtosmt.org/issues/mto.04.10.1/mto.04.10.1.wibberley3.html
5. Hook, J.: Enharmonic systems: a theory of key signatures, enharmonic equivalence and diatonicism. J. Math. Music **1**(2), 99–120 (2007)
6. Mandelbaum, J.: Multiple division of the octave and the tonal resources of 19-tone temperament. Ph.D. dissertation. Indiana University (1961)
7. Hasegawa, R.: Due Corde. Self-published (2002)
8. Blackwood, E.: The Structure of Recognizable Diatonic Tunings. Princeton University Press, Princeton (1985)
9. Web-MIDI website. https://www.midi.org/developer-web-midi-info. Accessed 15 Jan 2022
10. Open-Source Microtonal Project website. https://microtonality.net/. Accessed 15 Jan 2022

Short Papers

Identifying Metric Types with Optimized DFT and Autocorrelation Models

Matt Chiu[1(✉)] and Jason Yust[2]

[1] Eastman School of Music, Rochester, NY, USA
mchiu9@u.rochester.edu
[2] Boston University, Boston, MA, USA
jyust@bu.edu

Abstract. This paper explores the classification of metric types using different feature representations. Using weighted timepoint, DFT, and autocorrelation, we train feedforward neural networks to distinguish allemandes, courantes, sarabandes, and gavottes in the Yale-Classical Archives Corpus. Autocorrelation and DFT models perform better than a baseline, with DFT consistently better by a small amount.

Keywords: Discrete Fourier transform · Autocorrelation · Meter classification · Metric types · Neural networks

Music theorists typically define meter as an abstract hierarchy, either as a hierarchical accent pattern on an underlying pulse stream [9] or a containment hierarchy of timespans [17]; see [5]. This is sufficient to represent musical time signatures, but musical practice also recognizes metrical types with the same time signatures and/or metrical hierarchies, for which these kinds of theories are therefore too abstract. Traditional dance meters of eighteenth-century Western Europe are a convenient example of this: Metric types which sometimes share metrical hierarchies are nonetheless distinguishable in practice. In this paper we explore the classification of allemandes, courantes, sarabandes, and gavottes using machine learning methods and three feature representations, a baseline weighted timepoint representation, autocorrelation, and discrete Fourier transform (DFT). Autocorrelation may be understood as an interval-based representation, while DFT is a periodicity-based representation.

After a review of both techniques and a discussion on the corpus preparation, we report on three feedforward neural networks models trained on data from the Yale Classical Archives Corpus [16] using the three representations and evaluated the models based on their ability to classify the four different baroque dance types.

1 Procedure

1.1 DFT

The DFT transfers a signal from the time domain to the frequency domain. With a discrete time-domain signal represented as a vector $X = (x_1, x_2, \dots)$ of length N, the DFT is a complex-valued vector, $F(X)$, defined by

$$F_k(X) = \sum_{n=0}^{N-1} a_j e^{-i2\pi kn/N} = \sum_{j=0}^{N-1} x_j(\cos(2\pi kj/N) + i\sin(2\pi kj/N)) \quad (1)$$

Each place in the DFT vector, k, represents a periodic function of period N/k. We are only interested in the size of each of these, so we take the norm of each component, $|F_k(X)|$ (eliminating the phase) and divide by $|F_0(X)|$ so that all values range from 0 to 1. The DFT is an orthogonal transform; each of the $F_k(X)$ for $1 \leq k \leq N/2$ is independent of the others.[1] Therefore, it partitions the total weight of the time-domain signal into all the possible frequencies dividing the fixed period N. When k divides N of these will coincide with traditional metrical periodicities. The DFT has been used for meter detection in audio signals [7,11] and it has been used in music analysis to relate meter to form [4] and to describe rhythmic canons in Steve Reich's music [18,19].

1.2   Autocorrelation

Autocorrelation is a correlation of a signal with itself at every possible lag value. It acts like a rhythmic interval vector, listing the weighted number of occurrences of each rhythmic interval (temporal distance between onsets). More precisely the autocorrelation is a vector $R(X)$ defined by

$$R_k(X) = \frac{1}{\sigma_X^2} \sum_{i=0}^{N-1} (x_i - \bar{x})(x_{i+k} - \bar{x}) \quad (2)$$

A number of studies demonstrate the use of autocorrelation to identify meter in symbolic (score or MIDI) data [3,12,14,15] and audio [6].

Autocorrelation is closely related to the DFT. Specifically, it can be understood as *squaring the signal in the frequency domain* by appealing to the convolution theorem; see [2]. To make this precise, accounting for the normalization in Eq. 2, define X' as the zero-mean version of X, i.e. $X' = X - \bar{x}$. This affects only the zeroth DFT coefficient. Then

$$R(X) = \frac{1}{\sigma_X^2 N} F(|F(X')|^2) \quad (3)$$

The function $(1/N)F(X)$ is the inverse Fourier transform, so this means that after removing the phase information autocorrelation returns the data to the time domain. The autocorrelation is therefore a vector of time intervals, whereas the DFT is a vector of frequencies, but otherwise contain essentially the same information.

---

[1] For $k > N/2$, $F_k(X)$ and $F_{N-k}(X)$ have equal magnitude and opposite phase for a real-valued signal, X, by the aliasing principle.

1.3   Corpus and Data Preparation

The Yale Classical Archives Corpus (YCAC) is comprised of "salami slices" of MIDI performance data [16]. A slice occurs everywhere that a new note is introduced, or a note ceases to sound. We isolated pieces in the YCAC by Bach with "allemandes," "courante," "sarabande," or "gavotte" in the title. This procedure found 76 pieces, consisting of 45,344 pitch slices altogether (Table 1). In an attempt to emphasize newly introduced notes [13], notes that were contained in the immediately prior slice were removed. Sarabandes and courantes are in triple meters, usually 3/4, and allemandes and gavottes are in duple meter, usually 4/4 and 2/2. Each dance style also has a corresponding rhythmic character: sarabandes accent beat 2, gavottes have long pick ups (starting half-way through a measure), allemandes have quick, sixteenth note pick-ups, and courantes frequently contain metric ambiguities.

Table 1. Corpus

	Allemandes	Sarabande	Gavotte	Courante	Total
Piece count	24	22	8	22	76
Total slices	16,833	9,249	4,014	15,068	45,344
Ave. slices per piece	701	429	502	685	2,317
Length in ♩	4,228	2,924	1,366	4,981	13,468

1.4   Weighting, Windowing, and Training

Perceptual theories of meter finding (e.g. [9,10]) often claim that a variety of musical features influence meter induction by imparting "phenomenal accent" to time points. We devised a relatively simple weighting scheme on slices based on three factors: the number of notes, duration, and bass notes. Recall that notes are removed if they occur in a preceding slice, so the first factor counts the number of new pitches introduced at a given time point. The duration factor represents the assumption that slices of longer duration will tend to have more metrical weight ("agogic accent"). We include a weighting parameter, δ, which we multiply by the duration of each slice in quarter notes. The register factor reflects the assumption that new bass notes will tend to be metrically weighted. We define a bass note as the lowest note within γ quarter notes before or after the given onset time. We add a constant, τ, for any slice where a bass note occurs. There are thus has three adjustable parameters, δ, γ, and τ. Figure 1 shows a sample score fragment with calculated weights for selected slices with (δ, γ, τ) set to $(1, 2, 3)$. During the training process we tuned the parameters to assess their impact.

We transformed the extracted chord slices to time-series data, encoding each score as an array dividing the full duration of the piece into 32nd notes. We placed the rhythmic weight for each slice in the time-series array at its corresponding

Fig. 1. Sample score with selected slices, showing the data structure and time-point weighting procedure with $(\delta, \gamma, \tau) = (1, 2, 3)$.

onset position, and put zeros elsewhere. Table 1 shows the number of quarter notes in each of the four metric types. To prevent wraparound for the DFT, windows were zero-padded with 96 additional 0s at the end of each vector.

For each metric type, we extracted 1,000 random windows, each 12 quarter notes in length. Our corpus thus consisted of 4,000 time-series windows and 4,000 corresponding labels identifying the correct metric type. We separated 950 samples of each type for training data, leaving 200 windows, 50 windows of each metric type, for evaluating the models. We fed the three inputs – baseline weights, autocorrelations, and DFTs – into the same neural network architecture: an input layer of 192, and 2 hidden layers of 30 and 10 neurons respectively each using relu activation [1]. We also used the Adam optimization algorithm [8]. The models were trained with 10 epochs on the training corpus (of 3,800 windows), repeated for each different tuning of weighting parameters (δ, γ, τ).

2   Results

Table 2 reports the evaluation scores as categorical accuracy, the percent of correct predictions based on the input. We found, as in [12], that adjusting the weighting parameters (δ, γ, τ) only alters predictions modestly and with no obvious trends. Even eliminating the duration weightings (δ) and bass note weightings (τ) entirely does not reduce performance, except in one case (eliminating both features in the autocorrelation condition). Therefore, in the DFT condition, where identification was the best, it appears to be based entirely on the basic rhythm and number of new pitches.

Excluding trials with 0-weighted features, the autocorrelation models ranged from 66%–77% with an average of 72% accuracy, and the DFT models ranged from 70%–80% with an average of 75% accuracy. Both models performed consistently around the same level with the DFT model modestly better. The control averaged 46% accuracy. The confusion matrix in Table 3 shows that gavottes were better classified than all other types, probably because there were fewer gavottes in the data set, so the classifier was more likely to be trained on excerpts from

Table 2. Categorical accuracy predictions

δ		1							0	1	2	3	4	5	6	0	1	2	3	4	5	6
τ	0	1	2	3	4	5	6				1					0	1	2	3	4	5	6
$\gamma = 2$																						
Control	.51	.47	.47	.47	.41	.40	.40	.49	.47	.48	.48	.51	.49	.48	.47	.47	.40	.40	.47	.44	.46	
DFT	.80	.78	.80	.72	.68	.70	.73	.77	.78	.76	.79	.74	.76	.77	.79	.78	.75	.74	.75	.69	.73	
Autocorr.	.75	.73	.71	.71	.70	.74	.75	.73	.73	.71	.73	.72	.72	.69	.74	.73	.66	.73	.69	.69	.70	
$\gamma = 3$																						
Control	–	.49	.48	.44	.50	.47	.44	.50	.49	.56	.45	.50	.49	.44	–	.49	.52	.44	.47	.43	.40	
DFT	–	.79	.75	.75	.76	.76	.74	.80	.79	.79	.75	.75	.80	.81	–	.79	.75	.74	.75	.74	.71	
Autocorr	–	.72	.73	.76	.76	.77	.76	.73	.72	.77	.73	.70	.71	.74	–	.72	.73	.72	.75	.74	.72	

the same piece used in the test, and although these would not have been exactly the same window, they may have had similar traits.

Table 3. Confusion matrix for DFT/autocorrelation, all with $(\delta, \gamma, \tau) = (6, 3, 1)$.

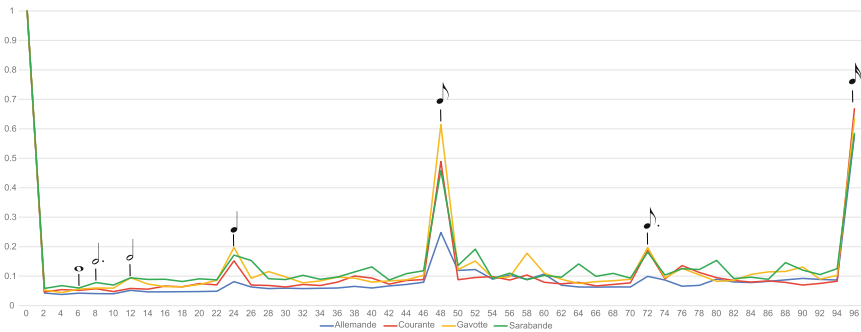|	Allemande	Courante	Gavotte	Sarabande	Accuracy ($n = 50$)
Allemande	38/34	6/7	0/2	6/7	76%/68%
Courante	4/9	32/35	3/2	11/4	64%/70%
Gavotte	0/0	0/5	50/44	0/1	100%/92%
Sarabande	2/5	7/7	0/4	41/34	82%/68%

Fig. 2. Average DFT values for each metric type at $(\delta, \gamma, \tau) = (6, 3, 1)$.

Figure 2 shows the average DFTs for $(\delta, \gamma, \tau) = (6, 3, 1)$ from which we can infer some of the differences that the classifier may have relied on to distinguish metric types.[2] The main differences are that allemandes are generally flat down to the sixteenth-note level, meaning that higher metrical levels were not very

---

[2] We include only even values here, because the zero padding produces distracting artifacts in the odd coefficients.

salient in these pieces. Higher metrical levels were better detected in sarabandes and gavottes, but differences relating to distinctions between duple and triple meter (in coefficients 6, 8, and 12) are weak at best. Therefore, the classifier is likely relying more on the salience of different metrical levels rather than differences between the duple and triple metrical hierarchies that would be emphasized in traditional metric theory.

References

1. Agarap, A.F.: Deep learning using rectified linear units (ReLU). arXiv preprint arXiv:1803.08375 (2018)
2. Amiot, E.: Music Through Fourier Space. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-319-45581-5
3. Brown, J.C.: Determination of the meter of musical scores by autocorrelation. J. Acoust. Soc. Am. **94**(4), 1953–1957 (1993)
4. Chiu, M.: Form as meter: metric forms through Fourier space. Master's thesis, Boston University (2018)
5. Cohn, R.: Preferatory note: how music theorists model time. J. Music Theory **65**(1), 11–16 (2021)
6. Eck, D.: Identifying metrical and temporal structure with an autocorrelation phase matrix. Music. Percept. **24**(2), 167–176 (2006)
7. Flannick, J.E., Hall, R.W., Kelly, R.: Detecting meter in recorded music. In: Sarhangi, R., Moody, R.V. (eds.) Renaissance Banff: Mathematics, Music, Art, Culture, pp. 195–202. Bridges Proceedings (2005)
8. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
9. Lerdahl, F., Jackendoff, R.: A Generative Theory of Tonal Music, Reissue, with a New Preface. MIT Press (1996)
10. Povel, D.J., Essens, P.: Perception of temporal patterns. Music. Percept. **2**(4), 411–440 (1985)
11. Sethares, W.A., Staley, T.W.: Meter and periodicity in musical performance. J. New Music Res. **22**(5), 149–158 (2001)
12. Eerola, T., Toiviainen, P.: MIDI Toolbox: MATLAB Tools for Music Research. University of Jyväskylä, Finland (2004)
13. White, C.W.: Meter's influence on theoretical and corpus-derived harmonic grammars. Indiana Theory Rev. **35**(1–2), 93–116 (2018)
14. White, C.W.: Autocorrelation of pitch-event vectors in meter finding. In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O.A. (eds.) MCM 2019. LNCS (LNAI), vol. 11502, pp. 287–296. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_23
15. White, C.W.: Some observations on autocorrelated patterns within computational meter identification. J. Math. Music **15**(2), 181–193 (2021)
16. White, C.W., Quinn, I.: The Yale-classical archives corpus. Empirical Musicology Rev. **11**(1), 50–58 (2016)
17. Yust, J.: Organized Time: Rhythm, Tonality, and Form. Oxford University Press, NY (2018)
18. Yust, J.: Steve Reich's signature rhythm and an introduction to rhythmic qualities. Music Theory Spectr. **43**(1), 74–90 (2021)
19. Yust, J.: Periodicity-based descriptions of rhythms and Steve Reich's rhythmic style. J. Music Theory **65**(2), 325–374 (2021)

Persistent Homology on Musical Bars

Victoria Callet$^{(\boxtimes)}$ (ID)

IRMA, UMR 7501, CNRS, Université de Strasbourg, Strasbourg, France
`victoria.callet@math.unistra.fr`

Abstract. This article presents a new way of building a filtered simplicial complex from a music piece and applying persistent homology in the context of musical analysis. Our approach consists of considering any musical score as the set of its musical bars, which we see as subsets of $\mathbb{R}^3$. With this definition, we may consider the Hausdorff distance between two musical bars, which gives us a point cloud from any score, and that allows us to build the associated Vietoris-Rips complex. We will then use barcodes to visualize persistent homology and give an illustration of our construction on a famous movie music piece.

Keywords: Musical bars · Filtered complex · Persistent homology · Barcodes · Musical analysis

1 Introduction

1.1 Persistent Homology

A filtration of a simplicial complex K is a nested sequence of sub-complexes $\emptyset = K^{-1} \subset K^0 \subset \ldots \subset K^N = K$ of K: we call K a **filtered complex**. A simple filtration is presented for instance in Fig. 1. Starting from a filtered complex, we can compute its simplicial homology (over $\mathbb{F}_2$) at each **time** of the filtration, and persistent homology gives information about inclusions between the various complexes, as explained in [7]. The associated homology groups $H_*(K^s)$ are characterized by their dimensions, which are called the **Betti numbers**. We can visualize persistent homology on a figure called a **barcode**, where the horizontal axis represents progress in the filtration and a bar that starts at time s and ends at time t corresponds to a generator of $H_*(K^s)$ that is still one for $H_*(K^{t-1})$ but not anymore at time t (see [4]). Barcodes allow us to immediately identify classes that **persist** during the filtration. For instance, barcodes associated to the filtered complex in Fig. 1 are presented in Fig. 2, in degrees 0 and 1.
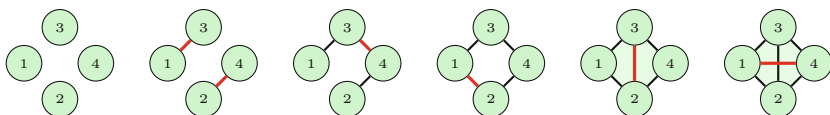

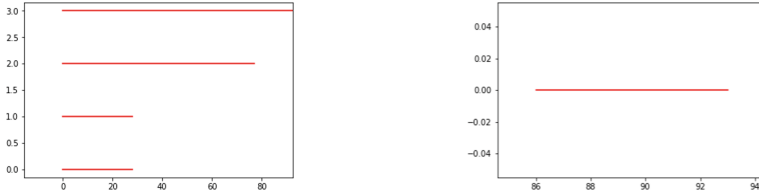**Fig. 1.** A filtered complex with 6 times of filtration.

Fig. 2. Barcodes for filtration of Fig. 1 in degree 0 (left) and degree 1 (right).

Persistent homology is commonly used in a recent field known as **Topological Data Analysis** (TDA), where the main idea is to use topological structures to study and sometimes "recognize" some objects. The general process is summarized in Fig. 3, where in our context the starting object is a musical score.
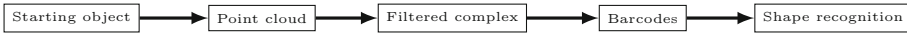

Fig. 3. Topological Data Analysis process.

1.2   Context and Problematic

TDA via persistent homology was already used in the context of musical analysis and especially in automatic classification of musical style, as we can see in [2,3,5] or [1]. The main and common issue is always the following problematic: **how should we associate a filtered complex with a given musical piece?**

The papers we cited above have different approaches to try to find a consistent answer to this question, using pitch-class sets complexes, time series or the Tonnetz. In our work, we chose to consider that a music piece can be represented by a set of distinct musical bars, which can be thought of as subsets of $\mathbb{R}^3$, as defined in Sect. 2.2. Our starting object is a musical score, from which we will extract a point cloud of $\mathbb{R}^3$ by considering its musical bars and the associated Hausdorff distance, as presented in Sect. 2.3. To build the filtration, we will use the Vietoris-Rips method, which is described in Sect. 2.1. Once we obtain a filtered complex, we can compute persistent homology, i.e. barcodes. In Sect. 3, we will illustrate our approach by giving an analysis of the French movie musical piece: *Comptine d'un autre été: l'Après-midi*, by composer Yann Tiersen. In order to allow musical interpretation, we will be interested in dimension 0 and 1 initially, and we will show that these barcodes, especially in degree 0, can capture structural information about the piece.

2   Persistent Homology on Musical Bars

2.1   Filtration: The Vietoris-Rips Method

The basic object of a Vietoris-Rips complex is a **point cloud**, that is, a set of data points with a metric over it.

Definition 1. *Let $X = \{x_1, \ldots, x_n\}$ be a point cloud and $\epsilon \geq 0$ be a **parameter**. The **Vietoris-Rips complex** $\mathcal{R}_\epsilon(X)$ is the simplicial complex with X as its set of vertices and $\sigma = \{x_1, \ldots, x_k\}$ is a k-simplex if and only if its vertices are pairwise close, that is, when $d(x_i, x_j) \leq \epsilon$ for all pairs x_i, x_j of σ.*

Figure 4 shows the classical construction of a Vietoris-Rips complex starting from a given point cloud X and a parameter ϵ. For two given parameters ϵ and ϵ' such as $\epsilon < \epsilon'$, there is the obvious inclusion $R_\epsilon(X) \hookrightarrow R_{\epsilon'}(X)$ and so by simply increasing the parameter ϵ, we will get a natural sequence of nested simplicial complexes, that is, a filtered complex. For instance, $R_0(X)$ is the 0-dimensional simplicial complex with n connected components while $R_\epsilon(X)$ will have only one connected component for ϵ large enough.
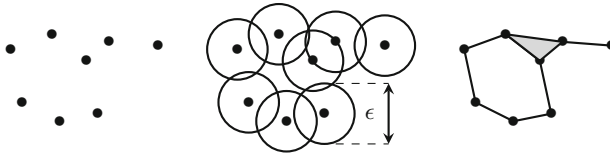

Fig. 4. The Vietoris-Rips method.

Remark 1. For all $\epsilon \geq 0$, we have a simplicial complex, but in practice we will discretize time by choosing a finite number of values for ϵ. In this paper, we choose to work with all values of the form $\epsilon = t\rho$, where $t \in \{0, 1, \ldots, 100\}$ and ρ is a fixed constant, so all our filtrations and also barcodes will be at the same **scale**, as we will see in Sect. 2.3.

2.2   Musical Bars of a Score

The main idea of our construction is to consider that a **musical score** $\mathcal{S}$ is simply a set of its musical bars: $\mathcal{S} = \{\mathcal{B}_1, \mathcal{B}_1, \ldots, \mathcal{B}_n\}$, where the indexing corresponds to the musical flow. In this context, our definition of a bar is the following one:

Definition 2. *A **musical bar** is a finite subset $\mathcal{B}$ of $\mathbb{R}^3$ where an element of $\mathcal{B}$ is called a **note** characterized by three coordinates:*

– *the **position**, which refers to its place in the bar*
– *the **duration**, expressed in beats*
– *the **pitch**, which is the value of the note in term of its fundamental frequency*

Example 1. Here is an example of a musical bar coding. The first two coordinates are determined by using the meter and the third is coded in *midicent*, which is related to the position of the note in a piano. Notice that if there are some rests in the bar, we can ignore them because the information is already encoded in the position of the next note.

