Corpus Research and Choro

Potential and Challenges for Digital Methods

Fabian C. Moss (Würzburg, Germany)

2 April 2024 – Universidade Federal do Rio de Janeiro

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

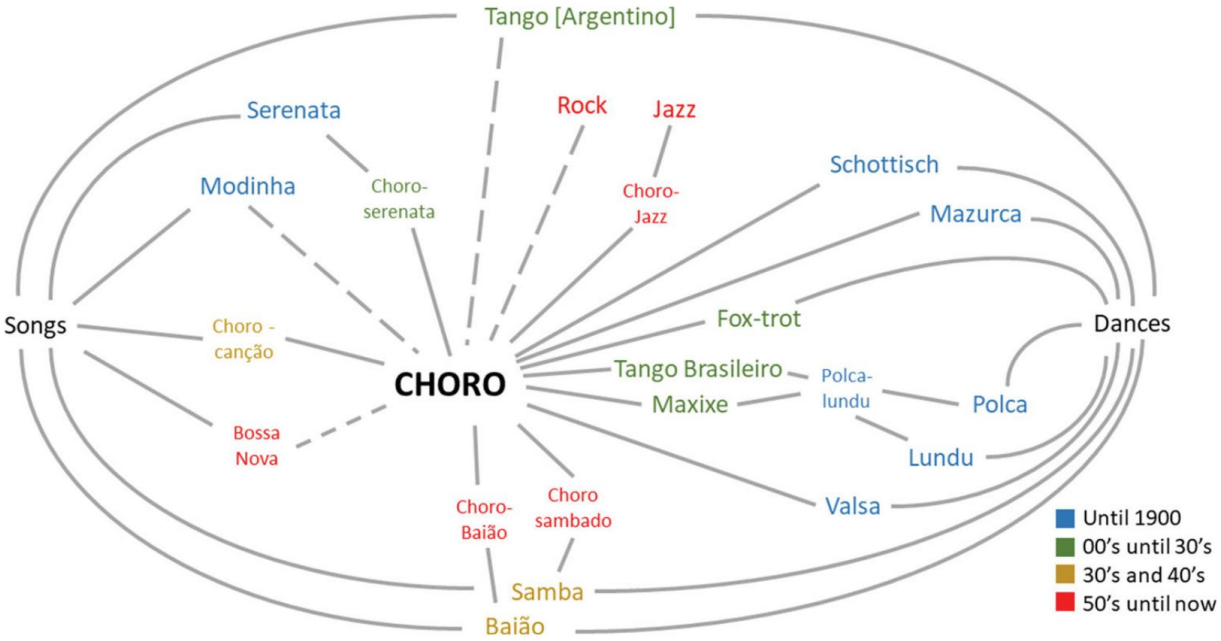
Bavarian
Research Alliance

About me

- from Cologne, DE
- background in mathematics & music education
- 2013: MA Musicology
- 2019: PhD Digital Humanities, Lausanne, CH
- 2022: Cultural Analytics, Amsterdam, NL
- Since 2023: Digital Music Philology and Music Theory, Würzburg, DE
- → Mathematical & Computational Music Analysis, Corpus Studies, Modeling

How I met your Choro

Willian Fernandes de Souza

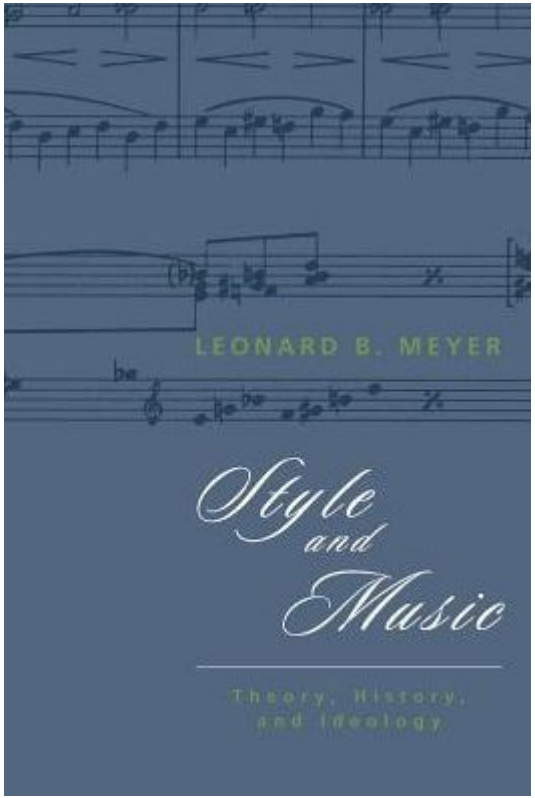


Moss, F. C., Souza, W. F., & Rohrmeier, M. (2020). Harmony and form in Brazilian Choro: A corpus-driven approach to musical style analysis. *Journal of New Music Research*, 49(5), 416–437.