 $\qquad \mathcal{B} = \{(0, 1/2, 71), (1, 2, 69), (3, 1, 72)\}$

2.3   The Score as a Point Cloud

If X and Y are two non-empty subsets of a metric space (M, d), it is possible to define a metric d_H called the **Hausdorff distance** between X and Y by:

$$d_H(X, Y) = \max \left\{ \sup_{x \in X} d(x, Y) \ ; \ \sup_{y \in Y} d(X, y) \right\}$$

with $d(x, Y) = \inf_{y \in Y} d(x, y)$. An illustration of this metric is given in Fig. 5.
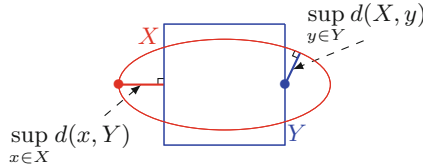

Fig. 5. Calculation of the Hausdorff distance between two metric spaces X (the ellipse) and Y (the square).

In our case, each musical bar of a score $\mathcal{S} = \{\mathcal{B}_1, \ldots, \mathcal{B}_n\}$ is a subset of $\mathbb{R}^3$ so we may naturally consider the Hausdorff distance between $\mathcal{B}_i$ and $\mathcal{B}_j$ for any i, j.

Definition 3. *Let $\mathcal{B}_i$ and $\mathcal{B}_j$ be two musical bars. The Hausdorff distance d_H between $\mathcal{B}_i$ and $\mathcal{B}_j$ is defined by*

$$d_H(\mathcal{B}_i, \mathcal{B}_j) = \max \left\{ \max_{n_i \in \mathcal{B}_i} \min_{n_j \in \mathcal{B}_j} d_1(n_i, n_j) \ ; \ \max_{n_j \in \mathcal{B}_j} \min_{n_i \in \mathcal{B}_i} d_1(n_i, n_j) \right\}$$

where $d_1(x, y) = \|x - y\|_1 = \sum_i |x_i - y_j|$.

Following up on Remark 1, we will consider the maximal distance $d_{\max}$ of all distances for a given score $\mathcal{S} = \{\mathcal{B}_1, \ldots, \mathcal{B}_n\}$ and define $\rho = \frac{d_{\max}}{100}$ as the **precision** we want to work with. We then consider for each $t \in \{0, 1, \ldots, 100\}$ the associated Vietoris-Rips complex $\mathcal{R}_{t\rho}(\mathcal{S})$. Furthermore, instead of speaking about "time t of the filtration", we will now say that we look at the filtration with an **error margin** of $t\%$. Indeed, we think of the presence of an edge between two musical bars as an indication that they are "similar", and the parameter t controls the way in which we choose to make this rigorous; for a small value of t, there are few edges which means that bars are finely distinguished, while for t large enough we allow coarser identifications.

We now have a point cloud from any musical score, so we have defined a new way to associate a filtered complex with a score by considering the Vietoris-Rips complex of Sect. 2.1. Thus, we are now able to compute persistent homology (barcodes) and the next section shows an example of a musical analysis using this approach. Recall the filtration and barcodes from Figs. 1 and 2 respectively; they correspond to the method we just defined applied to the little score with 4 distinct bars of Fig. 6, which was built arbitrarily to test it.

Fig. 6. The score from which we build filtration and barcodes of Figs. 1 and 2.

3   Application: Analysis of a Musical Piece

The idea is now to test our approach on a music piece to see in which way it is useful in the context of musical analysis. As we have already mentioned in the introduction, we are interested in the barcodes in degree 0 and 1. For computing homologies, Betti numbers and barcodes, we used our own algorithms and programs written with the help of the *SageMath* system (based on Python).

The musical piece we choose to analyse is taken from the soundtrack of the French movie *Le Fabuleux Destin d'Amélie Poulain*, directed by Jean-Pierre Jeunet (2001). The music is the famous *Comptine d'un autre été: l'Après-midi* for piano, composed and played by the minimalist composer Yann Tiersen (2001). The version we took for our analysis is extracted from [6].

The score has 53 muscial bars, some of which are repeated so it only contains 39 distinct bars, and it is split in two parts: actually, the music has 3 different themes in the first part that are played again one octave higher in the second one. These 3 themes and their respective structures are presented in Fig. 7. Notice that there are some repetitions of these themes in the original score, but here we suppress them in order to work with distinct musical bars only. Moreover, all the melody is constructed over 4 musical bars that are repeated and which constitute the musical accompaniment from Fig. 8.

Fig. 7. Part of each theme of *Comptine d'un autre été: l'Après-midi*. The first one goes from $\mathcal{B}_5$ to $\mathcal{B}_8$ and is repeated one octave higher from $\mathcal{B}_{22}$ to $\mathcal{B}_{25}$. The second goes from $\mathcal{B}_9$ to $\mathcal{B}_{12}$, then is repeated with extra notes from $\mathcal{B}_{13}$ to $\mathcal{B}_{16}$ and the whole 8 bars are played one octave higher from $\mathcal{B}_{26}$ to $\mathcal{B}_{33}$. The third one goes from $\mathcal{B}_{17}$ to $\mathcal{B}_{21}$ and is repeated one octave higher from $\mathcal{B}_{34}$ to $\mathcal{B}_{38}$, with a slight change in $\mathcal{B}_{38}$ to bring us to the end of the piece.

Fig. 8. The musical accompaniment of *Comptine d'un autre été: l'Après-midi*. These 4 bars consists of 4 arpeggiated chords $Em - G - B - D$ and are played once at the beginning of the score without any melody, from $\mathcal{B}_1$ to $\mathcal{B}_4$.
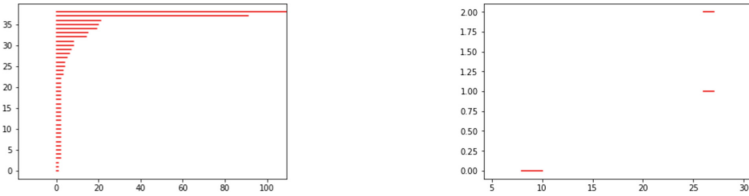

Fig. 9. Barcodes for *Comptine d'un autre été: L'Après-midi* in degree 0 (left) and degree 1 (right).

Let us look at the barcode in degree 0 from Fig. 9: there are several levels of analysis depending of the error margin we choose to take, and the main idea of persistent homology is to focus on the largest bars (those which *persist*), while the smallest ones can be considered as *noise*. In our case, there are 2 bars that stand out when we take an error margin larger than 21%, that means that the corresponding complex has only 2 connected components. One of them corresponds to the last musical bar $\mathcal{B}_{39}$ of the score, which only consists of the final chord Em played with whole notes, and the other is a large dimensional complex where all the musical bars are connected together. This first analysis shows that the barcode in degree 0 separates the end from the rest of the piece, which is a start. For $t\%$ with $t \leq 8$, there are only small bars so we ignore them as noise. Between 8% and 21%, there are 5, 6 or 7 classes that seem to last and more precisely, we found that with an error margin of 14%, the associated complex in the filtration looks like in Fig. 10, which is really remarkable: actually, there are 6 connected components and if we look at the vertices, we see that each one corresponds to a theme of the song, except for $\mathcal{B}_8$ and $\mathcal{B}_{28}$ which have a slight different structure than the rest of the first theme.
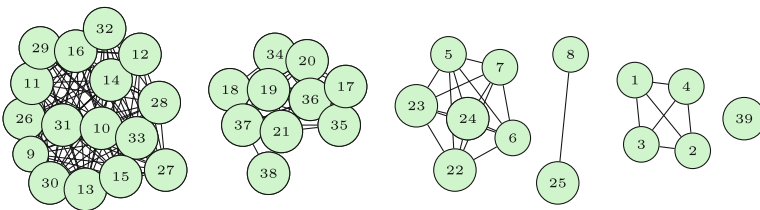

Fig. 10. The associated complex of *Comptine d'un autre été: l'Après-midi* with an error margin of 14%: each component characterizes a theme of the piece.

We may then conclude that, with a well-chosen error margin, the barcode in degree 0 captures the structure of this piece by separating its different themes in the associated complex.

On the other hand, barcode in degree 1 displays 3 different one-dimensional cycles that are presented in Fig. 11. Note that some edges of these cycles linked musical bars of one given theme to the same one octave higher, but not systematically and for now we are not able to interpret these cycles musically.
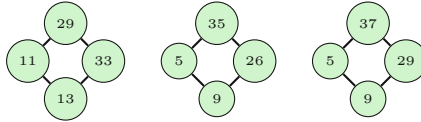

Fig. 11. One-dimensional cycles from *Comptine d'un autre été: L'Après-midi* that appear with an error margin of 8% (left) and 26% (middle and right).

4   Conclusion and Prospect

This paper has presented a new method to provide a filtered complex associated with a musical piece. Moreover, our construction reveals interesting results on the piece we chose to study. Actually, it shows that there is a way to capture the global structure of the piece by using barcodes in degree 0. In contrast, barcodes in degree 1 did not display an obvious musical interpretation, and we plan to focus on this dimension in our future work. In fact, one-dimensional cycles could be related to repeating patterns or musical loops in the score, and we are working on highlighting this interpretation from our construction. We also plan to apply our approach to a more general and diverse corpus of music data to see in which way it can capture the global structure of a music piece.

References

1. Bergomi, M.G., Baratè, A.: Homological persistence in time series: an application to music classification. J. Math. Music **14**(2), 204–221 (2020)
2. Bergomi, M.G., Baratè, A., Di Fabio, B.: Towards a topological fingerprint of music. In: Bac, A., Mari, J.-L. (eds.) CTIC 2016. LNCS, vol. 9667, pp. 88–100. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39441-1_9
3. Bigo, L., Andreatta, M.: Filtration of pitch-class sets complexes. In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O.A. (eds.) MCM 2019. LNCS (LNAI), vol. 11502, pp. 213–226. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_17
4. Ghrist, R.: Barcodes: the persistent topology of data. Bull. Amer. Math. Soc. **45**(1), 61–75 (2008)
5. Sethares, W.A., Budney, R.: Topology of musical data. J. Math. Music **8**(1), 73–92 (2014)
6. Tiersen, Y.: Comptine d'un autre été : L'après midi. In: Yann Tiersen, Six pièces pour piano, Vol. 2, pp. 5–7. Id Music (2004)
7. Zomorodian, A., Carlsson, G.: Computing persistent homology. Discrete Comput. Geom. **33**(2), 249–274 (2005)

Formal Structures of a Harmony in the Parabola

Edgar Armando Delgado Vega$^{(\boxtimes)}$ [ORCID]

15021 Lima, Peru
edelve91@gmail.com

Abstract. We develop a geometric analog of musical harmony from the group law of the affine parabola. First, we associate musical notes and intervals with points of a parabola. Immediately, we can define the usual affine and linear transformations for musical chords in module theory. Subsequently, we show that the actions of the groups T/I in PK-nets, PLR, $UTTs$, and JQZ behave identically to the circle space. Then, we propose to recreate the Planet-4D model, the study of musical distance and the DFT for subsets of points on the parabola. We believe that we have an innovative and motivational perspective to approach the parabola in a musical meaning.

Keywords: Parabola · Group law · Pitch-class set theory · Affine transformations · Neo-Riemannian theory · Music Fourier space

1 Introduction

Due to the cyclical nature of musical objects, the circle is a conventional locus to represent them. However, it is possible to define for a parabola $\mathcal{P}$ the finite ring geometric structure $\mathcal{P}(\mathbb{Z}/n\mathbb{Z})$ that behaves similarly to the classical circular pitch class space. We will see that the harmony of the circle is a kind of base layer to the harmony of the parabola. Thus, inspired by the isomorphic structure $\mathcal{P}(\mathbb{Z}/n\mathbb{Z}) \cong (\mathbb{Z}/n\mathbb{Z}, +)$ proved in [15], we will bijectively associate each point of the affine parabola $y = x^2$ with a musical note or interval of the chromatic scale.

2 The Group Law on the Parabola

Let Λ be a commutative ring with unity. The group law on the parabola $\mathcal{P}(\Lambda) = \{y = x^2 : x, y \in \Lambda\}$ is defined by taking a fixed point as a neutral element (the vertex of the parabola) which we denote by $N = (x_N, x_N^2) = (0, 0)$. Now, let $P = (x_P, x_P^2)$ and $Q = (x_Q, x_Q^2)$ be any two points on the parabola $\mathcal{P}$. The sum $P \oplus Q = R = (x_R, x_R^2)$ is the point of intersection with the parabola of the line parallel to PQ passing through vertex N. Algebraically, the addition of the group of points is given by

$$P \oplus Q = (x_P, x_P^2) + (x_Q, x_Q^2) = (x_P + x_Q, (x_P + x_Q)^2). \tag{1}$$

The proof of the group axioms in Definition (1) can be found in Lemmermeyer [15, p. 42] (see Shirali [18, pp. 31−32] for a general definition whose neutral element is any point on $\mathcal{P}$).

3   Harmonic Polygons over a Parabola

For abstract musical purposes, we are interested in associating points to notes and intervals of the chromatic scale. Thus, we have $\mathcal{P}(\mathbb{Z}_{12}) = \{(0,0),(1,1),(2,4),(3,9),(4,4),(5,1),(6,0),(7,1),(8,4),(9,9),(10,4),(11,1)\}$.
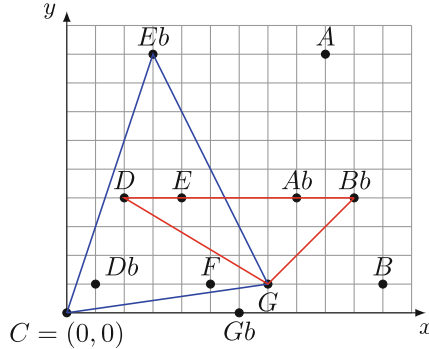

Fig. 1. Neo-Riemannian function PR, transposition by an interval of seven semitones $+(7,1)$, transform C-minor triad (blue polygon) to G-minor triad (red polygon). (Color figure online)

In Fig. 1 we describe harmonic progressions by drawing polygons on the parabola. From a geometric and metaphorical perspective, the abscissa is the base layer of the harmony in the circle $\mathbb{Z}_{12}$, while the ordinate is the harmonic layer which belongs to the parabola. We could express this idea as $(x,y) = (\text{circle}, \text{parabola})$. Therefore, the y-coordinate can be understood as a harmony added or attached to the harmony of the circle that corresponds to the x-coordinate. For example, the C-minor triad in the circle harmony $x = \{C, Eb, G\}$ has the parabolic layer $y = \{C, A, C\sharp\}$. With this interpretation we also have two sets of intervals: between the notes of the ordinates and, between the layers of the circle and the parabola. Furthermore, this point of view allows the algoritmic composition if we consider affine transformations, e.g. $y = 2x^2 + 1$, where the C-minor triad in the parabola varies to $y = \{C\sharp, G, D\sharp\}$.

4   The Ring and Field Law on the Parabola

For the rest of the analogous definitions we need a richer structure than a group. Thus, to the group $\mathcal{P}(\mathbb{Z}_n)$ we can also equip the structure of a finite ring with unity by the multiplication operation

$$P * Q = (x_P, x_P^2) * (x_Q, x_Q^2) = (x_P \cdot x_Q, x_P^2 \cdot x_Q^2). \tag{2}$$

Proposition 4.1. *The set of points* $(\mathcal{P}(\mathbb{Z}_n), \oplus, *)$ *with addition and multiplication defined by (1) and (2) forms a commutative ring with unity point* $(1,1)$.

The proof of Proposition 4.1 is straightforward if the projection on the x-axis is established also for the multiplication. Definition (2) follows from the geometric operation [15, p. 56] which equip a field for the parabola over $\mathbb{Q}$, and we are considering the ring $\mathcal{O}_{\mathcal{P}(\mathbb{Q})}$ of such rational field. Take the fixed point $M = (1,1)$. Let us draw a line between two points P and Q and see the intersection point R with the y-axis. Then, let us choose the intersection $S = P*Q$ of the line through R and M over $\mathcal{P}(\mathbb{Q})$. This field structure can extend the possibility of also modeling continuous spaces from a physical perspective of music if we take $\mathcal{P}(\mathbb{R})$.
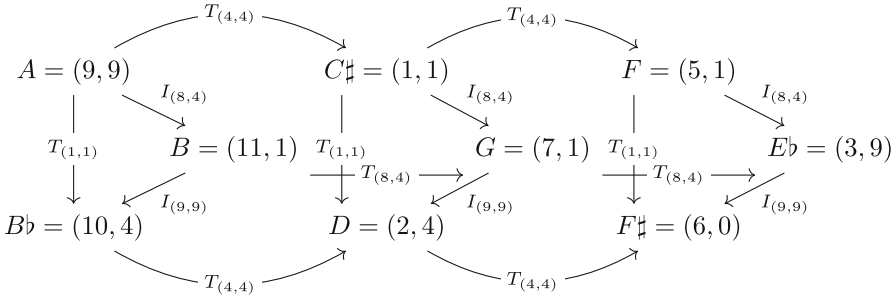
5   Parabola over a Module and Affine Transformations

Since $\mathcal{P}(\mathbb{Z}_n)$ forms a commutative ring by Definitions (1) and (2), we can observe it as a module over itself $_{\mathcal{P}(\mathbb{Z}_n)}\mathcal{P}(\mathbb{Z}_n)$ or a module with scalar action $[n] \in \mathbb{Z}_n$ given by $\cdot : \Lambda \times \mathcal{P}(\Lambda) \longrightarrow \mathcal{P}(\Lambda)$, $([n], P) \longmapsto [n] \cdot P = \underbrace{P \oplus P \oplus P \oplus \cdots \oplus P}_{[n]\text{-times}}$.

Proposition 5.1. *The points of the parabola over* $\mathbb{Z}_n$ *with addition and scalar action form a* $\mathbb{Z}_n$*-module* $\mathcal{P}(\mathbb{Z}_n)$.

The proof of Proposition 5.1 is straightforward. With this structure on the parabola $\mathcal{P}(\mathbb{Z}_{12})$, we can transform D-major triad into D-aug triad under a morphism that takes $((2,4),(6,0),(9,9)) \mapsto ((2,4),(6,0),(10,4))$, i.e., $\varphi : (P,Q,R) \mapsto (P,Q,[2]Q - P)$. In fact, we can rewrite all affine homomorphisms common in music theory. For instance, following [3], symmetries of consonance and dissonance in counterpoint, e.g. $e^{(2,4)}[5]((3,9)) = (5,1)$. If we consider the ring structure, we can represent counterpoint intervals as linear polynomials in $\mathcal{P}(\mathbb{Z}_{12})[X]$, for example a minor third $(7,1) \oplus (3,9)X$.

6   Group Actions over Parabolic Music

The musical groups T/I [10], PLR [6,7,9] and $UTTs$ [13] can act in the usual way on sets of points of the musical parabola $\mathcal{P}(\mathbb{Z}_{12})$. Consider first the elements in the T/I group that reveal underlying symmetries between notes of chords in PK-nets [17]. Transposition of a note Q of the parabola is defined as $T_P(Q) = P \oplus Q$, while inversions is given by $I_P(Q) = -Q \oplus P$. Thus, rewriting the musical PK-net analysis in [17, p. 36], we have

$$A = (9,9) \qquad\qquad C\sharp = (1,1) \qquad\qquad F = (5,1)$$

with diagram arrows labeled $T_{(4,4)}$, $I_{(8,4)}$, $T_{(1,1)}$, $B = (11,1)$, $G = (7,1)$, $E\flat = (3,9)$, $T_{(8,4)}$, $I_{(9,9)}$, $B\flat = (10,4)$, $D = (2,4)$, $F\sharp = (6,0)$, $T_{(4,4)}$.

In the context of Neo-Riemannian theory, the composition PR acts on the E-minor chord as a $\{(4,4),(7,1),(11,1)\} \mapsto \{(7,1),(10,4),(2,4)\}$. In fact, it can be generalized on a simplicial Tonnetz model [5,22] as it is observed in Fig. 1. Suppose we have an unfolded space $\mathcal{K}[2,4,6]$, then

$$PR \cdot \{(4,4),(6,0),(10,4)\} \mapsto \{(6,0),(8,4),(0,0)\}.$$

Similarly, we can reinterpret the uniform triadic transformation of E-major triad to the A-minor triad:

$$((4,4),+) \xrightarrow{\;U=(-,(5,1),(10,4))\;} ((9,4),-).$$

Another group action, in this case non-contextual, that we can use for $\mathcal{P}(\mathbb{Z}_{12})$ is JQZ [14] redefining $J = I_{(7,1)}, Q = I_{(11,1)}$ and, $Z = I_{(4,4)}$. Then,

$$ZJZ \cdot \{(5,1),(8,4),(0,0)\} \mapsto \{(1,1),(5,1),(8,4)\}.$$

On the other hand, it would be interesting to explore algebraic or formal relationships in a three-dimensional Tonnetz [12] or in a Cube Dance [8].

7   Parabolic Planet-8D and Metric

The points of a parabola behave similarly to their numerical analogues as we can observe in the conmutative diagrams below. For a field $K = \mathbb{Q}, \mathbb{R}$, or $\mathbb{C}$ and a ring $R = \mathbb{Z}, \mathbb{Z}_n$, following [15, p. 42], the morphisms ϕ_x and ψ_x can be understood as an injection into $y = x^2$, or as a geometrical projection on the x-axis. This properties would allow us to define a metric that emulates voice leading definitions [19] or the related problems for a multi-set metric [11].

$$
\begin{array}{ccc}
\mathcal{P}(K) \longrightarrow \mathcal{P}(R) & \qquad (x,x^2) \longrightarrow (x \mod n, x^2 \mod n) \\
\phi_x \uparrow \qquad \psi_x \uparrow & \qquad \downarrow \phi_x^{-1} \qquad\qquad \downarrow \psi_x^{-1} \\
K \longrightarrow R & \qquad (x) \longrightarrow (x \mod n)
\end{array}
\;.
$$

Now let us define the following isomorphism through the decomposition into direct sums of groups: $\mathcal{P}(\mathbb{Z}_{12}) \cong \mathbb{Z}_{12} \cong \mathbb{Z}_3 \bigoplus \mathbb{Z}_4 \cong \mathcal{P}(\mathbb{Z}_3) \bigoplus \mathcal{P}(\mathbb{Z}_4)$.

One of the models for visualization of harmonic relationships between pitch classes is Planet-4D [4]. We see the formal possibility of reconstructing the model in a space of four complex dimensions $\mathbb{C}^2 \times \mathbb{C}^2$. We define the same isomorphism of the direct product of cyclic groups and roots of unity but modified over the points of the parabola under multiplication. Thus, we have the isomorphisms

$$\mathcal{P}(\mathbb{Z}_3) \cong \{(1,1), (e^{\frac{2\pi i}{3}}, e^{\frac{-2\pi i}{3}}), (e^{\frac{-2\pi i}{3}}, e^{\frac{2\pi i}{3}})\}.$$

$$\mathcal{P}(\mathbb{Z}_4) \cong \{(1,1), (i, -1), (-1, 1), (-i, -1)\}.$$

Consider the F note associated with the element $(2,1) \in \mathbb{Z}_3 \times \mathbb{Z}_4$. Then, on the parabolic planet we have $(2,1) \cong (2,1,1,1) \cong (e^{\frac{-2\pi i}{3}}, e^{\frac{2\pi i}{3}}, i, -1)$. The bijection of an element of the direct product $\mathcal{P}(\mathbb{Z}_3) \times \mathcal{P}(\mathbb{Z}_4)$ to return to the parabola $\mathcal{P}(\mathbb{Z}_{12})$ is defined in imitation of [2] by sending the points $(P, Q) \mapsto 4P - 3Q$.

8   The Discrete Fourier Transform in a Parabolic World

The importance of Discrete Fourier Transform for the mathematical music theory is due to the fact that it helps to reveal hidden periodic qualities behind subsets of rhythms and scales [1]; even analyze harmony from a geometric perspective [20,21]. The DFT is built over a space of distributions $\mathbb{C}^{\mathbb{Z}_n}$. The analog for points P_i of the parabola is defined by the function $\mathcal{P}(\mathbb{Z}_{12}) \to \mathbb{C}^{2n}$, $f \mapsto (f(P_0), f(P_1), \ldots, f(P_{n-1}))$. Thus, we define the DFT of a subset of points $P \subset P(\mathbb{Z}_n)$ as the transformation of its characteristic function

$$\mathcal{F}_P = \widehat{f}_k = \sum_{x_P, x_P^2 \in P} \left(e^{\frac{-2\pi i k x_P}{n}}, e^{\frac{-2\pi i k x_P^2}{n}} \right). \tag{3}$$

Note that in the Definition (3) the sum is parabolic. For example, let $P = \{C, E\flat, G\flat, B\flat\flat\}$, the fourth Fourier coefficient of P produces $\widehat{f}_4 = (1,1)+(1,1)+(1,1)+(1,1) = (4,4)$. It is immediate to rewrite the convolution product for a set of points of a parabola, which mathematically describes musical operations such as multiplication of Boulez chords, intervallic content or rhythmic canons. Let f, g be characteristic functions, i.e., $f = (P_0, P_1, \ldots, P_{n-1})$, of subsets P and Q, respectively. The circular convolution is given by

$$f * g(k) = \sum_{n_{PQ}=0}^{n_{PQ}-1} f(k - n_{PQ})g(n_{PQ}), \tag{4}$$

for all $k \in \mathbb{Z}_{12}$. Note that n_{PQ} is the indexed position of the points in f, g. It follows analogously from (3) and (4) the identity that relates convolution and DFT for each k, $\widehat{f * g}(k) = \widehat{f}(k)\widehat{g}(k)$.

9   Conclusions

We have seen that musical harmony can be represented as polygons in a parabola. Although the geometry of the parabola, seen as a layer on top of the harmony of the circle, operates analogously in analytic approaches, it is possible to extend this perspective to layers defined by other equations maintaining a base formal structure, even with more dimensions, e.g. two ellipses whose integral points are isomorphic to the direct product $\mathbb{Z}_3 \times \mathbb{Z}_4$. In that sense, the arithmetic and geometric aspects of the affine transformations on $\mathcal{P}(\mathbb{Z}_n)$ and other curves can serve as a locus to generate musical ideas embedded in a mathematical environment. With regard to future research, the development of new geometric approaches to music theory can inspire technological and computational advances [16], which might also lead to new software developments for teaching and composition.

References

1. Amiot, E.: Music Through Fourier Space: Discrete Fourier Transform in Music Theory. Springer, Cham (2016)
2. Amiot, E., Baroin, G.: Old and new isometries between pc-sets in the Planet-4D model. Music Theor. Online, **21**(3), 2 (2015)
3. Arias-Valero, J.S., Agustín-Aquino, O.A., Lluis-Puebla, E.: On first-species counterpoint theory (2020). arXiv:2004.07983
4. Baroin, G.: The Planet-4D model: an original hypersymmetric music space based on graph theory. In: Agon, C., Andreatta, M., Assayag, G., Amiot, E., Bresson, J., Mandereau, J. (eds.) MCM 2011. LNCS (LNAI), vol. 6726, pp. 326–329. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21590-2_25
5. Bigo, L., Ghisi, D., Spicher, A., Andreatta, M.: Representation of musical structures and processes in simplicial chord spaces. Comput. Music J. **39**(3), 9–24 (2015)
6. Cohn, R.: Neo-Riemannian operations, parsimonious trichords, and their "Tonnetz" representations. J. Music Theor. **41**(1), 1–66 (1997)
7. Crans, A.S., Fiore, T.M., Satyendra, R.: Musical actions of dihedral groups. Am. Math. Mon. **116**(6), 479–495 (2009)
8. Douthett, J., Steinbach, P.: Parsimonious graphs: a study in parsimony, contextual transformations, and modes of limited transposition. J. Music Theor. **42**(2), 241–263 (1998)
9. Fiore, T.M., Noll, T.: Voicing transformations of triads. SIAM J. Appl. Algebra Geom. **2**(2), 281–313 (2018)
10. Forte, A.: The Structure of Atonal Music, vol. 304. Yale University Press, New Haven (1973)
11. Genuys, G.: Pseudo-distances between chords of different cardinality on generalized voice-leading spaces. J. Math. Music **13**(3), 193–206 (2019)
12. Gollin, E.: Some aspects of three-dimensional "Tonnetze". J. Music Theor. **42**(2), 195–206 (1998)
13. Hook, J.: Uniform triadic transformations. J. Music Theor. **46**(1–2), 57–126 (2002)
14. Jedrzejewski, F.: Non-contextual JQZ transformations. In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O.A. (eds.) MCM 2019. LNCS (LNAI), vol. 11502, pp. 149–160. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_12

15. Lemmermeyer, F.: Pell Conics, An Alternative Approach to Elementary Number Theory (2012). https://www.mathi.uni-heidelberg.de/~flemmermeyer/pell/bfc02.pdf. Accessed 11 Jan 2022

16. Mannone, M., Kitamura, E., Huang, J., Sugawara, R., Chiu, P., Kitamura, Y.: Cubeharmonic: a new musical instrument based on Rubik's cube with embedded motion sensor. In: ACM SIGGRAPH 2019 Posters. SIGGRAPH 2019, Association for Computing Machinery, New York (2019)

17. Popoff, A., Andreatta, M., Ehresmann, A.: Groupoids and wreath products of musical transformations: a categorical approach from poly-klumpenhouwer networks. In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O.A. (eds.) MCM 2019. LNCS (LNAI), vol. 11502, pp. 33–45. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_3

18. Shirali, S.A.: Groups associated with conics. Math. Gaz. **93**(526), 27–41 (2009)

19. Tymoczko, D.: Three conceptions of musical distance. In: Chew, E., Childs, A., Chuan, C.-H. (eds.) MCM 2009. CCIS, vol. 38, pp. 258–272. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02394-1_24

20. Tymoczko, D., Yust, J.: Fourier phase and pitch-class sum. In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O.A. (eds.) MCM 2019. LNCS (LNAI), vol. 11502, pp. 46–58. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_4

21. Yust, J.: Applications of DFT to the theory of twentieth-century harmony. In: Collins, T., Meredith, D., Volk, A. (eds.) MCM 2015. LNCS (LNAI), vol. 9110, pp. 207–218. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20603-5_22

22. Yust, J.: Generalized Tonnetze and Zeitnetze, and the topology of music concepts. J. Math. Music **14**(2), 170–203 (2020)

midiVERTO: A Web Application to Visualize Tonality in Real Time

Daniel Harasim[1(✉)], Giovanni Affatato[2], and Fabian C. Moss[1,3]

[1] École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
daniel.harasim@epfl.ch
[2] Politecnico di Milano, Milan, Italy
giovanni.affatato@mail.polimi.it
[3] University of Amsterdam, Amsterdam, The Netherlands
fabian.moss@uva.nl

Abstract. This paper presents a web application for visualizing the tonality of a piece of music—the organization of its chords and scales—at a high level of abstraction and with coordinated playback. The application applies the discrete Fourier transform to the pitch-class domain of a user-specified segmentation of a MIDI file and visualizes the Fourier coefficients' trajectories. Since the coefficients indicate different musical properties, such as triadicity and diatonicity, the application isolates aspects of a piece's tonality and shows their development in time. The aim of the application is to bridge a gap between mathematical music theory, musicology, and the general public by making the discrete Fourier transform as applied to the pitch-class domain accessible without requiring advanced mathematical knowledge or programming skills up front.

Keywords: Web application · Visualization · Discrete Fourier transform · Tonality · MIDI

1 Introduction

Music analysis requires a high degree of expertise to derive insights about latent tonal structures in a composition. Mathematical music theory provides analytical tools by generalizing from concrete pieces to more abstract concepts and definitions. For example, many interesting musical entities can be described as subsets of cyclic groups, such as chords, scales, and repeating rhythms. In recent years it became increasingly popular among music theorists to study such entities by applying the discrete Fourier transform (DFT) to the pitch domain [3,10,12], in contrast to the time domain to which the DFT is most commonly applied (e.g., in signal processing and music information retrieval). For a general description of

this method's mathematical details and music-theoretical interpretations see for instance [2]. This usage of the DFT essentially maps subsets of cyclic groups to six complex numbers (i.e., the Fourier coefficients 1 to 6) and many interesting properties, such as evenness [1,4], balancedness [7], and diatonicity [13], can be identified and studied with the Fourier coefficients.

Since the DFT can be applied to music in symbolic formats (e.g., MIDI) without prior interpretation of the musical material by a music theorist, it is well suited for distant-reading approaches in corpus studies, such as the comparison of different pieces' tonal organization at a high level of abstraction [13]. However, this requires advanced mathematical and computational skills, which often hinders music historians, students, and enthusiasts from engaging with it. Harnessing the power of such complex methods thus remains restricted to only a small group of researchers.

This paper presents the open-source web application midiVERTO which addresses this problem and aims to bridge the intradisciplinary methodological divide between mathematical music theory, musicology, and the general public. The application enables its users to visualize the tonality of a piece of music viewed through the lens of the DFT without the requirement to understand its formal details. The app thus contributes to an exchange of knowledge and techniques for music analysis between experts and a broader audience. In this spirit, we continue the work by Thomas Noll [8] and Jennifer Harding [5] to make DFT analyses of musical entities more easily accessible and reproducible for a broader readership, especially students.

2   midiVERTO: Features and Technical Details

A screenshot of the application is shown in Fig. 1.[1]. Apart from a welcome page ("Home") and a step-by-step tutorial ("Docs"), the application features an interface consisting of four parts ("Analysis"): an input-output menu on the left, two visualization panels in the center, and a control menu at the bottom.

To visualize the tonal content of a piece encoded in MIDI format, a user can upload the MIDI file and select a duration value that is used to segment the music. For instance, a value of a quarter note leads to a division into many quarter-note segments. Durations can be specified either in note values or in seconds. The application then creates two interlinked visualizations. Note that the user-specified segmentation is only relevant for the visualization in the lower panel. For the one in the upper panel, a fixed segmentation is used to avoid unnecessarily long computation times.

The first visualization shows six *wavescapes* [11]. These are triangular plots closely related to keyscapes [9] and pitch scapes [6]. The k-th wavescape shows the values of the k-th Fourier coefficient in time and at different time scales. Magnitudes of the Fourier coefficients are indicated by opacity and their phases are shown according to a circular color mapping described below, together with the second visualization. The bottom of a wavescape corresponds to the respective coefficient's values for each segment. All other horizontal slices take groups

---

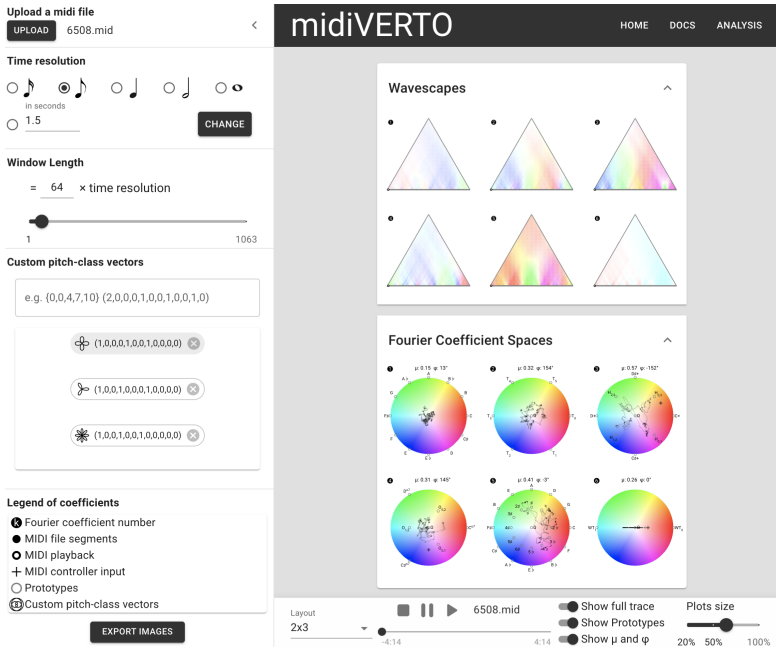[1] The app is accessible at https://dcmlab.github.io/midiVERTO.

Fig. 1. Screenshot of the web application midiVERTO with three components: input-output on the left, visualizations in the center, and control options at the bottom.

of segments into account. The higher the slice, the more segments are considered. The high levels of the k-th wavescape thus show the values of the k-th Fourier coefficient on large time scales, and the tip of each triangle represents the value of the respective Fourier coefficient for the pitch-class content of the entire piece. The coefficients have different interpretations, for instance triadicity and diatonicity for the third and fifth coefficients, respectively [13].

In contrast to the wavescape visualization, which is static, the second visualization is dynamic and coordinated with the playback of the MIDI file. For each of the six Fourier coefficients, a unit disk in the complex plane is shown and each angle is assigned a color in a circular color mapping. We call these disks *Fourier coefficient spaces*. Each disk shows the positions of prototypical pitch-class sets to provide orientation in the space. Those are sets that have high magnitudes in the respective coefficient, such as augmented triads in the third and diminished seventh chords in the fourth coefficient.

Each time point of the uploaded piece is assigned to six complex numbers. Similar to the wavescape visualization, the coefficient-space visualization can show values according to different time scales, but in contrast to wavescapes only with respect to one time scale at a time. A time scale corresponds to the length of a sliding window over segments and this length can be set by the user in the input-output menu. For each sliding window, all contained pitch classes are summed into one 12-dimensional pitch-class count vector which is subsequently normalized (L^1 norm) and mapped under the Fourier transform.

The normalization ensures that all coefficients' values are contained inside the unit circle. Very short window lengths tend to distribute the segments of the piece more evenly throughout the disks. In contrast, long window lengths collapse the segments into a small area, and even into a single point if the window covers the entire piece (i.e., the topmost point of a wavescape). Therefore, it is beneficial to try different window lengths in order to identify an appropriate level of detail for the concrete piece under analysis, for instance one where the piece moves along smooth paths in some coefficient spaces.

The user can play, pause, and skip through the MIDI file using the control menu at the bottom of the app. While the piece is being played, the coefficient values of the current sliding window are displayed as white dots in the coefficient spaces. The control menu contains additional options, such as for hiding the prototypical pitch-class sets and for the adjustment of the visualizations' sizes and layouts. A user can also manually input pitch-class multisets or distributions into midiVERTO. Those can be either typed into a text field in the input-output menu or played through an external MIDI controller, such as a keyboard or music notation software. This feature can be particularly useful to compare specific chords and scales of interest, for example in educational contexts.

midiVERTO was developed as a purely browser-based application in order to minimize maintenance requirements and to facilitate a wide adoption by the community. That is, no installation is required and no server application needs to be maintained since the application is entirely client-based. This also allows the application to run across platforms and lowers the entry barrier for users that might not be confident with technical installation instructions. The application is written in javascript using the libraries react for state management, material-ui for adaptive interface design, and Canvas API and SVG markup for visualization. The source code is publicly available on Github under a GPL3 licence, and the Github Pages service is used to deploy the application.[2]

3   A Brief Case Study

To showcase midiVERTO's capabilities, we use the main theme of the musical *Phantom of the Opera* by A. L. Webber with a resolution of an 8th note and a window length of 300 times the resolution.[3] Each point thus represents a length of 37.5 whole notes. With these parameters, several important aspects of the piece's global tonality are visualized intuitively in the wavescapes and the Fourier coefficient spaces. The wavescapes 3 and 5 have the strongest colors among all 6, indicating that diatonic scales as well as (augmented) triads play and important role for the overall harmonic organization of the piece. Therefore, we focus only on these two Fourier coefficients (see Fig. 2).

On the lower levels of the 3rd wavescape, the colors blue – green – yellow – pink (–green–) pink change in regular time intervals, following the primary key changes of the piece. Tracing this sequence of colors in the third coefficient

---

[2] Source code available at https://github.com/DCMLab/midiVERTO.

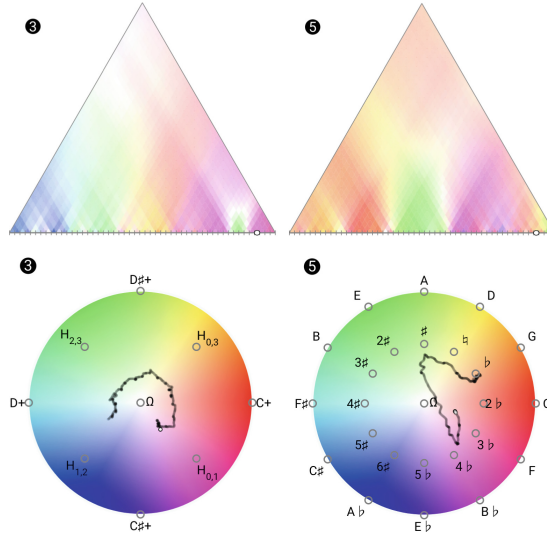[3] MIDI file taken from https://bitmidi.com/uploads/6508.mid.

Fig. 2. Visualizations of the 3rd and the 5th Fourier coefficient for the *Phantom of the Opera* theme (resolution: 8th note, time window: $300 \times \text{resolution} = 37.5$ whole notes). The unit disks are labeled with their prototypes: augmented triads are denoted with a +, hexatonic scales by two contained pitch classes (e.g., $H_{2,3} = \{2, 3, 6, 7, 10, 11\}$), singleton pitch classes by a representative, and diatonic scales by their key signatures. (Color figure online)

space, we observe that these colors correspond to the hexatonic scales $H_{1,2}$, $H_{2,3}$, $H_{0,3}$, and $H_{0,1}$, respectively (in clockwise direction; see figure caption for explanation of notation). The overall harmonic trajectory of the piece thus moves in descending minor thirds, and at the end it briefly moves across the plane to its hexatonic pole (the green interruption in the pink area).

The color progression on the lower levels of the 5th wavescape is roughly orange/red - green - pink, showing that the initial, middle, and final parts of this piece modulate through different keys. Note that the red and pink areas are adjacent to each other but opposite to the green area, again showing a symmetrical organization in terms of keys. Analogous observations can be made by following the path in the 5th coefficient space, because the window size of 300 eighth notes is adjusted to the required abstraction level.

4   Conclusion

Interactive applications like midiVERTO can help to promote mathematical tools for music analysis among a broader audience, such as music theorists, students, and practitioners. They can guide musical intuition and lead to new ways of hearing without the need of studying a complex methodology up front. Furthermore, they can spark interest, generate curiosity, and inspire to learn more

about the inner workings of techniques such as the DFT. Moreover, we expect that the application will enable scholars and students of music to employ this powerful method in their work and teaching, and for public engagement. Beyond purely music-theoretical contexts, the application is also suited to teach and study visualization techniques for complex cultural data in the Digital Humanities and Cultural Analytics.

References

1. Amiot, E.: David Lewin and maximally even sets. J. Math. Music **1**(3), 157–172 (2007). https://doi.org/10.1080/17459730701654990
2. Amiot, E.: Music Through Fourier Space: Discrete Fourier Transform in Music Theory. Computational Music Science, Springer International Publishing (2016). https://doi.org/10.1007/978-3-319-45581-5
3. Amiot, E.: Entropy of fourier coefficients of periodic musical objects. J. Math. Music **15**(3), 235–246 (2021). https://doi.org/10.1080/17459737.2020.1777592
4. Harasim, D., Noll, T., Rohrmeier, M.: Distant neighbors and interscalar contiguities. In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O.A. (eds.) MCM 2019. LNCS (LNAI), vol. 11502, pp. 172–184. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_14
5. Harding, J.D.: Computer-aided analysis across the tonal divide: cross-stylistic applications of the discrete fourier transform. In: Music Encoding Conference 2020 (2020)
6. Lieck, R., Rohrmeier, M.: Modelling hierarchical key structure with pitch scapes. In: Proceedings of the 21st International Society for Music Information Retrieval Conference, Montréal, pp. 811–818 (2020). https://doi.org/10.5281/zenodo.4245558
7. Milne, A.J., Bulger, D., Herff, S.A.: Exploring the space of perfectly balanced rhythms and scales. J. Math. Music **11**(2–3), 101–133 (2017). https://doi.org/10.1080/17459737.2017.1395915
8. Noll, T.: Insiders' choice: studying pitch class sets through their discrete fourier transformations. In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O.A. (eds.) MCM 2019. LNCS (LNAI), vol. 11502, pp. 371–378. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_32
9. Sapp, C.S.: Visual hierarchical key analysis. Comput. Entertainment **3**(4), 1–19 (2005). https://doi.org/10.1145/1095534.1095544
10. Tymoczko, D., Yust, J.: Fourier phase and pitch-class sum. In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O.A. (eds.) MCM 2019. LNCS (LNAI), vol. 11502, pp. 46–58. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_4
11. Viaccoz, C., Harasim, D., Moss, F.C., Rohrmeier, M.: Wavescapes: a visual hierarchical analysis of tonality using the discrete fourier transform. Musicae Scientiae (in press). https://doi.org/10.1177/10298649211034906
12. Yust, J.: Schubert's harmonic language and fourier phase space. J. Music Theor. **59**(1), 121–181 (2015). https://doi.org/10.1215/00222909-2863409
13. Yust, J.: Stylistic information in pitch-class distributions. J. New Music Res. **48**(3), 217–231 (2019). https://doi.org/10.1080/09298215.2019.1606833

Quantum-Musical Explorations on $\mathbb{Z}_n$

Thomas Noll[1][(✉)] and Peter Beim Graben[2]

[1] Departament de Creació i Teoria Musical, Escola Superior de Música de Catalunya, Barcelona, Spain
`thomas.mamuth@gmail.com`
[2] Bernstein Center for Computational Neuroscience Berlin, Berlin, Germany
`peter.beimgraben@b-tu.de`

Abstract. Motivated through recent applications of quantum theory to the music-theoretical conceptualisation of tonal attraction, the paper recapitulates basic facts about quantum wave functions over the finite configuration space $\mathbb{Z}_n$, and proposes a particular musical application.