Style

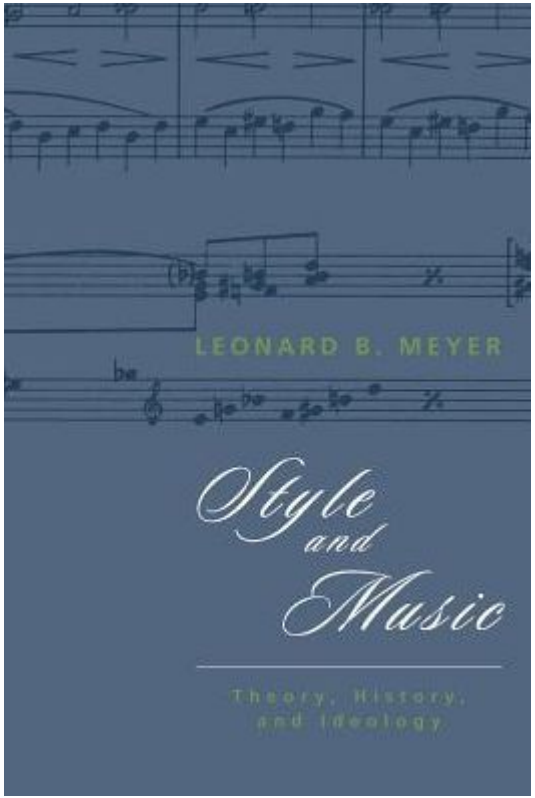
Style

L. B. Meyer (1989). *Style and Music: Theory, History, and Ideology.* Chicago University Press



Style

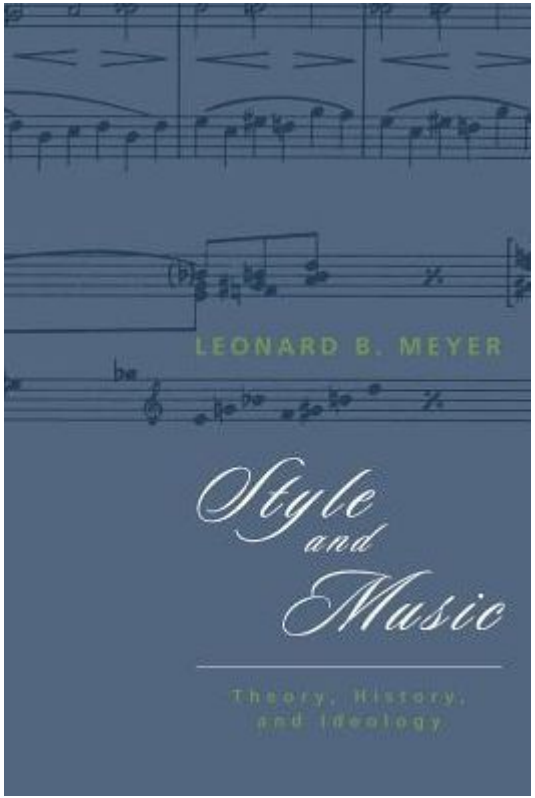
"**Style** is a replication of **patterning**, whether in human behavior or in the **artifacts** produced by human behavior, that results from a series of **choices** made within some set of **constraints**."



L. B. Meyer (1989). *Style and Music: Theory, History, and Ideology.* Chicago University Press

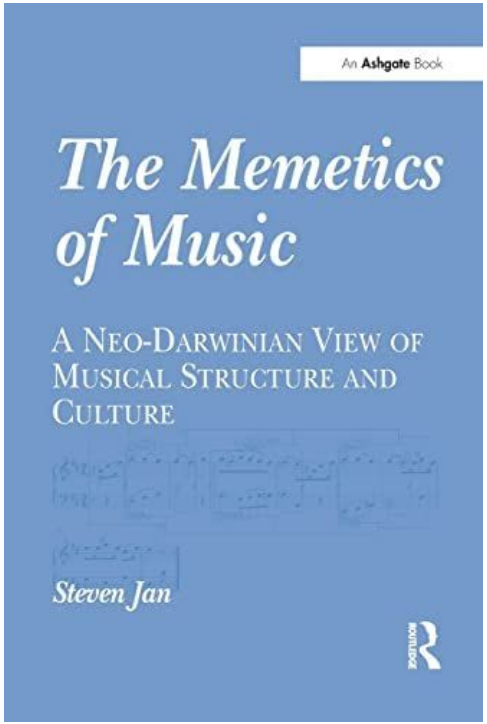
Style

"[**Style analysis** is] to **describe** the patternings replicated in some group of works, to **discover and formulate** the rules and strategies that are the basis for such patternings, and to **explain** in the light of these constraints how the characteristics described are related to one another."



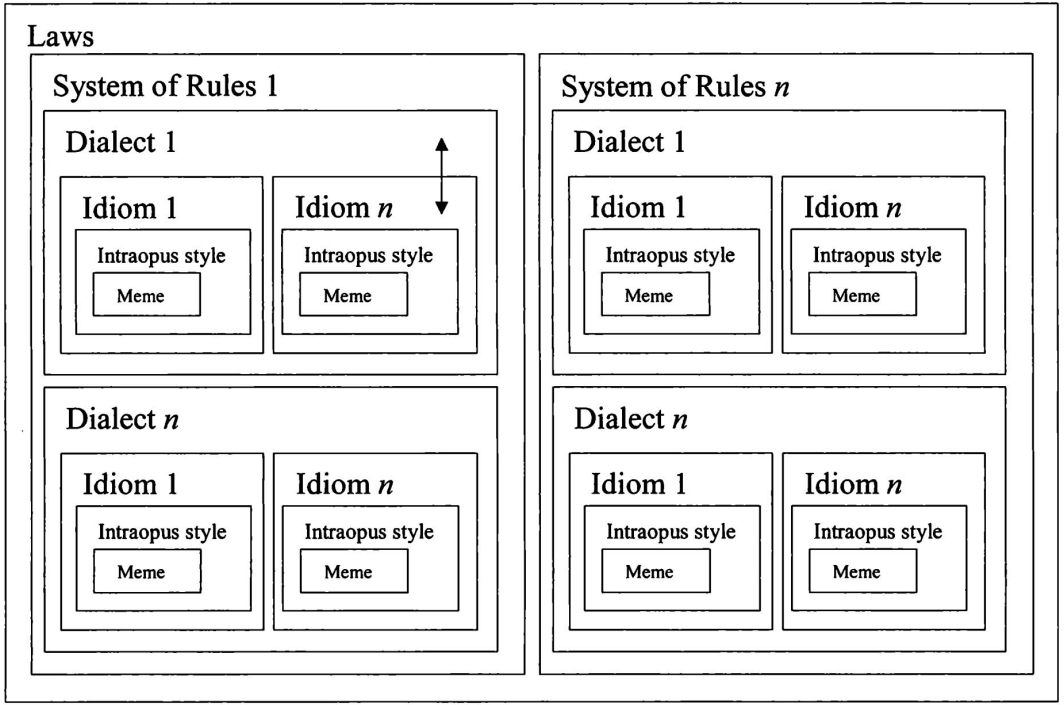
L. B. Meyer (1989). *Style and Music: Theory, History, and Ideology.* Chicago University Press

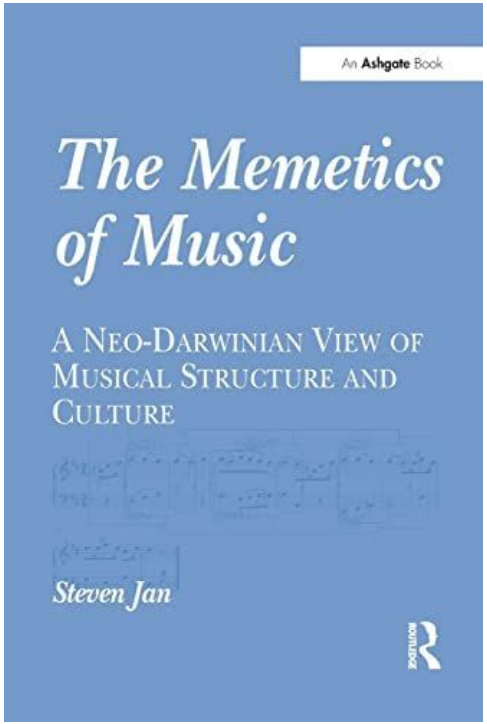
Style as a hierarchy of patterns



Steven Jan (2007). *The Memetics of Music: A Neo-Darwinian View of Musical Structure and Culture.*

Style as a hierarchy of patterns

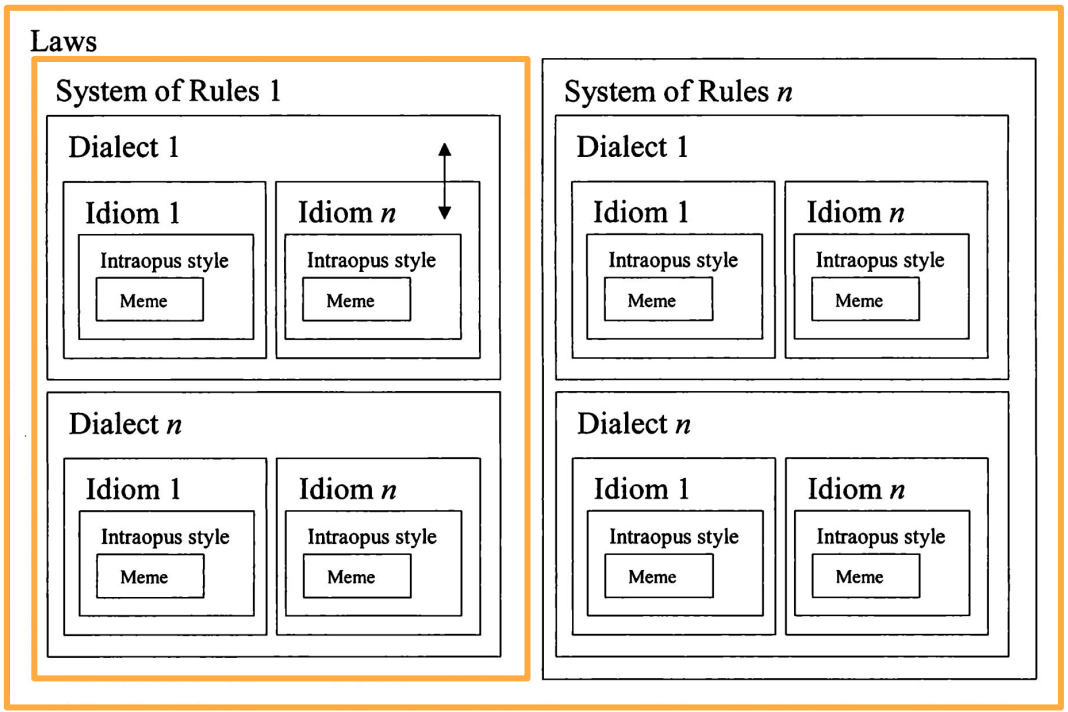


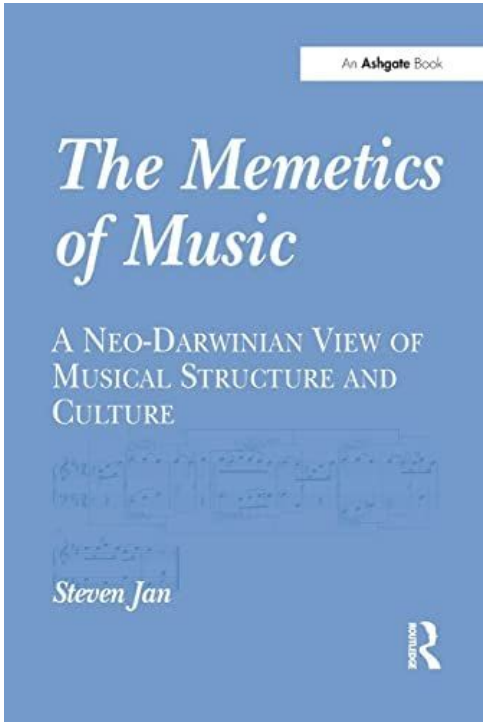


Steven Jan (2007). *The Memetics of Music: A Neo-Darwinian View of Musical Structure and Culture.*