After an introduction of position and momentum operators, the Fourier transform as well as the translation and ondulation operators, particular attention is plaid to the Quantum Harmonic Oscillator via its Hamilton operator and its eigenstates. In this setup the time development of chosen wave functions is applied to the control of moving sound sources in a Spatialisation scenario.

Keywords: Quantum theory · Music theory · Pitch class profiles

1 Motivation

A new quantum-theoretical approach to the study of musical tones (c.f. [2,7,9]) motivates the present attempt for an integration of other mathematical approaches into this new line of investigation. These new ideas may possibly open productive theoretical links between statistical approaches to music cognition on the one hand and structural mathematical approaches to music on the other. Up to now connections between these two areas are not yet highly sought-after and both areas suffer from deficiencies, which exhibit a remarkable complementary: Statistical approaches treat histograms of and transition matrices between possible musical events as these were already fully valid models of musical reality, while mathematicians build nice but somewhat empty spaces of musical objects, wherein no events actually happen. Under the quantum perspective pitch class profiles are interpreted as probability density functions of underlying quantum wave functions, which may inhabit the "empty" spaces of the mathematical music theorists. And this entails the possibility to gain explanatory power for the constitution of empirically derived pitch class profiles from these wave functions alongside with the Hermitian and unitary operators acting on them, and last not least from the Schrödinger equation. Needless to say, that these wave functions are not intended to be interpreted in a literal physical way. The wave functions are defined on spaces of musical tones or higher musical objects, not in physical space.

The starting point for the new quantum-theoretical approach in [9] (and follow up papers [2,7]) is the modelling of *tonal attraction* by means of a suitable match of the Krumhansl-Kessler pitch class profiles. It turns out that the circle-of-fifths ordering of the twelve pitch classes allows to build such a match from a continuous wave function on $\mathbb{R}/\mathbb{Z}$ exemplifying cosine-similarity. The present study is further motivated by a potential conceptual bifurcation within the quantum-theoretical framework. We observed, that there is an alternative possibility to match the Krumhansl-Kessler pitch class profiles, namely by starting from a Gaussian wave function on $\mathbb{R}$, which represents the ground state of a quantum harmonic oscillator.
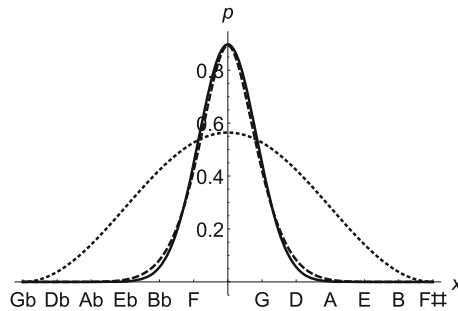

Fig. 1. Three attraction kernels $p(x) = |\psi(x)|^2$ of the Krumhansl-Kessler experimental data for C major (centered at tone C of the quint group), obtained from their respective wave functions $\psi(x)$. Solid: Gaussian wave packet, dashed: deformed cosine similarity, dotted: default cosine similarity.

In search of an analogy to the situation in physics we would view the (continuous) circle or line of fifths in the role of a configuration space for the "position representation" of wave functions ϕ. And consequently the question is on the table what the musical meaning of the associated "momentum representation" might be. We reflect about this question with the awareness that the Fourier Transform, which mediates between the two representations, already plays a productive role in recent approaches to the study of pitch classes and pitch class profiles (see [1,12–16]), for example. But while we deal with the Fourier Transform $\hat{\phi}$ of wave functions ϕ we would categorize the objects of study in these investigations as Fourier Transforms $\widehat{|\phi|^2}$ of probability density functions $|\phi|^2$. A second conceptual difference in these investigations consists in the finite configuration space $\mathbb{Z}_{12}$ as opposed to $\mathbb{R}/\mathbb{Z}$ or $\mathbb{R}$. But this difference is not an obstacle for an integration. The quantization result in [7] actually provides confidence into the suitability of the finite-dimensional approach also from within the quantum approach. Therefore, the most straight-forward first step towards an integration of the pre-established Fourier approach into the new quantum approach consists in the study and musical interpretation of wave functions on $\mathbb{Z}_n$. And particular attention has to be paid to the role of phases. And we approach this project

with the idea in mind to encode the parameter of *tone width* together with the parameter of *tone height* (as studied by David Clampitt and the first author in [3]) which is further inspired by Martin Ebeling's proposal to model musical tones on the complex plane [5].

The present poster is intended as a preparatory "mathe-musical warmup" with the purpose to get the Quantum theory on $\mathbb{Z}_7$ and/or $\mathbb{Z}_{12}$ at our fingertips. Here we postpone the search for answers to the motivating questions in favour of a plain sailing musical playground, where every wave function with moderate parameters and its time development for a given Energy operator can be realised and musically explored. In this scenario the parameters of amplitude and phase are interpreted in terms of the loudnesses and angular positions of a cycle of sound sources in a spatialisation scenario.

2 Quantum Theory on $\mathbb{Z}_n$

The mathematical foundations of quantum theory in n dimensions have been thoroughly investigated in recent years. We draw upon [4,6,10,11]. In this section we recapitulate elementary knowledge.

Quantum states are described in terms of wave functions $\psi : \mathbb{Z}_n \to \mathbb{C}$, which we will identify with vectors $\psi \in \mathbb{C}^n$. In the position representation the residue classes $0, ..., n-1 \in \mathbb{Z}_n$ denote positions, while they denote momenta in the momentum representation.
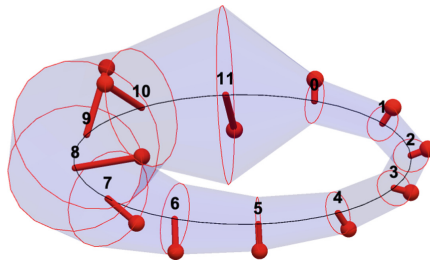

Fig. 2. Representation of a wave function over $\mathbb{Z}_{12}$. The lengths of the twelve needles represent amplitudes and their directions represent phases.

Linear operators $F : \mathbb{C}^n \to \mathbb{C}^n$ are represented through $n \times n$-matrices with complex coefficients, accordingly. The can represent active and passive transformations (i.e. active transformations of the wave functions themselves or passive coordinate transformations of one and the same wave function).

We start with the consideration of the *position operator Q*. Following [4] we define it as a diagonal $n \times n$ matrix Q_a with the n diagonal entries and eigenvalues $\{-\frac{a(n-1)}{2}, \ldots, \frac{a(n-1)}{2}\}$. The indices $j \in \{-\frac{(n-1)}{2}, \ldots, \frac{(n-1)}{2}\}$ are centered around 0 and are integers for odd n and half-integers for even n. The real scaling factor $a > 0$ is a length unit.

$$Q_a = a \cdot \begin{pmatrix} -\frac{n-1}{2} & 0 & \dots & 0 & \dots & 0 & 0 \\ 0 & -\frac{n-1}{2}+1 & \dots & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots\dots\dots & & \dots & \dots \\ 0 & 0 & \dots & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots\dots\dots & & \dots & \dots \\ 0 & 0 & \dots & 0 & \dots & \frac{n-1}{2}-1 & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 & \frac{n-1}{2} \end{pmatrix}$$

The normalized eigenstates of Q_a are the 'δ-functions" with precisely one non-vanishing coordinate, i.e.

$$\begin{aligned} \varphi_{-\frac{n-1}{2}} &= (1, 0, \dots, 0), \\ \varphi_{-\frac{n-1}{2}+1} &= (0, 1, 0, \dots, 0), \\ &\dots \\ \varphi_0 &= (0, \dots, 0, 1, 0, \dots, 0), \\ &\dots \\ \varphi_{\frac{n-1}{2}} &= (0, \dots, 0, 1). \end{aligned}$$

The exponential $M = exp(\frac{2\pi i}{n}Q)$ is known as the associated *Modulation-* or *Undulation operator*. M is an unitary operator and its n eigenvalues are either the n-th roots of unity or the odd $2n$-th root of unity. The determinant $det(M)$ is either 1 or -1. The latter happens, when n is even and $\frac{n-1}{2}$ the half of an odd number. The Fourier transform mediates between the position representation and the momentum representation of the wave functions, and it is therefore considered to be a passive transformation. Let $\omega(k) = exp(\frac{2\pi ik}{n})$, $k = 0, \dots, n-1$ denote the n'th root of unity. They form the coefficients of the Fourier transform:

$$F = \frac{1}{\sqrt{n}} \begin{pmatrix} \omega(0) & \omega(0) & \dots & \omega(0) & \dots & \omega(0) & \omega(0) \\ \omega(0) & \omega(-1) & \dots & \omega(-k) & \dots & \omega(2) & \omega(1) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \omega(0) & \omega(-k) & \dots & \omega(-k^2) & \dots & \omega(2k) & \omega(k) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \omega(0) & \omega(2) & \dots & \omega(2k)) & \dots & \omega(-4) & \omega(-2) \\ \omega(0) & \omega(1) & \dots & \omega(k) & \dots & \omega(-2) & \omega(-1) \end{pmatrix}$$

F is a unitary operator, satisfying $F^* = F^{-1}$. Its eigenvalues are $i, -1, -i, 1$. Their multiplicities depend on n and can be characterized in terms of the residue n mod 4 (see [10], p. 273).

The vectors $F \cdot \varphi_k$ are the momentum representations of the position eigenstates. Analogously we have a momentum operator P, and a basis of associated eigenstates $\phi_0, \phi_1, \dots \phi_{n-1}$, which in the momentum representation take the simple form 'δ-functions"

$$F \cdot \phi_0 = (1, 0, \dots, 0), \; F \cdot \phi_1 = (0, 1, 0, \dots, 0), \quad \dots \quad , F \cdot \phi_{n-1} = (0, \dots, 0, 1)$$

and the momentum operator in the momentum representation takes the diagonal form

$$F \cdot P \cdot F^* = \begin{pmatrix} 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & n-1 \end{pmatrix}.$$

The position representations of the eigenstates of P are the exponential circle functions:

$$\begin{aligned} \phi_0 &= \tfrac{1}{\sqrt{n}}(1,1,\ldots,1), \\ \phi_1 &= \tfrac{1}{\sqrt{n}}(1,\omega^1,\omega^2,\ldots,\omega^{n-1}), \\ &\quad \cdots \\ \phi_k &= \tfrac{1}{\sqrt{n}}(1,\omega^k,\omega^{2k},\ldots,\omega^{k(n-1)}), \\ &\quad \cdots \\ \phi_{n-1} &= \tfrac{1}{\sqrt{n}}(1,\omega^{-1},\omega^{-2},\ldots,\omega^{1}). \end{aligned}$$

The exponential $T = exp(-\frac{2\pi i}{n}P)$ is known as the *Translation operator*. T is a permutation matrix, and hence orthogonal (and hence unitary). Its n eigenvalues are the n-th roots of unity. The determinant $det(T)$ is either 1 (for odd n) and -1 (for even n).

$$T = \begin{pmatrix} 0 & 0 & 0\ldots & 0 & 0 & 1 \\ 1 & 0 & 0\ldots & 0 & 0 & 0 \\ 0 & 1 & 0\ldots & 0 & 0 & 0 \\ \ldots\ldots & \ldots & \ldots\ldots & & & \\ 0 & 0 & 0\ldots & 0 & 1 & 0 \end{pmatrix}.$$

T and M generate the *Heisenberg group*.

3   Exploring the Finite Quantum-Harmonic Oscillator

In the context of an ongoing investigation the authors found a suitable definition for a tonal attraction kernel in terms of a Gaussian wave function (as a possible alternative to the deformed cosine kernel in [9]). This finding brings the quantum harmonic oscillator into the spotlight of interest, whose ground state is a Gaussian. The mentioned investigations assume a continuous configuration space $\mathbb{R}$. But in connection with the already established Fourier approach in music theory it seems worthwhile to explore this important and well-studied physical example also in the finite-dimensional scenario. Although there is no analogue to the Schrödinger Equation, several constructions can be based on the study of Eigenvalues and Eigenfunctions. We start by inspecting the Hamilton operator $H = \frac{1}{2}(P^2 + a^2 Q^2)$ with parameter a (abstractly) measuring the impact of the potential energy against the normalised kinetic energy P^2. The excited states $\xi_0, \ldots, \xi_{n-1}$ can be obtained as the eigenfunctions of H and they can be ordered in accordance with the raising positive real eigenvalues of H. We observed in the case $n = 7$, that the choice of symmetric position eigenvalues (to both sides of 0) ensures that the excited states are also eigenfunctions of the Fourier transform

F, which in turn motivates the inspection of a finite analogue for the Bargmann transformation (e.g. [8], Sect. 14.4), where the excited states ξ_k are chosen as a basis and are mapped to the associated monomials $z \mapsto z^k : \mathbb{C} \to \mathbb{C}$. Figure 2 shows the first excited state ξ_1 for the case $n = 12$.

The Hamilton-Operator gives rise to the unitary *time evolution operator* $U(t) = exp(-iHt)$ and allows the study and musical exploration of the time developments of individual wave functions. A crucial open problem for their interpretation in the music theoretical context of pitch class profiles ($n = 12$) or scale degree profiles ($n = 7$) is the interpretation of the phases. On the one hand, building on [9] it seems plausible to interpret the pitch class profiles as probability density functions of underlying wave functions. On the other hand, this would imply that the established application of the finite Fourier-Transform to pitch class profiles, is not the quantum-theoretical change of perspective from the position to the momentum representation. While these questions need to be addressed in future investigations, it is useful to explore the finite wave functions and their time developments in practical musical experiments.

An auspicious musical application of the time development of finite quantum wave functions is the control of sound sources in a spatialisation scenario. The dimension n of the wave function is the number of sound sources, which are supposed to move in a horizontal plane. As an illustration I will show some experiments with the Max/MSP library *Spat*[1] in combination with *Mathematica*[2]. The time development of a given wave function is encoded in a textfile and using a *Coll*-Object in connection with a metronome at control rate, the Max/MSP-patch interprets the magnitudes and phases at every time stamp in terms of distances and azimuths of the individual sound sources. The Spatialisateur calculates the resulting outputs for the available arrangement of a circle of loudspeakers. In conjunction with our poster presentation we will demonstrate this scenario through the usage of a binaural synthesis instead of the multichannel version. A five-dimensional application is part of a musical piece (of the first author) entitled *The Backside of the Stroboscope* which is dedicated to Jack Douthett.

References

1. Amiot, E.: Music Through Fourier Space. Discrete Fourier Transform in Music Theory, Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45581-5
2. Blutner, R., beim Graben, P.: Gauge models of musical forces. J. Math. Music **15**(1), 17–36 (2021). https://doi.org/10.1080/17459737.2020.1716404
3. Clampitt, D., Noll, T.: Modes, the height-width duality, and Handschin's tone character. Music Theor. Online, 1**7**(1) (2011). http://www.mtosmt.org/issues/mto.11.17.1/mto.11.17.1.clampitt_and_noll.html
4. De La Torre, A.C., Goyeneche, D.: Quantum mechanics in finite-dimensional Hilbert space. Am. J. Phys. **71**(1), 49–54 (2003)

---

[1] https://forum.ircam.fr/projects/detail/spat/.
[2] https://www.wolfram.com/mathematica/.

5. Ebeling, M.: Tonhöhe: Physikalisch - Musikalisch - Psychologisch - Mathematisch. Peter Lang, Frankfurt a.M (1999)
6. Feichtinger, H..G., Hazewinkel, M., Kaiblinger, N., Matusiak, E., Neuhauser, M.: Metaplectic operators on $\mathbb{C}^n$. Q. J. Math. **59**(1), 15–28 (2007). https://doi.org/10.1093/qmath/ham023
7. beim Graben, P.: Musical pitch quantization as an eigenvalue problem. J. Math. Music **14**(3), 329–346 (2020). https://doi.org/10.1080/17459737.2020.1763488
8. Hall, Brian C..: Quantum Theory for Mathematicians. GTM, vol. 267. Springer, New York (2013). https://doi.org/10.1007/978-1-4614-7116-5
9. Beim Graben, P., Blutner, R.: Quantum approaches to music cognition. Quantum Theory Mathematicians **91**, 38–50 (2019)
10. Vourdas, A.: Quantum systems with finite Hilbert space. Rep. Prog. Phys. **67**(3), 267–320 (2004). https://doi.org/10.1088/0034-4885/67/3/r03
11. Vourdas, A., Banderier, C.: Symplectic transformations and quantum tomography in finite quantum systems. J. Phys. Math. Theor. **43**(4), 042001 (2010). https://doi.org/10.1088/1751-8113/43/4/042001
12. Yust, Jason: Applications of DFT to the theory of twentieth-century harmony. In: Collins, Tom, Meredith, David, Volk, Anja (eds.) MCM 2015. LNCS (LNAI), vol. 9110, pp. 207–218. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20603-5_22
13. Yust, J.: Schubert's harmonic language and Fourier phase space. J. Music Theory **59**(1), 121–181 (2015). https://doi.org/10.1215/00222909-2863409
14. Yust, J.: Harmonic qualities in Debussy's Les sons et les parfums tournent dans l'air du soir. J. Math. Music **11**(2–3), 155–173 (2017). https://doi.org/10.1080/17459737.2018.1450457
15. Yust, Jason: Probing questions about keys: tonal distributions through the DFT. In: Agustín-Aquino, Octavio A.., Lluis-Puebla, Emilio, Montiel, Mariana (eds.) MCM 2017. LNCS (LNAI), vol. 10527, pp. 167–179. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-71827-9_13
16. Yust, J.: Geometric Generalizations of the Tonnetz and Their Relation to Fourier Phases Spaces, Chapter 13, pp. 253–277. World Scientific (2018). https://doi.org/10.1142/9789813235311_0013. https://www.worldscientific.com

The Mystery of Anatol Vieru's Periodic Sequences Unveiled

Luisa Fiorot[ID], Alberto Tonolo[ID], and Riccardo Gilblas[(✉)][ID]

Dipartimento di Matematica "Tullio Levi-Civita", Università degli Studi di Padova, Padua, Italy
`riccardo.gilblas@math.unipd.it`

Abstract. In [10], Anatol Vieru proposed a compositional technique based on an algorithmic manipulation of periodic sequences in $\mathbb{Z}_{12}$. This technique was translated in mathematical terms in ([3,4,8]). Two mathematical problems arose starting from the so called Vieru's sequence V: period of primitives and proliferation of values. In this paper we announce, providing only the sketch of the proofs, the solution of these questions in a purely algebraic way.

Introduction

In the *Book of Modes* [10], the romanian composer Anatol Vieru collects periodic sequences by iteratively applying a finite sum operator starting from the constant sequence (6) on $\mathbb{Z}_{12}$, corresponding to the triton interval. Then he decodes from each sequence a musical aspect, giving rise to a composition: *Zone d'Oubli*.

This was the starting point for a prolific math-music research area ([1–4,8]). Vieru highlighted two remarkable phenomena about the particular sequence (so called Vieru's sequence):

$$V = (2, 1, 2, 4, 8, 1, 8, 4) \in \mathbb{Z}_{12}$$

originated from the initial sequence $(2, 1)$ corresponding to Messiaen's second mode of limited transpositions. Vieru repeatedly applied to V the operator Σ_8 (see Eq. 1). He noticed that in the obtained sequences the period tends to increase and it is always a power of 2. Moreover the values 4 and 8 tend to proliferate among the coefficients of the sequences (recovering in some cases more than 99% of the coefficients [4]). In [4] the authors faced the problem using the Fitting Lemma and explicit computations, providing the main reference for this work, but leaving open the two problems. In a recent article submitted to a mathematical journal, we completely solved these questions. The main new idea consists in linking periodic sequences to binomial coefficients, which have been studied using Kummer's Theorem [7] and the generalisation of Lucas' Theorem ([5,6]). In this paper, we announce these results providing only a sketch of the proofs.

1   Anatol Vieru's Periodic Sequence: A New Formalization

Let us recall some definitions.

A sequence $f \in \mathbb{Z}_m^{\mathbb{N}} := (\mathbb{Z}/m\mathbb{Z})^{\mathbb{N}}$ is called *periodic* if there exists $j \geq 1$ such that $\theta^j(f) = f$ where θ is the shift operator defined by

$$\theta(f)(n) := f(n+1) \quad \forall n \in \mathbb{N}.$$

The minimal $j \geq 1$ satisfying this condition is called the *period* of f and it is denoted it by $\tau(f)$ (we use the notation $(a_0, \ldots, a_{n-1})$ for a sequence of period n). The set $P_m := \bigcup_{j \geq 1} \ker(\theta^j - \mathrm{id})$ of all periodic sequences over $\mathbb{Z}_m$ is a $\mathbb{Z}_m$-module with point-wise sum and multiplication.

Let us consider on P_m the operators $\Delta := \theta - \mathrm{id}$ (discrete derivation) and Σ_c for $c \in \mathbb{Z}_m$ (discrete integration) defined as

$$\Sigma_c f(n) := \begin{cases} c & \text{if } n = 0 \\ f(n-1) + \Sigma_c f(n-1) & \text{if } n > 0. \end{cases} \tag{1}$$

We will write Σ instead of Σ_0 to keep the notation clean. We denote by (c) the constant sequence (i.e., periodic sequence of period 1) having all entries equal to $c \in \mathbb{Z}_m$. Hence $\Sigma_c f = \Sigma f + (c)$ and $\Delta(\Sigma_c f) = f$ for every $f \in P_m$ and $c \in \mathbb{Z}_m$. Observe that in particular $\Sigma_c(f_1 + f_2) = \Sigma f_1 + \Sigma_c f_2$ for any $f_1, f_2 \in P_m$ and so $\Sigma_c^s(f_1 + f_2) = \Sigma^s f_1 + \Sigma_c^s f_2$ for any $s \geq 1$.

Given the constant sequence $f = (c)$ with $c \neq 0$, we have: $\Sigma f(0) = 0, \Sigma f(1) = c$ and iterating $\Sigma f(n) = nc$ so $\Sigma f = (0, c, 2c, \ldots, (m-1)c)$ has period m. More generally one has:

Lemma 1. *If (c) is a constant sequence in P_m, then $\Sigma^s(c)(n) \equiv_m c\binom{n}{s}$.*

The period never decreases when applying Σ. Indeed the following holds:

Lemma 2. *Given $f \in P_m$ of period τ, let us denote by $\mathrm{tr}(f) := \sum_{i=0}^{\tau-1} f(i)$. For each $c \in \mathbb{Z}_m$ the period of $\Sigma_c f$ is $h\tau$ where h is the minimum positive integer such that $h \cdot \mathrm{tr}(f) \equiv 0 \mod m$.*

We say that a periodic sequence $f \in P_m$ is *nilpotent* (resp. *idempotent*) if there exists $n \geq 1$ such that $\Delta^n f = 0$ (resp. $\Delta^n f = f$). These two kinds of sequences are called resp. *reducible* and *reproducible* sequences in [2]. The *nilpotency* (resp. *idempotency*) *index* of f is the minimal n satisfying the previous condition. We denote by N_m^Δ and I_m^Δ the $\mathbb{Z}_m$-submodules of nilpotent resp. idempotent sequences.