Style as a hierarchy of patterns

Music Cognition

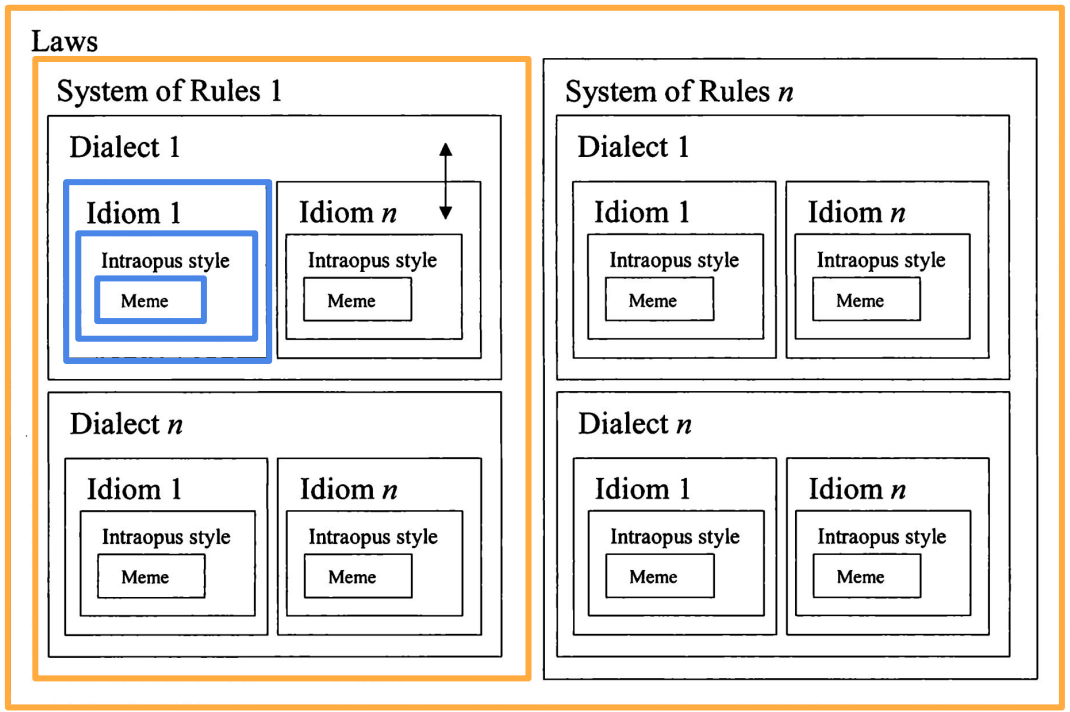


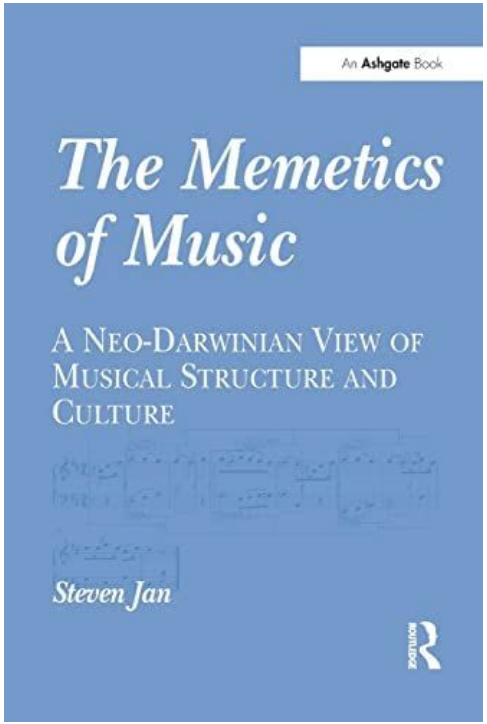


Steven Jan (2007). *The Memetics of Music: A Neo-Darwinian View of Musical Structure and Culture.*

Style as a hierarchy of patterns

Music Cognition Music Theory

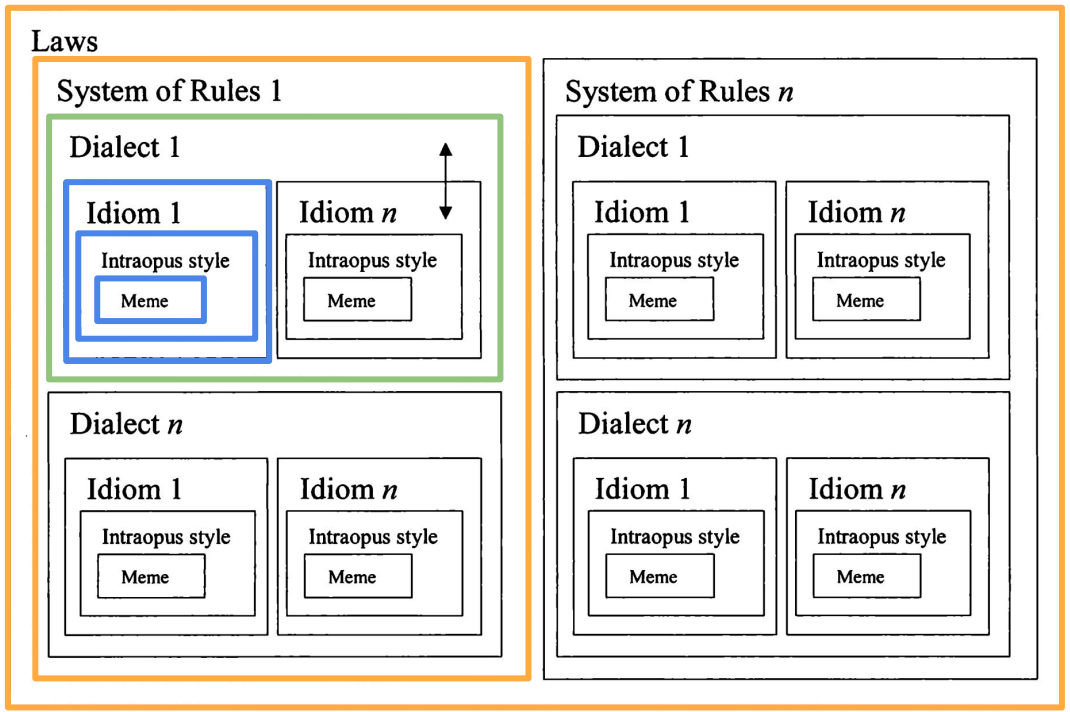


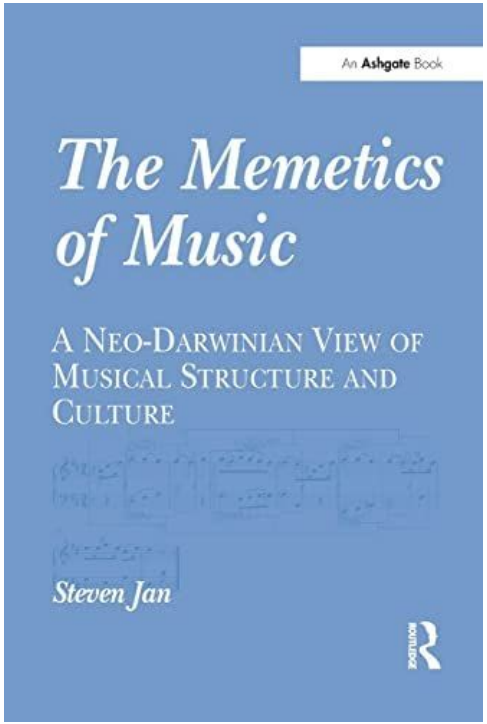


Steven Jan (2007). *The Memetics of Music: A Neo-Darwinian View of Musical Structure and Culture.*

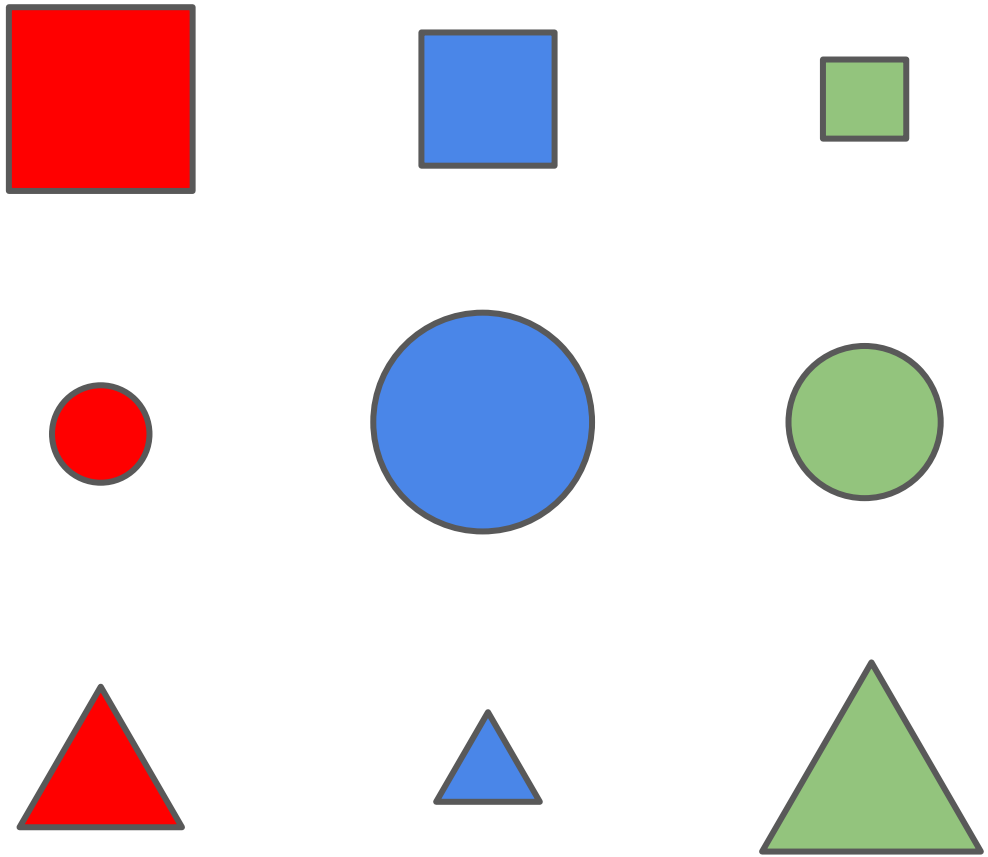
Style as a hierarchy of patterns

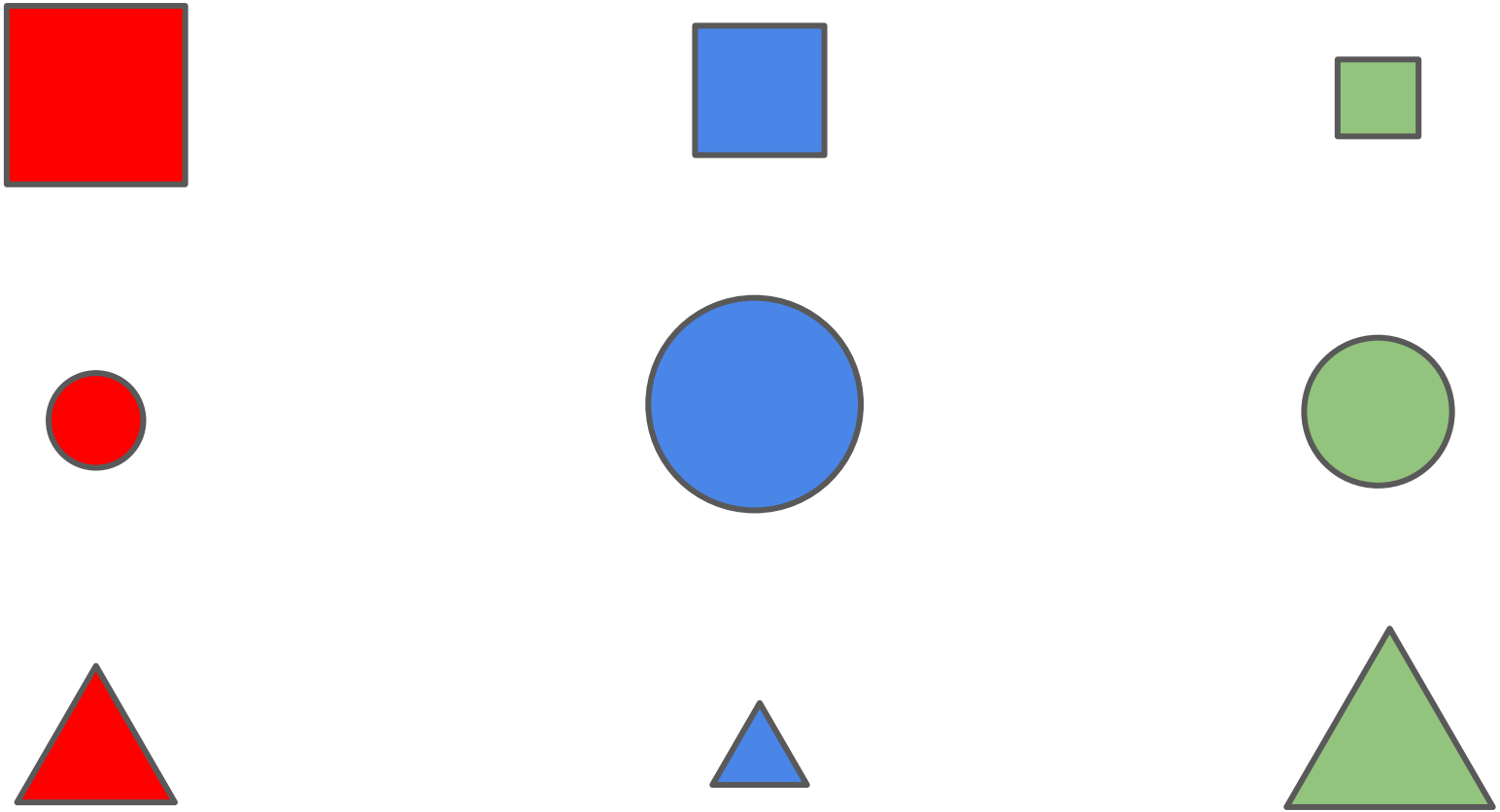
Music Cognition Music Theory Music Stylometry