Example 1. 1. Consider the sequence $f = (0, 1, 2, 3) \in P_4$. We have:

$$\begin{aligned} \Delta f = {} & \theta f - f = (1, 2, 3, 0) - (0, 1, 2, 3) = (1) \\ \Delta(1) = {} & \theta(1) - (1) = (1) - (1) = (0). \end{aligned}$$

Hence f is nilpotent with nilpotency index 2.
2. The sequence $g = (2, 1) \in P_3$ is idempotent of idempotency index 1, since:

$$\Delta g = \theta g - g = (1, 2) - (2, 1) = (2, 1) = g.$$

2   Decomposing P_m

We use three decompositions: the decomposition into primes, the decomposition in nilpotent and idempotent parts and lastly the decomposition of nilpotent sequences using constants.

2.1   Decomposition with Primes

Given $m \in \mathbb{N}$, $m \geq 2$ with prime factorization $m = \prod_{i=1}^{t} p_i^{\ell_i}$, the group isomorphism $\mathbb{Z}/m\mathbb{Z} \to \bigoplus_{i=1}^{t} \mathbb{Z}/p_i^{\ell_i}\mathbb{Z}$ gives rise to an isomorphism of abelian groups (see [2, Th. 5])

$$P_m \longrightarrow \bigoplus_{i=1}^{t} P_{p_i^{\ell_i}}$$
$$f \longmapsto (f_{p_i})_{1 \leq i \leq t}$$

where f_{p_i} is the projection of f in $P_{p_i^{\ell_i}}$ and we will call it the p_i-part of f. Its inverse is given by the Chinese remainder theorem.

Lemma 3. ([2, Prop. 6, Prop. 13]) *Following the previous notation, f is nilpotent (resp. idempotent) if and only if its p_i-part f_{p_i} is nilpotent (resp. idempotent) for any $1 \leq i \leq t$. The nilpotency (resp. idempotency) index coincides with the maximum of the nilpotency (resp. idempotency) indices of f_{p_i} and the period $\tau(f) = \mathrm{lcm}\{\tau(f_{p_i})\}_{1 \leq i \leq t}$.*

Example 2. Since $\mathbb{Z}/12\mathbb{Z} \simeq \mathbb{Z}/3\mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z}$, we obtain $P_{12} \simeq P_3 \oplus P_4$ and Vieru's sequence $V = (2, 1, 2, 4, 8, 1, 8, 4)$ decomposes as:

$$V_3 = (2, 1) \in P_3, \qquad V_2 = (2, 1, 2, 0, 0, 1, 0, 0) \in P_4.$$

2.2   Decomposition in Nilpotent and Idempotent Part

Let us recall [4, Prop. 1] that, by the Fitting Lemma, $P_m = I_m^{\Delta} \oplus N_m^{\Delta}$. The primes decomposition and Lemma 3 imply the following isomorphisms:

$$P_m = \bigoplus_{i=1}^{t} I_{p_i^{\ell_i}}^{\Delta} \oplus N_{p_i^{\ell_i}}^{\Delta} \qquad I_m^{\Delta} = \bigoplus_{i=1}^{t} I_{p_i^{\ell_i}}^{\Delta} \qquad N_m^{\Delta} = \bigoplus_{i=1}^{t} N_{p_i^{\ell_i}}^{\Delta}.$$

Thus we can always reduce to study sequences on $\mathbb{Z}_{p^{\ell}}$.

Lemma 4. *If $f \in P_{p^{\ell}}$, then:*

1. *[4, Th. 3] $f \in N_{p^{\ell}}^{\Delta}$ if and only if $\tau(f) = p^t$ for $t \in \mathbb{N}$;*
2. *if $f \in N_{p^{\ell}}^{\Delta}$ with period p^t and nilpotency index η, then $\eta \leq \ell p^t$.*

2.3   Decomposition of Nilpotent Sequences Using Constants

The nilpotent sequences decompose in sums of primitives of constant sequences:

Lemma 5. *A nilpotent sequence $f \in N_m^\Delta$ of nilpotency index η can be written in a unique way as*

$$f = c_0 + \Sigma^1 c_1 + \cdots + \Sigma^{\eta-1} c_{\eta-1}$$

for suitable constants $c_0, \ldots, c_{\eta-1} \in \mathbb{Z}_m$.

Applying the previous decompositions to Vieru's sequence

$$V = (2, 1, 2, 4, 8, 1, 8, 4) \in P_{12}$$

we find the 2-part $V_2 = (2, 1, 2, 0, 0, 1, 0, 0) \in P_4$ and the 3-part $V_3 = (2, 1) \in P_3$. In this case V_2 is nilpotent of index 5 while V_3 is idempotent with index 1.

Therefore the nilpotent and idempotent components are:

$$\widetilde{V_2} = (6, 9, 6, 0, 0, 9, 0, 0) \equiv (-3) \cdot V_2 \mod 12, \qquad \widetilde{V_3} = (8, 4) \equiv 4 \cdot V_3 \mod 12.$$

Vieru repeatedly used the operator Σ_8 applied to $V = \widetilde{V_2} + \widetilde{V_3}$ in order to generate new periodic sequences. Since $\Sigma_8 \widetilde{V_3} = \widetilde{V_3}$, we get

$$\Sigma_8^s V = \Sigma^s \widetilde{V_2} + \Sigma_8^s \widetilde{V_3} = \Sigma^s \widetilde{V_2} + \widetilde{V_3}. \tag{2}$$

Thus we are reduced to study the operator Σ applied to the nilpotent sequence V_2 in P_4. In particular the period of $\Sigma_8^s V$ in P_{12} coincides with the period of $\Sigma^s V_2$ in P_4. Since $\Sigma_8 \widetilde{V_3} = \widetilde{V_3} = (8, 4)$, the proliferation of the values $8, 4$ in $\Sigma_8^s V$ in P_{12} is equivalent to the proliferation of the value 0 in $\Sigma^s V_2$ in P_4.

Finally, the last decomposition provided by Lemma 5 gives $V_2 = (2) + \Sigma(3) + \Sigma^2(2) + \Sigma^3(3) + \Sigma^4(2)$.

3   Unveiling the Period and the Proliferation of Values

3.1   Period of the Primitives of Vieru's Sequence

The study of the period is based on the following lemma:

Lemma 6. *For every $s \in \mathbb{N}$, the sequence $\Sigma^s(2) \in P_4$ has period 2^{k_s} while $\Sigma^s(3) \in P_4$ has period 2^{k_s+1} where $k_s := \lfloor \log_2(s) \rfloor + 1$ is the number of figures in the representation of s in base 2.*

Proof. We prove the statement for the primitives of (3) proceeding by induction on the primitive index s. As observed before Lemma 1, the period of $\Sigma(3)$ is $4 = 2^{k_1+1}$. Suppose the statement true for s, let us prove it for $s + 1$.

- If $k_{s+1} = k_s$, it is possible to show that $\mathrm{tr}(\Sigma^s(3)) = 0$ using Kummer's Theorem [7]. By Lemma 2 one obtains

$$\tau(\Sigma^{s+1}(3)) = \tau(\Sigma^s(3)) = 2^{k_s} = 2^{k_{s+1}}.$$

– If $k_{s+1} = k_s + 1$, again by Kummer's Theorem one gets $\text{tr}(\Sigma^s(3)) = 2$. By Lemma 2 one obtains

$$\tau(\Sigma^{s+1}(3)) = 2 \cdot \tau(\Sigma^s(3)) = 2 \cdot 2^{k_s} = 2^{k_s+1} = 2^{k_{s+1}}.$$

Reducing (2) $\in P_4$ to (1) $\in P_2$ via the isomorphism $2\mathbb{Z}/4\mathbb{Z} \xrightarrow{\sim} \mathbb{Z}/2\mathbb{Z}$ the previous argument proves the statement.

The following result solves completely the problem of the period of Vieru's sequence $V = (2, 1, 2, 4, 8, 1, 8, 4)$.

Theorem 1. *The period of $\Sigma^s V$ is $2^{k_{s+3}+1}$ (with k_{s+3} is the number of figures of $s+3$ in base 2). Notice that the period changes whenever $s = 2^r - 3$ for $r \geq 2$.*

Proof. As previously observed, the period of $\Sigma^s V$ coincides with the period of $\Sigma^s V_2$. The period of the sequences

$$\Sigma^s V_2 = \Sigma^s(2) + \Sigma^{s+1}(3) + \Sigma^{s+2}(2) + \Sigma^{s+3}(3) + \Sigma^{s+4}(2)$$

clearly divides the least common multiple of the periods of its summands, which coincides with the period of $\Sigma^{s+3}(3)$. Using the previous theorem and Lemma 2, an accurate analysis permits to prove that they are in fact equal.

3.2   Proliferation of Values

Let us study the proliferation of 0 in $\Sigma^s V_2$ with $s \geq 1$. This will allow us to evaluate the number of $4, 8$ in the primitives of Vieru's sequence V. Rather than the absolute value of occurrences inside the period, it is much more interesting to study the ratio with respect to the period. In [4], the authors explicitly computed the first 61 primitives of V and they remarked that *"at level 61 of period 128, more than 90% of the elements belong to the set $\{4, 8\}$. This percentage dramatically decreases in the following level which is the last one having period equal to 128"*. The level 61 above corresponds to $s = 59$ and it has period 128 by Theorem 1, as confirmed by the computation in [4].

Lemma 7. *Let us denote by $z(s)$ the number of zeros in the sequence $\Sigma^{s-3} V_2$ inside its period. Then for every $r \geq 3$ the following inequalities hold:*

$$z(2^r - 1) < z(2^r - 2) = 2^{r+1} - 8.$$

More precisely, one has:

$$\Sigma^{2^r-5} V_2 = (\underbrace{0, \ldots, 0}_{2^r-5}, 2, 3, 1, 0, 0, \underbrace{0, \ldots, 0}_{2^{r-1}-4}, 2, 2, 0, 0, \underbrace{0, \ldots, 0}_{2^{r-1}-5}, 2, 1, 3, 0, 0).$$

Proof. The proof is based on Kummer's Theorem [7] and on the generalization of Lucas' Theorem proved by Davis and Webb in [5, Th. 1]. The second statement is proved by a component-wise analysis. In the under-braced positions all summands of V_2 are equal to zero by Kummer's Theorem. The remaining 14 coefficients can be explicitly computed using Davis and Webb result.

The following result solves completely the problem of the proliferation of $\{4, 8\}$ in Vieru's sequence $V = (2, 1, 2, 4, 8, 1, 8, 4)$.

Theorem 2. *The ratio between the number of $\{4, 8\}$ in the primitives of $\Sigma_8^{2^r-5}V$ (with $r \geq 3$) and their period is $\frac{2^{r+1}-8}{2^{r+1}}$, which tends to 1 for $r \to \infty$.*
More precisely, $\Sigma_8^{2^r-5}V$ is equal to:

$$(\underbrace{8, 4, 8, 4, \ldots, 8}_{2^r-5}, 10, 11, 1, 8, 4, \underbrace{8, 4, 8, 4, \ldots, 4}_{2^{r-1}-4}, 2, 10, 8, 4, \underbrace{8, 4, 8, 4, \ldots, 8}_{2^{r-1}-5}, 10, 5, 7, 8, 4).$$

Proof. Since the number of $\{4, 8\}$ in $\Sigma_8^{2^r-5}V$ coincides with the number of zeros in $\Sigma^{2^r-5}V_2$, the first part of the statement follows from the previous lemma and Theorem 1. The explicit form for $\Sigma_8^{2^r-5}V$ follows from the previous lemma and Eq. (2), where

$$\Sigma^{2^r-5}\widetilde{V_2} \equiv (-3) \cdot \Sigma^{2^r-5}V_2 \mod 12.$$

Remark 1. It is nice to compare the formula of Theorem 2 for $r = 3, 4, 6$ with the explicit computation of the corresponding levels $5, 13, 61$ provided in [4, App. A].

4   Recursive Formulas for the Number of $\{4, 8\}$ in $\Sigma_8^{s-3}V$

In the last days we proved a recursive formula for the number $z(s)$ of zeroes in the $(s-3)$-primitive of V_2 in P_4. The number $z(s)$ coincides also with the number of $\{4, 8\}$ in $\Sigma_8^{s-3}V$ in P_{12}. We chose this shift by -3 in order to have all sequences of the same period 2^{r+2} when $2^r \leq s < 2^{r+1}$. In this interval of primitives, one can compute the percentage of $\{4, 8\}$ as $z(s)/2^{r+2}$.

We need first to introduce a tuple $\mathfrak{d}_r$ of integers. Denote by $\mathrm{wt}(\ell)$ the Hamming weight of ℓ, i.e. the number of 1's in the binary expansion of ℓ. Then we set

$$\mathfrak{d}_r(m) = 2^{\mathrm{wt}(2^r+2^{r-1}-1-m)+1}.$$

We will be mainly interested in the values of $\mathfrak{d}_r(m)$ when $2^r + 2^{r-2} + 3 \leq m < 2^r + 2^{r-1} - 1$. For brevity we write

$$\mathfrak{d}_r := (\mathfrak{d}_r(m))_{2^r+2^{r-2}+3 \leq m < 2^r+2^{r-1}-1}.$$

For $\mathfrak{d}_r$ the following equalities hold:

$$\mathfrak{d}_5 = (4, 8, 4, 4) \quad \text{and} \quad \mathfrak{d}_{r+1} = (2 \times \mathfrak{d}_r, 4, 2^{r-1}, 2^{r-2}, 2^{r-2}, \mathfrak{d}_r) \ \forall r \geq 5.$$

We can now enunciate the recursive formula. Define:

$$u_i := 2^{r-i} \qquad i = 1, 2, 3$$
$$t := s - u_1 \qquad 2^r \leq s < 2^{r+1}$$
$$(c_1, c_2, c_3, c_4) := 2^{r-3}(12, 8, 10, 11)$$
$$(c'_1, c'_2, c'_3, c'_4) := 2^{r-3}(12, 10, 11, 12).$$

The initial condition for the recursive formula is the 2^5-tuple $(z(s))_s$ for $2^5 \leq s < 2^6$:

$$(32, 48, 64, 88, 64, 80, 88, 92, 64, 80, 88, 104, 92, 104, 108, 94,$$
$$78, 88, 96, 108, 96, 104, 108, 110, 102, 108, 112, 118, 114, 118, 120, 64).$$

In this interval, the period of the sequences $\Sigma_8^{s-3} V$ is constantly equal to 128.

For $2^r \leq s < 2^{r+1}$ with $r \geq 6$, the 2^r-tuple $(z(s))_s$ concides with:

$$(\underbrace{2z(t), \ldots, 2z(t)}_{2^{r-2}-1}, \underbrace{z(t-u_3)+c_1, \ldots, z(t-u_3)+c_4}_{4}, \underbrace{2z(t)-\mathfrak{d}_r(s), \ldots, 2z(t)-\mathfrak{d}_r(s)}_{2^{r-2}-4},$$

$$\underbrace{z(t-u_2)+c_1', \ldots, z(t-u_2)+c_4'}_{4}, \underbrace{z(t-u_1)+2^{r+1}, \ldots, z(t-u_1)+2^{r+1}}_{2^{r-1}-4}, \underbrace{2z(t-u_1)}_{1}).$$

Let us recall that $t = s - 2^{r-1}$ and so in the tuple above the first coefficient is computed using $s = 2^r$, the second one using $s = 2^r + 1$, the last one using $s = 2^{r+1} - 1$.

References

1. Ancelotti, N.: On some algebraic aspects of Anatol Vieru Periodic Sequences. Tesi di Laurea Triennale in Matematica, Università degli Studi di Padova, Relatore L. Fiorot
2. Andreatta, M., Vuza, D.T.: On some properties of periodic sequences in Anatol Vieru's modal theory. Tatra Mt. Math. Publ. **23**, 1–15 (2001)
3. Andreatta, M., Agon, C., Vuza, D.T.: Analyse et implémentation de certaines techniques compositionnelles chez Anatol Vieru, pp. 167–176. Marseille, Actes des Journées d'Informatique Musicale (2002)
4. Andreatta, M., Vuza, D.T., Agon, C.: On some theoretical and computational aspects of Anatol Vieru's periodic sequences. Soft Comput. **8**(9), 588–596 (2004). https://doi.org/10.1007/s00500-004-0382-7
5. Davis, K.S., Webb, W.A.: Lucas' theorem for prime powers. European J. Combin. **11**(3), 229–233 (1990)
6. Fine, N.J.: Binomial coefficients modulo a prime. Amer. Math. Monthly **54**, 589–592 (1947)
7. Kummer, E.: Über die Ergänzungssätze zu den allgemeinen Reciprocitätsgesetzen. J. Reine Angew. Math. **44**, 93–146 (1852)
8. Lanthier, P., Guichaoua, C., Andreatta, M.: Reinterpreting and extending Anatol Vieru's periodic sequences through the cellular automata formalisms. In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O.A. (eds.) MCM 2019. LNCS (LNAI), vol. 11502, pp. 261–272. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_21
9. Mariconda, C., Tonolo, A.: Discrete Calculus. U, vol. 103. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-03038-8
10. Vieru, A.: The Book of Modes. Editura Muzicala, Bucharest (1993)

Benford's Law and Music Note Frequencies

Sybil Prince Nelson(✉) , Brian Wickman, Jack Null, and Eric Gazin

Washington and Lee University, Lexington, VA 24450, USA
`sprincenelson@wlu.edu`

Abstract. We considered musical note frequencies of the 88 keys of the piano and found that they are Benford distributed. We extended our focus beyond the 88 keys and found the connection to Benford holds for the lowest note measuring at 16.35 Hertz (Hz) to the highest note of 7902.13 Hz. We next investigated whether the distribution holds within specific types of music. We found that classical music such as a random sample of songs from the Romantic period adhere to the Benford distribution while modern music such as a random sampling of songs from the 2000s do not. We also coined a term called "Naturalness" to assess how well a song adheres to the Benford distribution.

Keywords: Benford distribution · Logarithmic distribution of first digits · Classical music

1 Introduction

Greek mathematician and scientist Pythagoras said: "There is geometry in the humming of strings, there is music in the spacing of the strings". Ever since Pythagoras demonstrated that there were simple numerical ratios that produced all the intervals necessary to create a musical scale, there have been countless connections made between mathematics and music [1]. Counting, rhythm, keys, intervals, patterns, symbols, harmonies, time signatures, overtones, tone, pitch and even the notations of composers are all connected to mathematics [2]. This paper introduces yet another connection: musical note frequencies are Benford distributed.

For this paper, we first considered the formula for frequency of music notes and showed that it is a sequence whose limit is Benford. Next, we considered the 88 keys of the piano and their note frequencies in Hertz. Figure 5 provides the 88 frequencies. We also compared a collection of classical songs from the Romantic period of music and compared them to a random selection of songs from the 2000s. We downloaded the classical music MIDI files from http://www.kunstderfuge.com and the modern songs from bitmidi.com and freemidi.com. Each of these websites provide a resource for music files. We used R to extract the notes used in each song. We found that the 88 keys of the piano are Benford

distributed. When we extend our focus and look at notes beyond those found on a piano, we find that those frequencies are even closer to the Benford distribution. When comparing the works from the Romantic period to the songs of the 2000s, we find that Romantic period works are much more Benford. This paper will proceed as follows: first, we will provide a summary of Benford's Law. Then we will present our analysis and results of the distribution of note frequencies. Next, we will discuss what conclusion can be made based on our analysis as well as possible future steps.

1.1   What is Benford's Law

Benford's Law began as the empirical observation from Newcomb in 1881 that distribution for first digits in any set of numbers is not uniform or symmetric [3]. Specifically, the frequency of 1s is higher than the frequency of 2s and so on. This observation was later supported by Benford in 1938 with a data set of over 20,000 numbers from naturally occurring data sets such as the atomic weights of elements [4]. He was able to show that the probability of a digit being $d(d = 1, 2, \cdots, 9)$ was equal to

$$Prob(d) = \log_{10}(1 + \frac{1}{d})$$

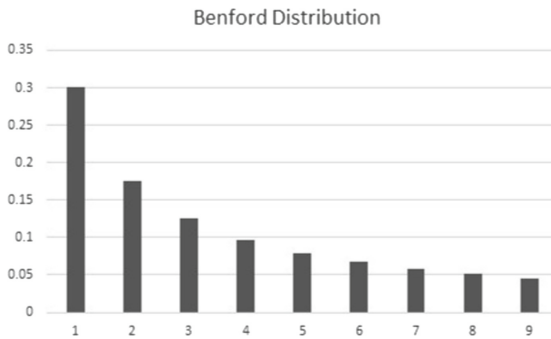This was then named Benford's Law. The first digit probabilities follow the specific distribution shown in Fig. 1.

Fig. 1. Distribution of Benford's law

The Benford distribution has been found in many collections of numerical data across various fields of natural and social sciences [5] and even in music [6]. This paper further explores its connection to music. To measure how far a song deviates from Benford, we use a measure for Delta found in Berger and Hill's *An Introduction to Benford's Law.* [7]

$$\Delta = 100 \cdot (max)_{d=1}^{9}|Prob(D_1 = d) - \log_{10}(1 + \frac{1}{d})|$$

where D_1 denotes the first digit of a number, D_2 denotes the second and so on. For example, $D_1(2568) = 2$ while $D_2(2568) = 5$

The Δ values simply measure how well a data set adheres to the Benford distribution in that if $\Delta = 0$, it is perfectly Benford. Essentially, it compares the frequency of each digit with that of the corresponding digit in the Benford distribution. The largest deviation becomes the Δ value for the set. One limitation of this method is that the digit $d = 1, ..., 9$ that determines the Δ significantly affects whether a χ^2 test of Goodness of Fit is rejected. For example, a Δ of 4.72 creates a p-value of 0.69 if the Δ results from the first digit. A Δ of 4.72 creates a p-value of 0.007 if it results from the seventh digit. This means that we would fail to reject that our observed distribution is Benford if the Δ is from the first digit while we would reject if the same Δ was found in the seventh digit. A future paper address this issue creating a classification system for determining the "Benfordness" of a song.

Not all data sets follow a Benford distribution. One of the criteria for those that do is that it is a naturally occurring data set and one that is not subjective or artificially derived. Because of this, we have determined that the closer a dataset or song adheres to Benford the more "naturalness" it has. We do not quantify this quality as good or bad or leading to a more popular or less popular song. It is merely a measure to assess the music.