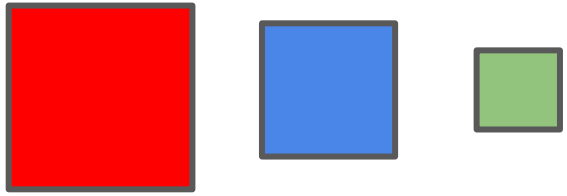


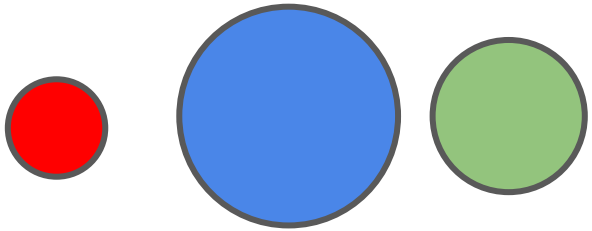


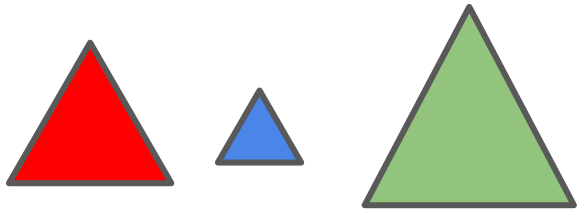
Steven Jan (2007). *The Memetics of Music: A Neo-Darwinian View of Musical Structure and Culture.*

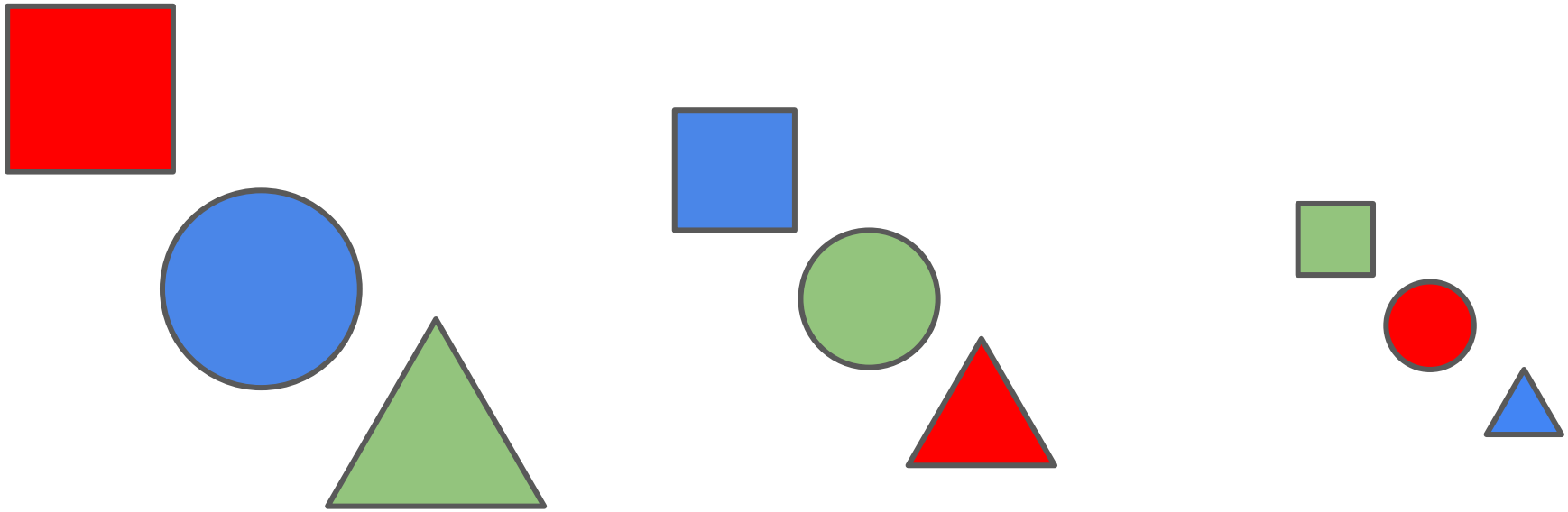












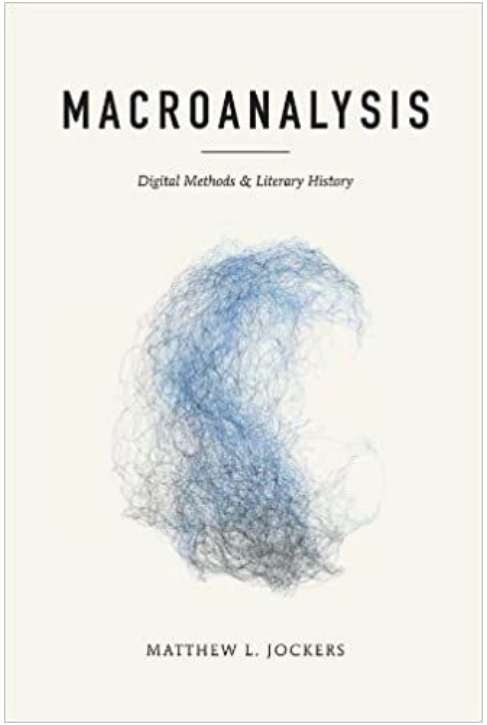
How we study the world depends on our world view!

I work with an office full of sadists

u/NoTick • 9d

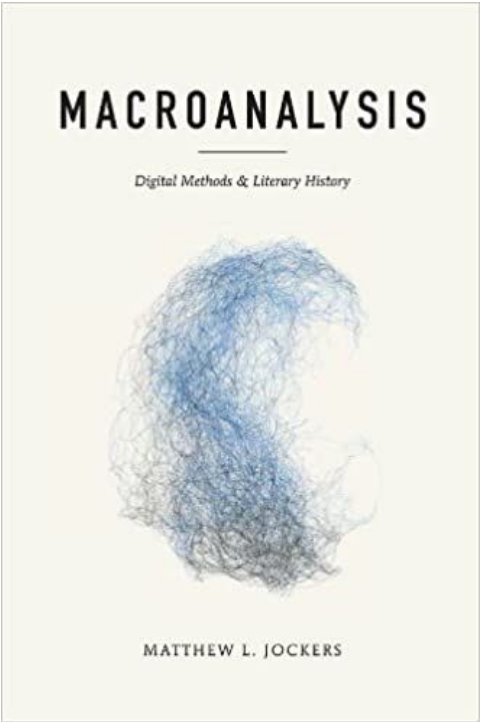
https://www.buzzfeed.com/pablovaldivia/cut-cake-fails

Digital Methods



Jockers, M. L. (2013). *Macroanalysis: Digital Methods and Literary History.* University of Illinois Press.

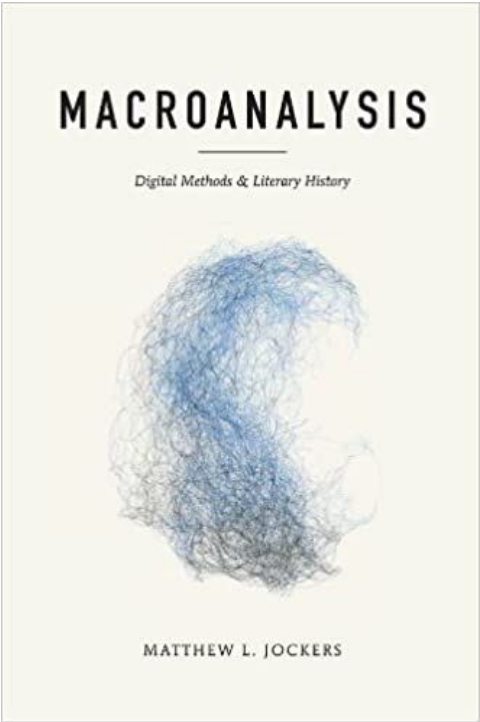
Digital Methods



"**Technology** has certainly changed some things about the way [music scholars] go about their work, but until recently change has been mostly at the level of simple, even anecdotal search."

Jockers, M. L. (2013). *Macroanalysis: Digital Methods and Literary History.* University of Illinois Press.

Digital Methods



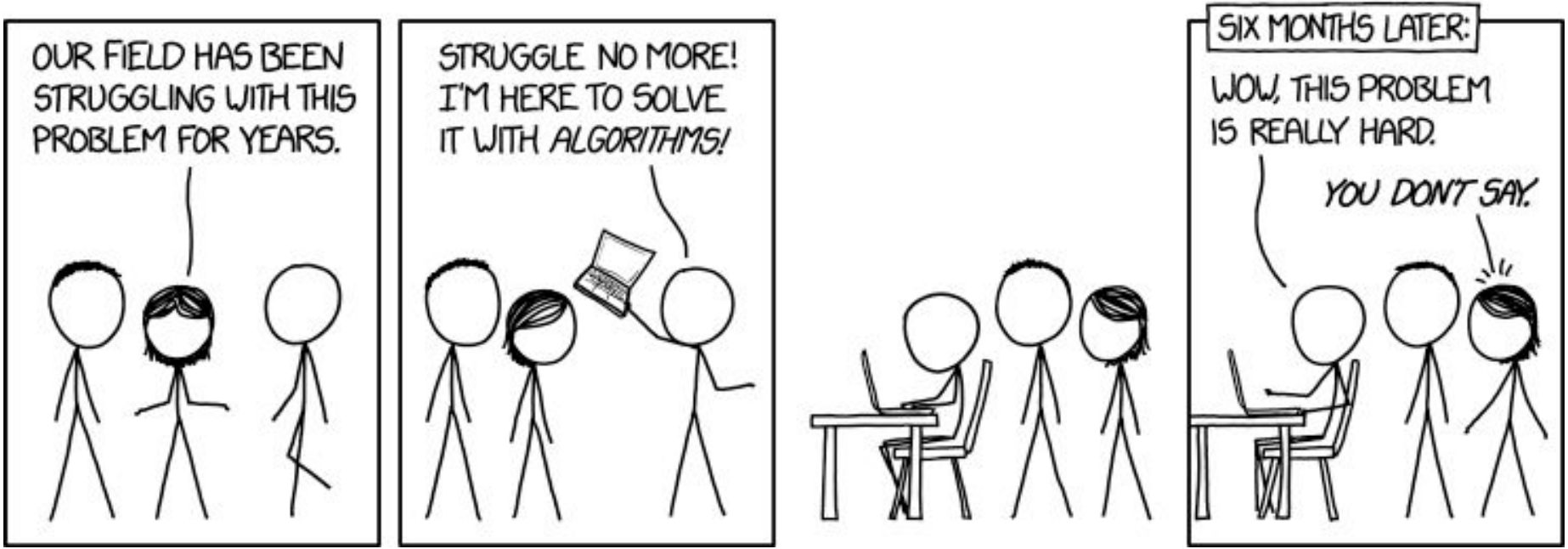
"**Technology** has certainly changed some things about the way [music scholars] go about their work, but until recently change has been mostly at the level of simple, even anecdotal search."