2   Numeric Proof

We used the formula $f_n = 440 \cdot (2)^{\frac{n-69}{12}}$ to find the frequency in Hertz of piano keys. This formula corresponds to the note values assigned by MIDI. Note 69 is the A above middle C so it has a frequency 440 Hz, while middle C is assigned note 60 and has a frequency of 261.63. The sequence of frequencies is Benford. Its deviation (Δ) from Benford goes to 0 as N$\to \infty$ Figs. 2, 3, 4 show that as N$\to \infty$ or the number of notes increases the distribution of the first digit gets more and more Benford. The deviation from Benford measured with Δ gets closer to 0 (Figs. 6 and 7).

Table 1. Individual probabilities for each digit and Δ for each value of N

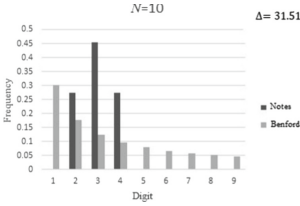| N	P(d = 1)	P(d = 2)	P(d = 3)	P(d = 4)	P(d = 5)	P(d = 6)	P(d = 7)	P(d = 8)	P(d = 9)	Δ
10	0	0.275	0.45	0.275	0	0	0	0	0	31.51
100	0.24	0.16	0.15	0.12	0.09	0.08	0.07	0.05	0.04	6.11
1000	0.3	0.178	0.125	0.097	0.079	0.066	0.058	0.051	0.046	0.19
10000	0.301	0.1756	0.1251	0.097	0.0791	0.0671	0.058	0.0511	0.0459	0.05

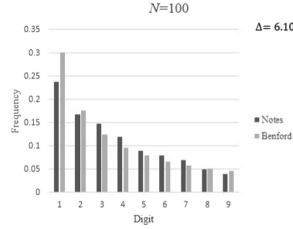**Fig. 2.** The first 10 values of the note frequency formula

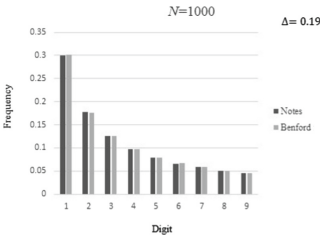**Fig. 3.** The first 100 values of the note frequency formula

Fig. 4. The first 1000 values of the note frequency formula
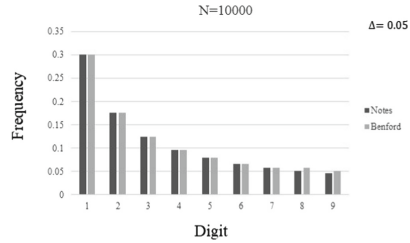

Fig. 5. The first 10000 values of the note frequency formula

As seen in Table 1, for every natural number m, and $d_1 \in 1, 2, ..., 9$ and all $d_j \in 1, 2, ..., 9, j \geq 2$

$$\lim_{N \to \infty} \frac{\text{total}\# 1 \leq n \leq N : D_j(f_n) = d_j \text{for} j = 1, 2, ..., 9}{N}$$

$$= \log(1 + (\sum_{j=1}^{m} 10^{m-j} d_j)^{-1})$$

where $f_n = 440 \cdot (2)^{\frac{n-69}{12}}$ and D_j is the first digit of a number corresponding to $j = 1, 2, ..., 9$. Thus, by definition f_n or the sequence of note frequencies is Benford.

3 Piano Note Frequencies

Next, we focused on just the 88 keys of a piano. For piano note frequencies, we also performed a Chi-squared test, Pearson correlation and looked at Euclidean distance. All of these measures provide even more justification for piano note frequencies being Benford distributed. The Chi-Squared p-value for null hypothesis of no difference between the distributions is 0.9755. Thus, we fail to reject the null and conclude that the distribution of the first digit of piano note frequencies
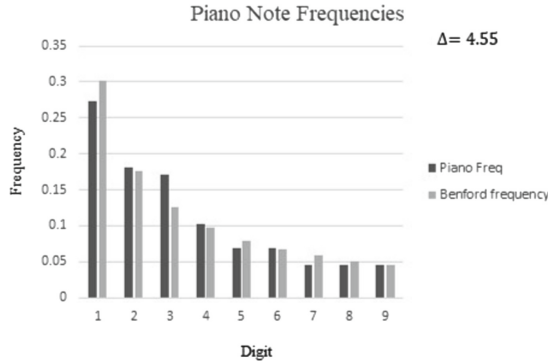
Fig. 6. Piano note frequency distribution

is no different from the Benford distribution. Pearson correlation test indicates that there is a correlation between the distribution of the first digit of piano note frequencies and the Benford distribution. The correlation is 0.97 with a p-value < 0.001. The normalized Euclidean distance is measured as follows:

$$d^* = \frac{\sqrt{\sum_{i=1}^{9}(b_i - e_i)}}{\sqrt{\sum_{i=1}^{8} b_i^2 + (1 - e_9)}}$$

The closer the statistic is to 0 the more similar the distributions are. The Euclidean distance in this instance is 0.055 indicating there is no difference between the distribution of the first digit of piano note frequencies and the Benford distribution.

4 Application

In an attempt to apply what we have noticed about the distribution of piano note frequencies, we compared a sample of music from two different time periods in music history.

Wilcoxon rank sum test shows that there is a significant difference between the median delta value of the Romantic period and that of the 2000s with a p-value of 0.04. The paper Benford's Law in Music History showed that Western music became more Benford as it moved through history from the Medieval period to the Romantic [8]. One theory as to why this is the case is that Medieval music is known for mostly monophonic chants used in sacred worship [1]. Over time, music became more complex down through the Romantic period which is known for the imagination and virtuosity of its composers. The modern music time period did not maintain this trend as shown in the figure. Our sample of modern songs was less Benford than the songs from the Romantic period. This might be due to the often repetitive nature of modern music, but more research is needed.
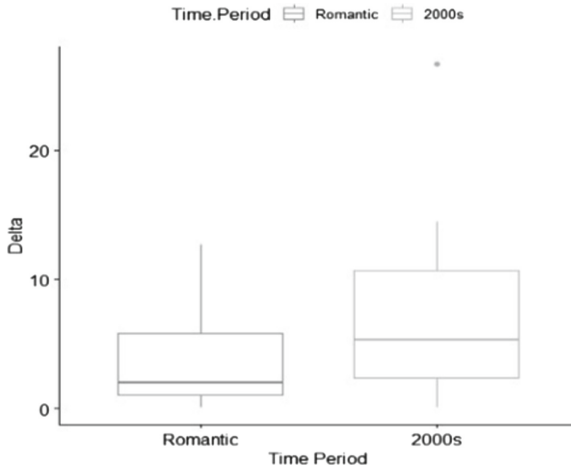
Fig. 7. Comparison of the romantic period to modern music

5 Conclusion

In conclusion, this paper shows that the frequency of musical notes is Benford distributed. Further, classical music such as the works of the Romantic period is more Benford distributed than our collection of modern music. Previous research has shown that music became more Benford as it progressed through history. [8] This trend did not continue through to modern music or music created after World War I. This may be because of the repetition in many modern songs or because of the vast differences found in types of modern music. It may be possible to find trends within specific genres such as Rock or Jazz.

Further research is needed to determine whether how "Benford" a song is will reflect in its popularity. It may even be possible to find some sort of correlation between a song that is Benford distributed and how well or poorly it performs on the charts or in awards programs.

These results cannot be used to quantify what makes a song more popular or more enjoyable. It merely speaks to how well it adheres to the Benford distribution. Since piano note frequencies have a deviation from Benford of $\Delta = 4.55$, we may perhaps conclude that songs near the same Delta value are more "natural". It is possible that it reflects a human component to the music. Thus, we are conducting research in comparing AI created music versus music created by composers. If this holds true, using the Benford distribution may be useful as a first step in identifying digitally manipulated audio recordings just as it is used as a first step in identifying tax fraud or assess the quality of data sets [5].

References

1. Rosenstiel. Leoni: Schirmer History of Music. Schirmer Books (1982)

2. American Mathematical Society: Mathematics and Music. www.ams.org/publicoutreach/math-and-music
3. Newcomb, S.: Note on the frequency of use of the different digits in natural numbers. Am. J. Math. **4**(1), 39–40 (1881). https://doi.org/10.2307/2369148
4. Benford, F.: The law of anomalous numbers. Proc. Am. Philos. Soc. **784**, 551–72 (1938). https://www.jstor.org/stable/984802
5. Li, F., et al.: Application of Benford's law in data analysis. J. Phys. Conf. Ser. **1168**, 032133 (2019). https://doi.org/10.1088/1742-6596/1168/3/032133
6. Khosravani, A., et al.: Emergence of Benford's law in music. J. Math. Sci. Adv. Appl. **54**(1), 11–24 (2018)
7. Berger, A., Hill, T.: An Introduction to Benford's Law. Princeton University Press, New Jersey (2015)
8. Nelson, S.P., et al.: An evaluation of music's adherence to benford's law throughout history. J. Math. Sci. Adv. Appl. **67**(1), 73–84 (2021)
9. Crocetti, E., Randi, G.: "Using Benford's Law as a First Step to Assess the Quality of the Cancer Registry Data" Frontiers in Public Health (2016). https://doi.org/10.3389/fpubh.2016.00225

Altered Chord Alternatives

Lauren C. Ruth[(✉)]

Mercy College, 555 Broadway, Dobbs Ferry, NY 10522, USA
LRuth@mercy.edu
https://www.mercy.edu/directory/lauren-ruth

Abstract. Motivated by the empirically pleasing sound of $E^\flat 7\text{sus}4$ as a substitution for G7 in the II–V–I progression Dm7–G7–CM7, we advocate for the use of this chord and four other uncommon four-note chord substitutions in jazz, taking Schönberg's *schwebend* as justification and explaining the details of our MATLAB code for analyzing a four-note chord's "diatonic citizenship."

**Keywords:** Altered scale · Schönberg · *Schwebend* · MATLAB · Jazz

1 Introduction

Jazz musicians often speak of chord extensions and substitutions as lending different "colors" to their playing. This paper takes a small step towards articulating these "colors" in a mathematically precise way. We hypothesize that perceived "color" is a function of external diatonic context ("In what scales does this chord live?") and internal intervallic structure ("What intervals does this chord contain?"). Our computer-generated chord diagrams elucidate both of those features. For each chord, a 12-tone ring illustrates its intervals; and underneath the 12-tone ring, we list the scales where the chord "lives." We refer to this collection of scales as the chord's *diatonic citizenship*.

We propose extending the practice of altered chords in jazz to what we call *alternative altered chords*. Section 2 summarizes the use of altered chords in jazz and concludes with our definition of alternative altered chords. Section 3 provides theoretical justification for the use of these alternative altered chords from Schönberg's work [5]. In Sect. 4, we describe our MATLAB code for determining a four-note chord's diatonic citizenship. Section 5 lists the alternative altered chords and their diatonic citizenship, and Sect. 6 outlines future directions.

2 Motivation

Here is a II–V–I progression, ubiquitous in jazz, in C major:
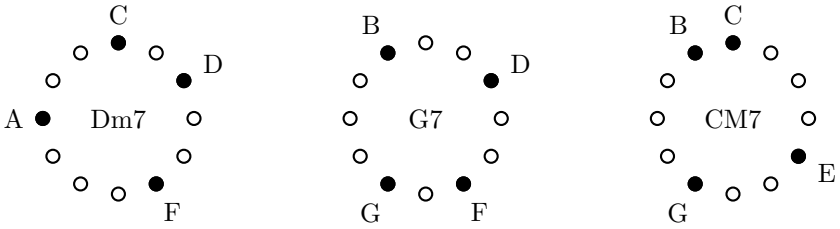
Comping and improvising over a II–V–I progression are among the first skills an aspiring jazz musician learns. After gaining familiarity with this progression, the musician graduates to playing the G altered scale over the G7 chord, rather than playing the G mixolydian scale. G mixolydian is the fifth mode of the C major scale (G, A, B, C, D, E, F), and G altered is the seventh mode of the $A^\flat$ melodic minor scale (G, $A^\flat$, $A^\sharp$, B, $D^\flat$, $D^\sharp$, F). In jazz, the *essential chord tones* of a seventh chord are the 1, 3, and 7. An *altered* G7 chord is created by altering any non-essential chord tone in G7, *i.e.* adding combinations of the $\flat 9$, $\sharp 9$, $\flat 5$, or $\sharp 5$ to combinations of the 1, 3, or 7 of G7. Below are common examples of altered G7 chords.
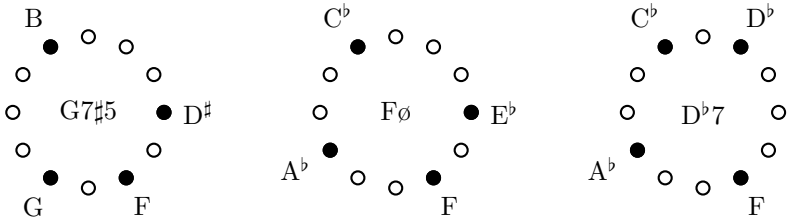
An altered G7 chord represents a compromise between

(i) creating chromatic voice-leading opportunities between Dm7 and CM7, which is accomplished by plugging the semitonal gaps in the C major scale;

(ii) preserving the dominant function of G7 by retaining its essential chord tones G, B, and F; and

(iii) maintaining at least one common tone between each pair of chords (*harmonisches Band*, [5]).

Let us turn to an example lacking (ii) and (iii). Consider the following II–V–I in C major, with $E^\flat 7\text{sus4}$ substituted for G7.
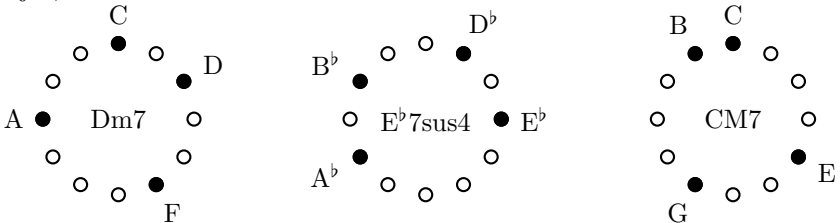
E$^\flat$7sus4 consists of exactly the $\flat$9, $\sharp$9, $\flat$5, and $\sharp$5 of G7, with none of the essential G7 chord tones. Notice how E$^\flat$7sus4 plugs four out of the five semitonal gaps left by the C major scale, which form a G$^\flat$ major pentatonic scale (G$^\flat$, A$^\flat$, B$^\flat$, D$^\flat$, and E$^\flat$). There are five four-note chords available to us from this scale: E$^\flat$7sus4, and four others that lie outside the G altered scale—that is, outside the usual source for G altered chords.

Definition 1. *Let the tonic be C major. Then an **alternative altered chord** is one of the five four-note chords drawn from the notes G$^\flat$, A$^\flat$, B$^\flat$, D$^\flat$, and E$^\flat$.*

(Each alternative altered chord has no intersection with the C major scale. Note that only one of the five chords, E$^\flat$7sus4, could be considered a traditional altered chord, since it lives in the G altered scale.) This definition can be transposed. For example, if the tonic is G major, then to find the source of alternative altered chords to substitute for D7, begin a major pentatonic scale a tritone away from G: D$^\flat$, E$^\flat$, F, A$^\flat$, and B$^\flat$.

3   Theoretical Framework

The empirically pleasing sound of substituting E$^\flat$7sus4 for G7 in the context of a C major II–V–I progression suggests that our arsenal of altered chords could be extended to include those chords which *only* satisfy (i), and neither (ii) nor (iii), from the list of compromises in Sect. 2. As Schönberg writes in [5], pg. 133: "Quite certainly there are harmonic means, which at present have just not been theoretically determined, whose capacity for forming cadences or, far more, for admitting them, is just as great as that of IV, II, V and I."

If we were to use only triads, the V in a II–V–I would be necessary to solidify the key. We don't hear the B of C major without the V triad: II (D-F-A), V (G-B-D), I (C-E-G). But since jazz employs seventh chords, our key is already determined by the notes in the II (D-F-A-C) and the I (C-E-G-B). Thus we can dispense with key determination, connecting the two chords II and I via one chord that lies completely outside the key, using this external chord as "chromatic glue." Here is what we mean by the claim that E$^\flat$7sus4 (E$^\flat$-A$^\flat$-B$^\flat$-D$^\flat$) serves as "chromatic glue" between Dm7 and CM7: E$^\flat$7sus4 links D to C via D$^\flat$, and A to G via A$^\flat$; and as the movements C to B and F to E are already chromatic, the B$^\flat$ and E$^\flat$ serve to add momentary diversions.

On the subject of V substitutions, Schönberg brings up the III and the VII. He writes in [5], pg. 134: "First of all, looking for a substitute for V, we shall consider the suitability of III." Then he offers pros and cons of this substitution: It shares two common tones with I, a drawback; but it contains the leading tone, and it creates a nice root progression. "Yet, it is not commonly used; hence, we shall not use it much either, but shall remember why we do not: chiefly because it is not commonly used. That means, it could be used." He also considers substituting VII for V: "It does indeed determine the key, it does lead to the closing chord; but it, too, is not in common practice today, and so we shall disregard it." While his discussion only concerns diatonic substitutions, we include it as part of the theoretical justifcation for using alternative altered chords, because it

demonstrates Schönberg's acute awareness of the idiom *rules are made to be broken*: There is nothing inherently wrong with using an uncommon substitution. Indeed, the narrative arc of [5] involves establishing rules and discarding them.

As final justification for the alternative altered chords, which will be enumerated in Sect. 5, we quote Schönberg at length: "A piece can also be intelligible to us when the relationship to the fundamental is not treated as basic; it can be intelligible even when the tonality is kept, so to speak, flexible, fluctuating (*schwebend*). Many examples give evidence that nothing is lost from the impression of completeness if the tonality is merely hinted at, yes, even if it is erased. And [...] the analogy with infinity could hardly be made more vivid than through a fluctuating, so to speak, unending harmony, through *a harmony that does not always carry with it a certificate of domicile and passport carefully indicating country of origin and destination*" ([5], pp. 128–129, emphasis added).

So, our five alternative altered chords, the five four-note chords lying entirely outside C major, contribute this *schwebend* quality in a way that standard G7 altered chords (drawn from the G altered scale, containing at least one essential chord tone of G7, hence overlapping with the C major scale) cannot. We propose the V in a II–V–I jazz progression be replaced by a chord that merely serves as "chromatic glue," departing from the tonic entirely, losing any trace of dominant chord function, and maintaining no common tones between chords.

4   Methodology

Restrictions. We impose the following limitations on this investigation.

(A) We assume our II–V–I progression is in the key of C major. So, we are looking for G7 substitutions beyond those residing in the G altered scale (or the $A^\flat$ melodic minor scale). The reader can transpose our C major results to II–V–I progressions in other major keys.
(B) For choice of scale, we restrict our attention to major, melodic minor, harmonic minor, and harmonic major, since these are the scales most commonly associated with the jazz tradition, and since each scale in this four-scale cycle can be changed to the next by altering just one semitone (the third or the sixth note in the scale).
(C) We will be concerned with four-note chords, also known as tetrachords, for the simple reason that these are more easily playable on guitar (the author's instrument) than extensions of higher cardinality.

MATLAB Code Overview. We modifed the MATLAB code [3], which was described briefly in [2] and in greater detail in [4] (in the context of a random walk on a graph that generates modulation excercises for guitar). First, we build a large matrix, `chordsModesKeysScalesRoots`, containing all four-note combinations in 12-TET major, melodic minor, harmonic minor, and harmonic major.

```
1 M=[2 2 1 2 2 2 1];    % major intervals
2 m=[2 1 2 2 2 2 1];    % melodic minor intervals
```

```
3  hm =[2 1 2 2 1 3 1];   % harmonic minor intervals
4  hM =[2 2 1 2 1 3 1];   % harmonic Major intervals
5  scaleMatrix =[M; m; hm; hM];
6  scaleMatrix2 = cat (2 , scaleMatrix , scaleMatrix );
7  n =0;
8  chordsModesKeysScalesRoots =[];
9  for K =0:11    % goes through keys
10   for s =1:4    % goes through scales
11    for i =1:7    % goes through modes
12     for j =( i +1): i +6
13      for k =( j +1): i +6
14       for l =( k +1): i +6
15        n = n +1;
16        int1 = sum ( scaleMatrix2 (s ,i:j -1));   %------------
17        int2 = sum ( scaleMatrix2 (s ,j:k -1));   % four intervals
18        int3 = sum ( scaleMatrix2 (s ,k:l -1));   %
19        int4 = sum ( scaleMatrix2 (s ,l:i +6));   %------------
20        note1 = mod ( K + sum ( scaleMatrix2 (s ,1: i )) - ...
21         scaleMatrix2 (s , i ) ,12);              %------------
22        note2 = mod ( note1 + int1 ,12);          % four notes
23        note3 = mod ( note2 + int2 ,12);          %
24        note4 = mod ( note3 + int3 ,12);          %------------
25        % now write this data into the matrix :
26        chordsModesKeysScalesRoots ( end +1 ,:)=[ int1 ,
27         int2 , int3 , int4 ,  % four intervals
28         i ,     % mode
29         K ,     % key
30         s ,     % scale
31         note1 , note2 , note3 , note4 ]; % four notes
32 end    end    end    end    end    end
```

The matrix chordsModesKeysScalesRoots turns out to have dimensions 6720×11. Each row contains a chord built from the scale interval vectors M, m, hm, and hM. For example, row 222 of chordsModesKeysScalesRoots is

<center>[2 2 3 5 5 0 2 7 9 11 2]</center>

where the numbers [2 2 3 5] describe the intervals in the chord, the numbers [5 0 2] describe a mode (5), key (C), and scale (melodic minor) where this chord can be found, and the numbers [7 9 11 2] describe the notes: G, A, B, D. As another example, row 4437 of chordsModesKeysScalesRoots is

<center>[5 2 2 3 5 7 4 2 7 9 11]</center>

where the numbers [5 2 2 3] describe the intervals in the chord, the numbers [5 7 4] describe a mode (5), key (G), and scale (harmonic major) where this chord can be found, and the numbers [2 7 9 11] describe the notes: D, G, A, B. These examples show there are duplicate chord entries—as there should be, since the same note combinations can arise in multiple modes, keys, and scales. Next we build a smaller matrix grouping together the chords that are equal up

to permutation. (The only permutations appearing are circular shifts, due to the nature of the construction of the matrix `chordsModesKeysScalesRoots`.)

```
1  sameChordNotes=[];
2  m=0;
3  for i=1:size(chordsModesKeysScalesRoots,1)
4  % ^runs down rows of chordsModesKeysScalesRoots
5   if ismember(i,sameChordNotes)==0
6   % ^checks if row number has already been assigned
7    n=0; % resets sameChordNotes column to 0
8    m=m+1; % increments sameChordNotes row
9    for j=i:size(chordsModesKeysScalesRoots,1)
10   % ^checks all chords after current index
11    for k=0:3
12     if isequal(chordsModesKeysScalesRoots(j,[8:11]), ...
13      circshift(chordsModesKeysScalesRoots(i,[8:11]),k))==1
14      % ^checks if the notes are the same up to shift
15      n=n+1;
16      sameChordNotes(m,n)=j;
17  end    end    end    end    end
```

The matrix `sameChordNotes` turns out to have dimensions 393×32, meaning that there are 393 possible four-note chords in all keys of 12-TET drawn from the major, melodic minor, harmonic minor, and harmonic major scales (and that the same chord appears in at most 32 rows in `chordsModesKeysScalesRoots`). Each row of `sameChordNotes` lists the rows of `chordsModesKeysScalesRoots` that contain the same chord. For example, row 30 of `sameChordNotes` is:

```
[37 82 107 136 177 222 247 276 1137 1182 1207 1236 1277 1322 ...
1347 1376 3922 3947 3976 4017 4342 4367 4396 4437 0 0 0 0 0 0 ]
```
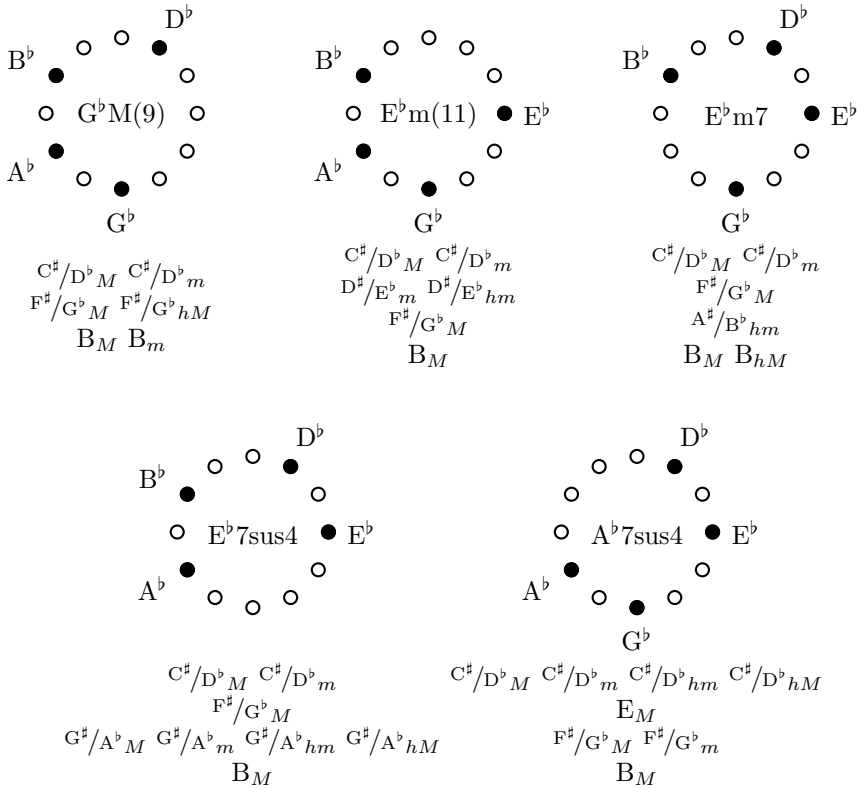
This is the row of `sameChordNotes` dedicated to (permutations of) the chord G-A-B-D, which is why it lists the rows 222 and 4437 of `chordsModesKeysScalesRoots` that we displayed earlier.

To create the lists of "diatonic citizenship" in Sect. 5, we wrote additional code that takes four notes as input, looks up their information in `sameChordNotes` and `chordsModesKeysScalesRoots`, and returns all the keys and scales where the four-note chord lives, formatted for LaTeX . We also wrote code that turns MATLAB output into vector graphics by way of PSTricks. Due to space limitations, we must omit this code. It is available upon request.

5   Results

Let us look at all five tetrachords that completely plug the semitonal gaps in the C major scale. (There is more than one possible name for each chord. Names below were assigned for convenience.) Beneath each chord, we list its diatonic citizenship, adopting italicized abbreviations for scale types: *M* for the major

scale, m for the melodic minor scale, hm for the harmonic minor scale, and hM for the harmonic major scale. For chord types, we use "M" for major triads and "m" for minor triads. Diatonic citizenship was determined according to the MATLAB code and scale restrictions in Sect. 4.

$\mathrm{G^{\flat}M}(9)$

$\mathrm{C^{\sharp}/D^{\flat}}_M$ $\mathrm{C^{\sharp}/D^{\flat}}_m$
$\mathrm{F^{\sharp}/G^{\flat}}_M$ $\mathrm{F^{\sharp}/G^{\flat}}_{hM}$
B_M B_m

$\mathrm{E^{\flat}m}(11)$ ● $\mathrm{E^{\flat}}$

$\mathrm{C^{\sharp}/D^{\flat}}_M$ $\mathrm{C^{\sharp}/D^{\flat}}_m$
$\mathrm{D^{\sharp}/E^{\flat}}_m$ $\mathrm{D^{\sharp}/E^{\flat}}_{hm}$
$\mathrm{F^{\sharp}/G^{\flat}}_M$
B_M

$\mathrm{E^{\flat}m7}$ ● $\mathrm{E^{\flat}}$

$\mathrm{C^{\sharp}/D^{\flat}}_M$ $\mathrm{C^{\sharp}/D^{\flat}}_m$
$\mathrm{F^{\sharp}/G^{\flat}}_M$
$\mathrm{A^{\sharp}/B^{\flat}}_{hm}$
B_M B_{hM}

$\mathrm{E^{\flat}7sus4}$ ● $\mathrm{E^{\flat}}$

$\mathrm{C^{\sharp}/D^{\flat}}_M$ $\mathrm{C^{\sharp}/D^{\flat}}_m$
$\mathrm{F^{\sharp}/G^{\flat}}_M$
$\mathrm{G^{\sharp}/A^{\flat}}_M$ $\mathrm{G^{\sharp}/A^{\flat}}_m$ $\mathrm{G^{\sharp}/A^{\flat}}_{hm}$ $\mathrm{G^{\sharp}/A^{\flat}}_{hM}$
B_M

$\mathrm{A^{\flat}7sus4}$ ● $\mathrm{E^{\flat}}$

$\mathrm{C^{\sharp}/D^{\flat}}_M$ $\mathrm{C^{\sharp}/D^{\flat}}_m$ $\mathrm{C^{\sharp}/D^{\flat}}_{hm}$ $\mathrm{C^{\sharp}/D^{\flat}}_{hM}$
E_M
$\mathrm{F^{\sharp}/G^{\flat}}_M$ $\mathrm{F^{\sharp}/G^{\flat}}_m$
B_M

Recall that we are interested in widening the palette of available "colors" in altered chords, where the vague notion of "color" could refer to a chord's internal intervallic structure or to its external diatonic context. First, regarding internal intervallic structure: Note that all of these alternative altered chords lack the tritone interval present in a V7 chord; and the seconds and fourths in these chords generally outnumber the major and minor thirds.

Second, regarding external diatonic context: Observe that the use of any of these alternative altered chords could be framed as temporary modulation from the key of C major into $\mathrm{C^{\sharp}/D^{\flat}}$ major, $\mathrm{C^{\sharp}/D^{\flat}}$ melodic minor, B major, or $\mathrm{F^{\sharp}/G^{\flat}}$ major—keys that are adjacent to C major ($\mathrm{C^{\sharp}/D^{\flat}}$, B), or directly across the circle of fifths ($\mathrm{F^{\sharp}/G^{\flat}}$). Therefore we may say that substituting alternative altered chords for G7 in the II–V–I progression Dm7–G7–CM7, as opposed to substituting the usual altered chords drawn from the G altered scale, maximizes Schönberg's *schwebend*, in the sense that $\mathrm{C^{\sharp}/D^{\flat}}$ major, $\mathrm{C^{\sharp}/D^{\flat}}$ melodic minor, B major, and $\mathrm{F^{\sharp}/G^{\flat}}$ major share fewer notes with C major than the G altered scale (which is $\mathrm{G^{\sharp}/A^{\flat}}$ melodic minor) shares with C major.

In conclusion, refering to the itemized list of compromises in Sect. 2, we advocate for V7 substitutions in jazz that let (i), chromatic voice leading, take precedence over (ii), preservation of the V7's dominant function, as well as over (iii), maintaining common tones between two chords. As Schönberg writes in [5], "harmonic usage is often created by coincidences of voice leading" (pg. 115).

6 Future Directions

Composition. By losing the essential chord tones of G7, the substitutions we propose effectively erase the V in the II–V–I cadence in C major. How would jazz sound if every V were replaced with these alternative altered chords? Do we need to preserve the essential chord tones of G7? (Recall that the essential chord tones are the 1, 3, and 7—so, the G, B, and F.) In other words, do we really need the V in a II–V–I, or do we just need a chord that adds tension by temporarily transporting us from the tonic while also serving as the chromatic glue between the II and the I? We will rewrite a few jazz standards using our alternative altered chords instead of dominant seventh chords, and we will disseminate the recordings online. Our code could be easily modified to investigate additional scales (*e.g.* octatonic), or to map out four-note chords in 31-TET and other non-standard tuning systems. We look forward to collaborating with microtonal guitarists on this front.

Analysis. This work is part of our long-term goal of applying the rich ideas in [1] and [6] to the jazz tradition. Specifically, we hope to explore the voice-leading dance of nearly even tetrachords—instead of nearly even triads—in and around (perfectly even) diminished seventh chords—instead of augmented triads—by analyzing examples from well-known jazz recordings, quantifying notions of "color" in chromatic harmony along the way.

References

1. Cohn, R.: Audacious Euphony: Chromatic Harmony and the Triad's Second Nature. Oxford Studies in Music Theory, Oxford University Press, USA (2012). https://books.google.co.in/books?id=rZxZCMRiO9EC
2. Ruth, L.: Beyond Nearly Even Chords: Graph-Based Modulation Exercises for Guitar, Mathematics of Music Interest Group meeting, Society for Music Theory Annual Meeting (virtual), 7 November 2021
3. Ruth, L.: Four-note Guitar Chord Modulation Exercise (2021). https://github.com/LCRuth/4noteGuitarChordModulationExercise
4. Ruth, L.: Beyond Nearly Even Chords: Graph-Based Modulation Exercises for Guitar. In: 21st Century Guitar Conference (virtual). https://21cguitar.com/schedule-2022, 17 March 2022
5. Schoenberg, A., Carter, R., Frisch, W.: Theory of Harmony. University of California (2010). https://books.google.com/books?id=pdDsQQAACAAJ
6. Tymoczko, D.: A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice. Oxford Studies in Music Theory, Oxford University Press (2010). https://books.google.com/books?id=1Jpq5BRLCNoC

Information Synthesis of Time-Geometry QCurve for Music Retrieval

Shannon Steinmetz$^{(\boxtimes)}$ and Ellen Gethner

University of Colorado, Denver 80204, USA
{shannon.steinmetz,ellen.gethner}@ucdenver.edu

Abstract. We expand information segmentation to include additional properties of music geometry. We establish a distinct metric for invariant chord structure (harmonic consistency) and models for conjunct melodic motion and acoustic consonance. We combine these with centricity to form a unified measure of music geometry. Using geometric predictors and the LSQOP method, we classify music/non-music with comparable results to AI/ML, between 76% and 92% f-score.

Keywords: Music geometry · Information geometry · Harmonic consistency · Harmonic leading · Centricity · Dissonance · Consonance · Time-geometry · Qcurve · Quant-curve · Music retrieval · Music detection · LSQOP

1 Introduction

Music information retrieval (MIR) dominates audio classification, rhythm, melody, genre and emotion (MER) [19]. MIR began with self-similarity [8], but focuses now on neural networks (NN, CNN, RNN) [10] and support vector machines (SVM) [20]. Artificial intelligence (AI) and machine learning (ML) have proven successful with f-score as high as 85% [12,14]; however, AI/ML is burdened by data availability, supervision and labeling. This also means pre-processing (e.g. CUSUM, MLR, GLR, KCD [5,11]) is a major factor in finding valid segments to process.

We use Tymoczko's properties of music [23] as the basis for geometric audio segmentation. Therefore, we further develop sufficient models for these properties as segmentation estimators [22]. If we combine geometrically segmented (time/geometry) audio curves of the same class such as music, we arrive at *principal curves*, which are the foundation for testing geometric variance (i.e. classifying audio). Using a technique called LSQOP we will demonstrate classification for music and non-music, comparable to AI/ML [9,15].

2   QCurve Transform

Steinmetz and Gethner developed centricity segmentation via three parameter Gamma and geodesic likelihood [1,2,22]. The following sections expand this by modeling additional properties where the result is what we call time/geometry, or *qcurve*.

Definition 1. *(QCurve) A* qcurve *is a time series consisting of positive unitary measures of musical geometry.*

3   Harmonic Consistency

Harmonic consistency states "harmonies in a passage of music, whatever they may be, tend to be structurally similar to one another [23]."

Definition 2 (Musical Structure). *Let* $K_{12} = (V, E)$, $V = \{0, 1, \cdots, 11\}$ *be a complete graph with vertices labeled* $\{0 = C, 1 = C\#, \cdots, 11 = B\}$. *Every distinct edge and cycle* C_r $3 \leq r \leq 12$ *having non-crossing edges in the embedded* K_{12} *is a* musical structure.

Definition 3. *Two non-empty sets of musical frequencies are* similar *if information gain due to geometric consistency is zero, or very small, such that distance* $D(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(i+1)}) \leq \epsilon$ *where* ϵ *is a fixed positive value.* $D(\cdot)$ *is dependent on probabilities* $p(x_1 \mid \boldsymbol{\theta}^{(i)})$, $p(x_2 \mid \boldsymbol{\theta}^{(i+1)})$ *from [22].*

Definition 3 expands on Cont's idea of similarity [3] except here, we depend on geometric divergence. There are trivial, but useful mapping between vertices of a 12-gon and $\mathbb{Z}_{12}$ using the complex unit circle $z_s = f(s) = e^{i\pi(15-s)/6}$, $s \in \mathbb{Z}_{12}$. Assume musical frequency $k \in \mathbb{R}$, where $T : \mathbb{R} \to \mathbb{Z}$ such that $T(k) = \{0, 1, 2, \ldots, 11\}$ (i.e. pitch class). We define $\boldsymbol{S_t}(k_i, k_j)$ as shortest distance between Z_{12} elements, or *integer separation* between frequencies. Due to chroma, there exists a distinct, linearly independent, invariant Euclidian distance for every 12-tone musical frequency pair, therefore frequencies of identical integer separation are similar, which we will prove.

Because $s = f^{-1}(z) = \frac{6 \arg(z)}{\pi}$ and $f^{-1}(a) = f^{-1}(b) \Rightarrow \frac{6 \arg(f(a))}{\pi} = \frac{6 \arg(f(b))}{\pi} \Rightarrow a = b$, f is injective. Let γ be inner angle difference between complex arguments. Due to injectivity, integer separation is modeled $\gamma = \gamma_i - \gamma_j = \frac{\boldsymbol{S_t}(k_1, k_2)\pi}{6}$.

Lemma 1. *Every pair of musical frequencies* k_i, k_j *has invariant Euclidian distance*

$$2 \sin \left(\frac{\boldsymbol{S_t}(k_i, k_j)\pi}{12} \right). \tag{1}$$

Proof. Let $z_i, z_j \in \mathbb{C}$, $z_i \neq z_j$. We define magnitude $|\cdot| \equiv ||\cdot||_2$. Due to the law of cosines $|z_i - z_j|^2 = |z_i|^2 + |z_j|^2 - 2|z_i||z_j| cos(\theta)$. Since $|z| = 1$, $|z_i - z_j|^2 = 2(1 - cos(\theta))$. If $\gamma_i = arg(z_i)$, $\gamma_j = arg(z_j)$, then by the dot product

$$\theta = cos^{-1}(cos(\gamma_i)cos(\gamma_j) + sin(\gamma_i)sin(\gamma_j)).$$

Substituting the angle and replacing identities gives $|z_i - z_j|^2 = 2(1 - cos(\gamma_i)cos(\gamma_j) + sin(\gamma_i)sin(\gamma_j))$. Multiply by $\frac{1}{4}$ and substitute haversine

$$\frac{|z_i - z_j|^2}{4} = \frac{1 - cos(\gamma_i - \gamma_j)}{2} \Rightarrow |z_i - z_j| = 2\,sin(\frac{\gamma_i - \gamma_j}{2}).$$

Constrained to the positive domain and substituting γ leaves

$$2\,sin\left(\frac{\boldsymbol{S}_t(k_i, k_j)\pi}{12}\right) = |z_i - z_j|.$$

Since rotation is unitary, all distinct pairs of musical frequencies have invariant Euclidian distance of this form. $\square$

Lemma 1 is chroma-agnostic, linearly dependent, dyad similarity, but removing homogeneity makes all dyad pairs linearly independent under this mapping, therefore

$$\text{tone distance} = \delta_t(k_i, k_j) = 2\,sin\left(\frac{\boldsymbol{S}_t(k_i, k_j)\pi}{12}\right) + \boldsymbol{S}_t(k_i, k_j). \qquad (2)$$

Tone distance is symmetric $\delta_t(k_i, k_j) = \delta_t(k_j, k_i)$, invariant under rotation and satisfies triangle inequality. Because $k_i = k_j$ implies $\delta_t(k_i, k_j) = 2 \cdot sin(0) + 0 = 0$, (2) is a *metric space*. It follows frequencies are similar if and only if, tone distance are equivalent. *Path distance* $= P_\delta(F_{(i)})$ is defined as the sum of tone distance, assuming frequencies $F_{(i)} = \{k_j\}_{j=1}^n$. Using logical reduction and brute force search, we find *no duplicate* path distances among all 12-tone chords. Temporally, inverse harmonic consistency $= \text{HC}^{-1} = \sigma(\boldsymbol{H}) \le M_h = 18.24$, where σ is standard deviation, $\boldsymbol{H} = \{P_\delta(F_{(i)})\}_{i=1}^\infty$ and $F_{(i)}$ are sequential. Given no discernable notes, noise, or silence $\text{HC}^{-1} = M_h$.

4   Harmonic Leading

Harmonic leading is the combined measure of harmonic consistency, voice leading and conjunct melodic motion (CMM). CMM is the tendency for "melodies to move by short distances from note to note [23]." Due to [4] *musical metric distance* $\Delta(k_1, k_2)$ can be leveraged assuming $\boldsymbol{P} \in \mathbb{R}^{m \times n}$, whose columns are frequencies sorted in ascending order with $\boldsymbol{P}_{ij} = 0$, when m differs. From this, we approximate conjunct melodic motion

$$\overline{\nabla}(\boldsymbol{P}) = \left|\sum_{j=1}^{n-1}\sum_{i=1}^{m} \min\{\Delta(\boldsymbol{P}_{ij}, \boldsymbol{P}_{i(j+1)})\}\right|, \qquad (3)$$

which also dampens overfit[1]. Assuming $\sigma(\cdot) = \text{stddev}$, $f = \text{frame count}$ and $\Delta(\cdot) \le M_v = 6$ [4]

$$\text{harmonic leading} = 1 - \frac{\sigma(\boldsymbol{H}) + \overline{\nabla}(\mathbf{P})}{M_v(f - 1) + M_h}. \qquad (4)$$

---

[1] Harmonic leading *overfit* is defined as acoustically perceivable chord transposition, or inversion.

As a side note, we tested harmonic leading on Harte's 16 songs [13] with moderate success of 58% f-score, but we were unable to match HCDF retrieval. We recommend an HCDF quantifier experiment using our information framework, which we leave as an open problem.

5   Consonance

Definition 4. Acoustic consonance *is the inverse of dissonant contributions between* 20 *and* 250 *Hz of center frequency, within critical band* β *[7, 17, 18].*

Given $F(k) = |FFT|$, we select $\tau = |k - \text{argmax}_x [F(k) \cdot F(k+x)]|$, $x \leq 250$ which implies

$$\text{congruence score} = C_\tau(k) = \frac{1}{\max\{|F|\}} \sum_{x=k-\beta/2}^{k+\beta/2} F(x) \, f_\tau(x), \tag{5}$$

correlating $f_\tau(x) = \max\{F\}/2 \left[cos(2\pi x/\tau) + 1 \right]$. We center on $k_c : |k_c - k| \leq \beta/2$, selecting $k_j : F(k_j) \geq \frac{\max(F)}{4}$ and observe $\boldsymbol{C} = \left\{ |C_{\tau_j}(k_j)| \right\}_{j=1}^n$ converge to a block wave as dissonant contribution increases. Therefore,

$$1 - \text{consonance} = \text{dissonance} = \frac{\max(\boldsymbol{C})}{2M_d} \sum_{j=1}^n D \cdot \boldsymbol{C}_j, \tag{6}$$

where max dissonance $= M_d \approx 10$ and $D = \text{sgn} \left[\cos(\frac{2\pi j}{s}) \right] + 1$ over bandwidth $s \leq \beta$. β is displacement between the two loudest frequencies in critical band.

6   LSQOP

Qcurves are synthesized using [22] and combinations of $AC = $ acoustic consonance, $C = $ centricity, $HL = $ harmonic leading and $MG = $ music geometry.

$$MG = 1 - \frac{2}{\pi} cos^{-1} \left(\frac{U \cdot V}{\sqrt{3} \, ||V||} \right) \tag{7}$$

is overlap between ideal $U = (1, 1, 1)$ and measured vector $V = $ (AC, C, HL). We observe ordinal separation between qcurves of differing classes, exposing an opportunity to train *principal qcurves* as a model for variation. Figure 1 illustrates least squares ortho projector (LSQOP) method, assuming input audio qcurve $q(x)$, principal qcurve trend-line segments $\{p_0, p_1\}$(music), $\{p_2, p_3\}$(non-music) and arbitrary point $d = (x, q(x))$. There exist $x_1 = p_0 + \text{Res}_{p_1 - p_0}(d - p_0)$ and $x_2 = p_1 + \text{Res}_{p_3 - p_2}(d - p_2)$ assuming $\text{Res}(\cdot)$ implies vector resolute. $\boldsymbol{P} = \frac{(0,1)(0,1)^T}{(0,1)^T(0,1)}$, $a_1 = ||\boldsymbol{P}d||$, $a_2 = ||\boldsymbol{P}x_1||$ and $a_3 = ||\boldsymbol{P}x_2||$, yields ratio

$$\mathfrak{m}(x) = \begin{cases} 1 & a_1 > a_2 \\ 0 & a_1 < a_3 \text{ or } a_3 > a_2 \\ \frac{||a_1 - a_3||}{||a_2 - a_3||} & \text{otherwise.} \end{cases} \tag{8}$$
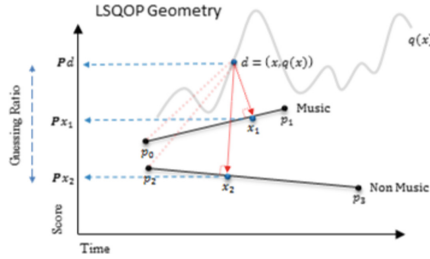
Fig. 1. Classification of audio qcurve $q(x)$, contrasted with trend lines from trained principal qcurves for music (top) and non-music (bottom).

Musical probability $\mathfrak{M} = E[m(x)]$ has binary pass/fail $\mathfrak{M} \geq \epsilon$ with expectation $E[\cdot]$ and non-negative *musical threshold* $\epsilon \leq 1$. Threshold varies depending on curve training.

7 Evaluation and Analysis

Verification of LSQOP and time/geometry involved classification experiments with GTZAN [24], TUT-17 (parts 1&2) [16], SWS1 (used by [6,21]), SWS2 (music), SWS3 (non-music) and SWS4 (non-music) data. The database contains 1556 music and 1441 non-music, totaling 2776 files. Custom data were created to fool LSQOP due to abnormal accuracy on GTZAN (100%) and TUT (92%). Custom data contains random quality, sample rate, content, size and non-thematic clips from *samplefocus.com*, *partnersinrhyme.com*, *bensound.com*, *freemp3cloud.com* and *BBC Sound Effects*. Custom non-music sets contain several categories (e.g. people, urban, construction, natural, office, animals, household, video games, military/war, etc.). Custom music sets are spread (mostly) even across several genres with famous, lesser known artists, synthesised and "poor" quality.

We performed three tests using LSQOP for music/non-music classification. The first and second test measures accuracy on all 2997 files. In the first we process the starting 10s of each file and in the second we process [.2, 3]s random samples from each file. Assuming notation (score/segment) (see [22]), (MG/HL) was effective with non-random samples at 96.4% accuracy for music and 70.78% classifying non-music. (MG/AC) was effective on random samples with 80% accuracy for music and 78% for non-music.

Table 1 is the result of an information retrieval (IR) exercise involving requests for music/non-music from a database (not all tests shown). Small randomly populated datasets are drawn from the entire clip database and each set is guaranteed to carry (roughly) equally distributed music and non-music. The results here are f-scores between 76% and 92%, averaging around 81.6%.

Table 1. Classification on randomly constructed subsets of the database. $\mathbf{M}$ = Music, $\mathbf{N}$ = Non-music, $\mathbf{n}$ = Num Random Samples, $\mathbf{p}$ = Precision, $\mathbf{r}$ = Recall, $\mathbf{f}$ = FScore. Training data shorthand: $A \equiv$ (SWS2/SWS4), $B \equiv$ (SWS2/TUT), $C \equiv$ GTZAN/SWS[1–3]. This table was inspired by [13].

Quantifier	Mn	Mp	Mr	Mf	Nn	Np	Nr	Nf
MG-C 10s	28	86%	92%	89%	28	93%	87%	90%
(AC/C)-B 10s	27	80%	92%	86%	27	92%	79%	85%
(MG/AC)-A 6s	22	100%	80%	89%	22	86%	100%	92%

8   Conclusions

We showed how 12-tone chroma-agnostic frequency pairs are mapped to a distinct, linearly independent, invariant measure of harmonic distance as a metric space. We provided a unique measure of musical chords (path distance) and discussed how to quantify harmonic consistency over successive frames. By combining this idea with work in voice leading we modeled conjunct melodic motion as harmonic leading. We then developed a straightforward approximation of acoustic consonance using psychoacoustic theory. From these measures we devised a unified score for musical geometry and showed how principal qcurves combined with LSQOP is effective in retrieval with consistent f-scores between 76% and 92%, averaging 81.6%.

Individually, the proposed musical property models have independent value for measuring structural content and acoustic quality analysis. The success of LSQOP is clear, but optimal time/geometry combinations require further experiments. Because audio data were successfully used to train accurate results against other audio, this opens the question of ideal data for use in widespread classification. We must also consider the disparity between music and non-music success, which itself can present a challenge to blind classification using the same geometric properties for music and non-music.

References

1. Abdel-All, N.H., Abdel-Galil, E.: Numerical treatment of geodesic differential. In: International Mathematical Forum, vol. 8, pp. 15–29 (2013)
2. Chen, W.W., Kotz, S.: The riemannian structure of the three-parameter gamma distribution (2013)
3. Cont, A., Dubnov, S., Assayag, G.: On the information geometry of audio streams with applications to similarity computing. IEEE Trans. Audio Speech Lang. Process. **19**, 837–846 (2010)
4. del Pozo, I., Gómez, F.: Formalization of voice-leadings and the Nabla algorithm. In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O.A. (eds.) MCM 2019. LNCS (LNAI), vol. 11502, pp. 352–358. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_30

5. Desobry, F., Davy, M., Doncarli, C.: An online kernel change detection algorithm. IEEE Trans. Signal Process. **53**, 2961–2974 (2005)
6. Gethner, S.S.E., Verbeke, J.: A view of music. In: Delp, D.M.K., Kaplan, C.S., Sarhangi, R. (eds.) Proceedings of Bridges 2015: Mathematics, Music, Art, Architecture, Culture, Phoenix, Arizona, Tessellations Publishing, pp. 289–294 (2015). http://archive.bridgesmathart.org/2015/bridges2015-289.html
7. Fishman, Y.I., et al.: Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. J. Neurophysiol. **86**, 2761–2788 (2001)
8. Foote, J.T., Cooper, M.L.: Media segmentation using self-similarity decomposition. In: Storage and Retrieval for Media Databases, vol. 5021, pp. 167–175. International Society for Optics and Photonics (2003)
9. Gimeno, P., Mingote, V., Giménez, A.O., Miguel, A., Lleida, E.: Partial auc optimisation using recurrent neural networks for music detection with limited training data. In: Interspeech, pp. 3067–3071 (2020)
10. Gururani, S., Summers, C., Lerch, A.: Instrument activity detection in polyphonic music using deep neural networks. In: ISMIR, pp. 569–576 (2018)
11. Gustafsson, F.: The marginalized likelihood ratio test for detecting abrupt changes. IEEE Trans. Autom. Control **41**, 66–78 (1996)
12. Hamel, P., Eck, D.: Learning features from music audio with deep belief networks. In: ISMIR, vol. 10, pp. 339–344. Citeseer (2010)
13. Harte, C., Sandler, M., Gasser, M.: Detecting harmonic change in musical audio. In: Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia, pp. 21–26 (2006)
14. Kataoka, M., Kinouchi, M., Hagiwara, M.: Music information retrieval system using complex-valued recurrent neural networks. In: SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 98CH36218), vol. 5, pp. 4290–4295. IEEE (1998)
15. Meléndez-Catalán, B., Molina, E., Gomez, E.: Music and/or speech detection mirex 2018 submission, music information retrieval evaluation eX-change (2018)
16. Mesaros, A., Heittola, T., Virtanen, T.: Tut acoustic scenes 2017, evaluation dataset (2017)
17. Moore, B.C., Glasberg, B.R.: Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J. Acoust. Soc. Am. **74**, 750–753 (1983)
18. Nordmark, J., Fahlén, L.E.: Beat theories of musical consonance. Q. Prog. Status Rep. **29**, 111–122 (1988)
19. Panda, R., Malheiro, R.M., Paiva, R.P.: Audio features for music emotion recognition: a survey, IEEE Trans. Affect. Comput. (2020)
20. Schedl, M., Gutiérrez, E.G., Urbano, J.: Music information retrieval: recent developments and applications. Found. Trends Inf. Retrieval. **8**(2–3), 127–261 (2014)
21. Steinmetz, S., Sonic imagery: a view of music via mathematical computer science and signal processing. University of Colorado at Denver (2016)
22. Steinmetz, S., Gethner, E.: On musical information geometry with applications to sonified image analysis. In: To Appear in International Conference on Pattern Recognition and Image Processing, Miami (2021)
23. Tymoczko, D.: A Geometry of Music, Oxford University Press, Oxford, 1 (ed.) (2011)
24. Zhang, X.: Gtzan, November 2019

Investigating Style with Scale Embeddings

Matt Chiu[✉]

University of Rochester - Eastman School of Music, Rochester, NY 14604, USA
mchiu9@u.rochester.edu

Abstract. In this paper, we use pitch-class vector *embeddings* to study scale relationships between composers. Recent research in natural language processing (NLP) has used machine learning to derive vector representations-known as embeddings—for words based on their co-occurrence. Borrowing from NLP, we use the word2vec algorithm to encode windows of pitch-classes, or *pitch-class vectors*, of music. We show that these embeddings not only replicate the well-known theoretical circle of fifths, but can also capture stylistic nuances between composers' use of scales.

Keywords: Scale theory · Embeddings · Word2vec · Vector space · Style

1 Introduction

In Natural Language Processing (NLP), the *word2vec* algorithm is a technique for deriving vector representations-known as *embeddings*-for words by iterating through a corpus [7]. Embeddings contain information about the syntactic placement and the semantic similarity of words. In symbolic music research, recent studies have explored musical embeddings for chords [6], and motivic fragments [1], as well as applications for harmonic tension [8], and music generation [2]. In this paper, we used word2vec to derive *scale embeddings* from pitch-class collections of different composers and compare embeddings between and within their styles.

2 Methodology: Word2vec, Encoding Procedure, and Training

2.1 Word2vec Algorithm

As said, the aim of the word2vec model is to generate dense vector representations (embeddings) for words based on their co-occurrence. Each unique word in a corpus is represented with a corresponding vector, with words that occur near one another in the corpus having similar embedding vectors. Since words with semantic similarity have similar syntactic placement, they also have similar embedding vectors (whether or not they not co-occur). In this paper, we used the

skip-gram version of word2vec: given a corpus W of words w with surrounding context-words c, the algorithm maximizes the likelihood of surrounding words:

$$\arg\max_{\theta} \prod_{w \in W} \prod_{c \in C} P(c|w;\theta) \tag{1}$$

where v_c and v_w are vector representations of words v and c respectively and C is the set of all contexts, the probability $P(c|w;\theta)$ is calculated with the softmax function:

$$P(c|w;\theta) = \frac{exp(v_c v_w)}{\sum_{c\text{τ} \in C} exp(v_{c\text{τ}} v_w)} \tag{2}$$

2.2   Encoding Procedure

As the word2vec algorithm is designed to parse words in a corpus, we needed both a corpus and a method for encoding musical objects as words-our musical vocabulary. For the corpus, we used the Yale Classical Archives Corpus, henceforth YCAC [11]. The YCAC is a spreadsheet of 13,769 midi files of works by 571 composers. Each midi file is parsed into "slices" containing the set of vertical pitches at any new pitch onset, time points T as quarter-note offsets from the beginning of the score, and other metadata.

For each piece, we extracted pitch classes using a sliding windowing procedure. Where a piece $a_t, a_{t+x}, \ldots, a_T$ consists of slices A at corresponding time points T, a window of length m is a subsequence $s_t = a_t, a_{t+x}, \ldots, a_{t \leq t+m}$. Each piece therefore contained $T - m$ windows, with each window consisting of m adjacent slices. Pitch classes in a window were then encoded as fixed-length, *pitch-class vectors* $PC = \{pc_0, pc_1, ..., pc_{11}\} \in \{0,1\}^{11}$ where the pitch class with subscript 0=C, subscript 1=C#/Db, etc. Present pitch classes in a window were represented as 1, and absent pitches were represented as 0, i.e. the C-major scale, for example, is represented as $\{1,0,1,0,1,1,0,1,0,1,0,1\}$. In NLP, this encoding method is also known as *one-hot* encodings. Note that since our methodology had no concept of tonic, pitch-class vectors represented multiple scales simultaneously: C major and A natural minor have the same collection and were represented identically. This is a significant simplification of scale and should not be taken as the end goal. Rather, the work here represents a crude proof-of-concept where future work might use weighted pitch-class vectors to distinguish pitch-class salience [3].

We then windowed pieces for four well-represented composers in the corpus: Mozart, Liszt, Saint-Saens, and Debussy (Table 1). To find an appropriate window length, we approximated the standard 7-diatonic-pitch scale by averaging the number of notes per window for windows of length 2–10 quarter notes (Fig. 1). Reflecting the historical narrative that chromaticism progressively increased over time, Mozart and Liszt use fewer pitches per window than Debussy and Saint-Saëns. Given the results of Fig. 1, window size m was set to 6 quarter-note beats for Mozart and Liszt and 5 for Saint-Saëns and Debussy.

Table 2 shows the total number of windows for each composer and the percent of major/natural minor and harmonic minor collections for each. 27% of windows

Table 1. Sampled yale classical archives corpus data.

Composer	No. of pieces in YCAC	No. of notes	Avg. notes per piece
Mozart	882	3,865,439	4,382.58
Liszt	125	806,025	6,448.2
Saint-Saëns	72	504,663	6,913.19
Debussy	39	170,773	4,378.79
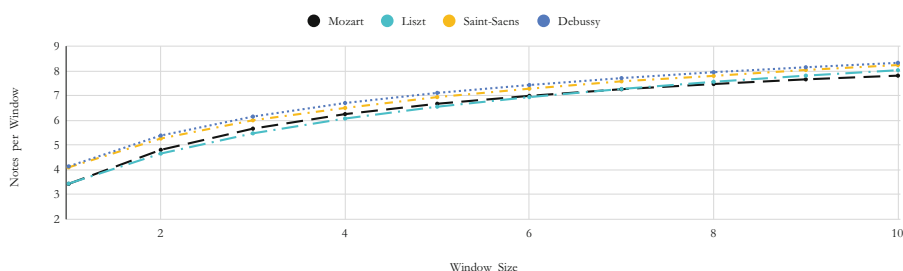
Fig. 1. Average number of notes with different window sizes, where window size is in quarter-note beats.

in the Mozart model were categorized as diatonic collections, whereas this was between 10%–14% for the other three composers, reflecting a larger variety of pitch-class collections for Liszt, Saint-Saëns, and Debussy. For a closer look, we looked at the top 5 most frequent pitch-class vectors for each composer (Table 3). Diatonic collections occupied the majority of positions in the table. For Liszt and Debussy, the most frequent pitch-class vector was the aggregate (all 12 pitches in a vector). In fact, the aggregate pitch-class vector for Liszt was more frequent than the next four diatonic sets combined. For Saint-Saëns, the most common pitch-class collection was an empty vector. These results show that Liszt and Debussy cycle through pitch classes at a faster notated rate than the others and that Saint-Saëns often has longer durations with no new pitch classes introduced.

Table 2. Frequency and percent of diatonic collections in windows.

Composer	Total no. of windows	Total major	Total harmonic minor	% diatonic
Mozart	464,229	98,490	11,806	27%
Liszt	82,179	6,442	2,282	11%
Saint-Saëns	40,643	4,645	1,093	14%
Debussy	13,923	1,931	125	15%

2.3 Model Parameters

Using the Gensim Python library, word2vec was then used to find embeddings for the resulting set of pitch-class vectors, mapping pitch-class vectors onto scale

Table 3. Top 5 frequent pitch-class vectors and their respective frequency sorted by composer.

Composer	#1 frequent	#2 frequent	#3 frequent	#4 frequent	#5 frequent
Mozart	C maj: 18,347	Bb maj: 17,580	D maj: 17,395	G maj: 16,189	Eb maj: 15,055
Liszt	All PCs: 3,187	F# maj: 927	C maj: 798	E maj: 726	A maj: 620
Saint-Saëns	No PCs: 1,067	C maj: 921	Eb maj: 853	All PCs: 689	E maj: 588
Debussy	All PCs: 487	E maj: 352	C maj: 258	A maj: 207	F maj: 186

embeddings. The scale embeddings here were set to $2^5 = 32$ dimensions. If a pitch-class vector c was within 6 beats of pitch-class vector w, it was included as a context pitch-class vector for w, notated as "Context Windows" and "Target Windows." For each target window, Word2vec maximizes the probability (using negative sampling) of context windows within 6 beats. We trained four models, one on each of the four composers, where each model iterated over the composer-corpus 20 times.

3   Properties of Embeddings

3.1   The Circle of Fifths According to the Mozart Model

Given the high dimensionality of embeddings, we used *t-Distributed Stochastic Neighbor Embedding* (t-SNE) to visualize the embeddings in a 2-dimensional space [10]. The left side of Fig. 2 shows the Mozart model's scale embeddings plotted in 2 dimensions. This figure resonates with music-theoretical claims about the circle of fifths (COF). The COF is a metric for scale distance: the further a scale on the COF, the more distant it is [5]. However, not all fifth-adjacent scales are equidistant. For example, A major is much closer to D major than it is to E major. This was likely the effect of absolute key: Mozart wrote more frequently in D major than in A major, and since pieces often modulated to their dominant in the classical period, A was drawn closer to D.

Including harmonic minor scales and clustering embeddings with euclidean distances still captured the COF (right side of Fig. 2). Each color in the figure are quadrants on the circle of fifths. There is only one minor scale–C# minor (marked with an asterisk)–located far away from its relative major.

3.2   Composer Embeddings Correlated with the Circle of Fifths

Calculating the cosine distance from C major to each other major scale for each of the four models resulted in Fig. 3. Beside the embedding distances, we plotted the distances according to the COF (normalized). COF distances were calculated based on a unit circle with 12 equidistant points from the center, measured with both angular and Euclidean metrics. Notably, each model in Fig. 3 makes an arch, signifying that distance from C gets further around the COF until reaching its diametrically opposed point (F#/Gb). The models roughly approximate the
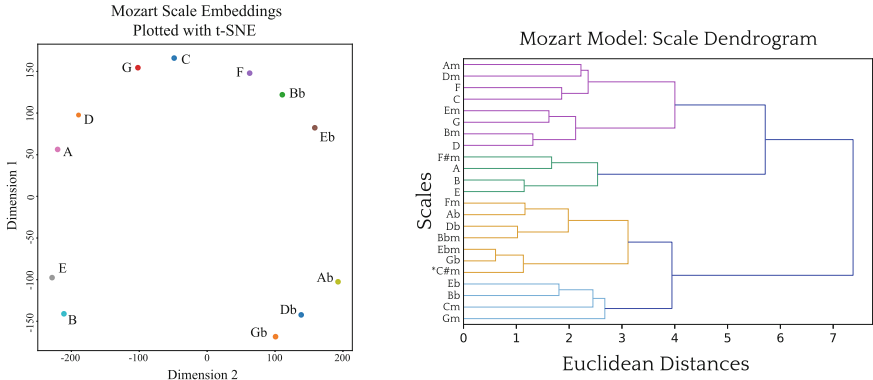
Fig. 2. Mozart model: major scale embeddings clustered with t-SNE (left) and major/minor scale embeddings clustered k-hierarchical clustering (right).

COF distances, and correlations also values also verify this claim. Correlations with the angular and Euclidean COF distances are, respectively, Mozart(.84, .9), Liszt(.78, .86), Saint-Saëns(.89, .94), and Debussy(.94, .96).
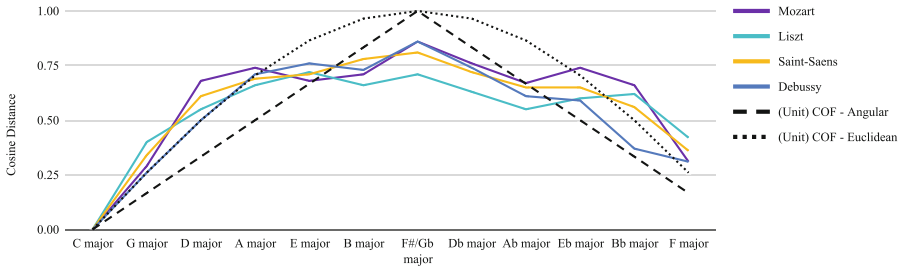

Fig. 3. Cosine distances from C-major scale embedding.

Examining Fig. 3 further reveals stylistic differences between composers. The Mozart model is relatively flat after two steps around the COF–approximately around .68. If we were to generalize this to other keys, this restates a well-known intuition: in the style of Mozart, fifths surrounding the tonic key are the most likely to be modulation goals. Despite having commensurate note-average-per-window values (Fig. 1), the Liszt model correlated less with the COF than the Mozart model (or any of the models, for that matter). This reveals his stylistic tendency to modulate to third-related keys [4,9]: Fig. 3 shows that Ab major–a scale 4 flats away on the COF–is closer to C than any other scale besides scales with keys with a single flat (F) or sharp (G).

Surprisingly, the Debussy model had a higher correlation with the COF (angular and Euclidean) than any of the other models. This could represent

the disentangling of the tonic-dominant key-relationship dichotomy: whereas other composers consistently modulate to fifth-related scales, drawing their fifth-related embeddings close together, the dominant's relationship is weighted less in Debussy's music.

4   Conclusion

We have shown that scale embeddings, encoded as pitch-class vectors, capture style-specific musical intuition about scale relationships within common-practice art music. The composers modeled here treat scales differently, resulting in nuanced distances between embeddings. Future avenues for research should first more accurately encode pitch-classes to correspond with scale-degree salience, and might then study the relationship between chord and scale embeddings, and, perhaps, how this interaction changes between and within composers over time.

References

1. Alvarez, A.A., Gómez-Martin, F.: Distributed vector representations of folksong motifs. In: Montiel, M., Gómez-Martin, F., Agustín-Aquino, O.A. (eds.) International Conference on Mathematics and Computation in Music, pp. 325–332. Springer, Heidelberg (2019). https://doi.org/10.1007/978-3-030-21392-3_26
2. Brunner, G., Wang, Y., Wattenhofer, R., Wiesendanger, J.: JamBot: Music theory aware chord based generation of polyphonic music with LSTMs. In: 2017 IEEE 29th International Conference on Tools with Artificial Intelligence, pp. 519–526. IEEE (2017)
3. Chiu, M.: Macroharmonic progressions through the discrete fourier transform: an analysis from Maurice Duruflé's requiem. Music Theory Online **27**(3) (2021)
4. Kopp, D.: Chromatic Transformations in Nineteenth-Century Music. Cambridge University Press, Cambridge (2002)
5. Krumhansl, C.L., Kessler, E.J.: Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. Psychol. Rev. **89**(4), 334–368 (1982)
6. Madjiheurem, S., Qu, L., Walder, C.: Chord2vec: learning musical chord embeddings. In: Proceedings of the Constructive Machine Learning Workshop at 30th Conference on Neural Information Processing Systems. Barcelona, Spain (2016)
7. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word represetations in vector space. arXiv preprint arXiv:1301.3781 (2013)
8. Nikrang, A., Sears, D.R., Widmer, G.: Automatic estimation of harmonic tension by distributed representation of chords. In: Aramaki, M., Davies, M.E.P., Kronland-Martinet, R., Ystad, S. (eds.) International Symposium on Computer Music Multidisciplinary Research, pp. 23–34. Springer, Cham (2017). https://doi.org/10.1007/978-3-030-01692-0_2
9. Riemann, H.: Große Kompositionslehre, vol. I. W. Spemann, Berlin (1902)
10. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**(11), 2579–2605 (2008)
11. White, C.W., Quinn, I.: The Yale-classical archives corpus. Empirical Musicology Rev. **11**(1), 50–58 (2016)

Author Index