"The **questions** we may now ask were previously inconceivable, and to answer these questions requires a **new methodology**, a **new way of thinking** about our object of study."

Jockers, M. L. (2013). *Macroanalysis: Digital Methods and Literary History.* University of Illinois Press.

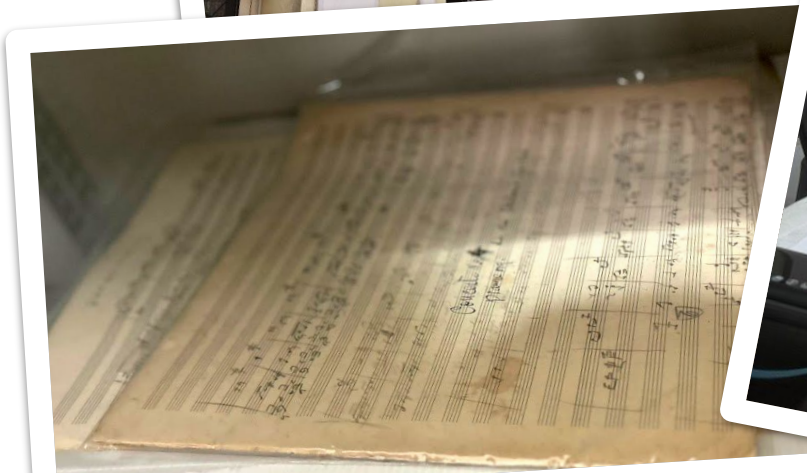
Music analysis and digital methods

Music analysis and digital methods



https://xkcd.com/1831/


38

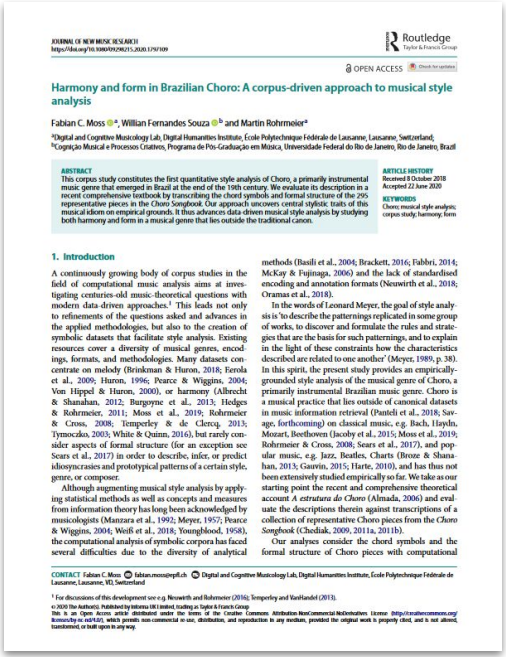


Challenges & Potential

Formats (manuscripts, prints, sketches, recordings…)	Sophisticated database models, symbolic-audio linking
Unknown information (composer, date…)	Machine Learning (statistical inference)
Size	Optical Music Recognition & ML
Musicology	Is a choro a “work”? Notation vs performance vs improvisation
Music Theory	Relation between structural elements
Copyright	Time?
…	…

Corpus Study

How I met your Choro



JOURNAL OF NEW MUSIC RESEARCH
https://doi.org/10.1080/09298215.2020.1747109

Routledge
Taylor & Francis Group

OPEN ACCESS

Harmony and form in Brazilian Choro: A corpus-driven approach to musical style analysis

Fabian C. Moss [a], Willian Fernandes Souza [b] and Martin Rohrmeier [a]

[a]Digital and Cognitive Musicology Lab, Digital Humanities Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; [b]Cognição Musical e Processos Criativos, Programa de Pós-Graduação em Música, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

ABSTRACT

This corpus study constitutes the first quantitative style analysis of Choro, a primarily instrumental music genre that emerged in Brazil at the end of the 19th century. We evaluate its description in a recent comprehensive textbook by transcribing the chord symbols and formal structure of the 295 representative pieces in the *Choro Songbook*. Our approach uncovers central stylistic traits of this musical idiom on empirical grounds. It thus advances data-driven musical style analysis by studying both harmony and form in a musical genre that lies outside the traditional canon.

ARTICLE HISTORY
Received 8 October 2018
Accepted 22 June 2020

KEYWORDS
Choro; musical style analysis; corpus study; harmony; form

1. Introduction

A continuously growing body of corpus studies in the field of computational music analysis aims at investigating centuries-old music-theoretical questions with modern data-driven approaches.[1] This leads not only to refinements of the questions asked and advances in the applied methodologies, but also to the creation of symbolic datasets that facilitate style analysis. Existing resources cover a diversity of musical genres, encodings, formats, and methodologies. Many datasets concentrate on melody (Brinkman & Huron, 2018; Ferola et al., 2009; Huron, 1996; Pearce & Wiggins, 2004; Von Hippel & Huron, 2000), or harmony (Albrecht & Shanahan, 2012; Burgoyne et al., 2013; Hedges & Rohrmeier, 2011; Moss et al., 2019; Rohrmeier & Cross, 2008; Temperley & de Clercq, 2013; Tymoczko, 2003; White & Quinn, 2016), but rarely consider aspects of formal structure (for an exception see Sears et al., 2017) in order to describe, infer, or predict idiosyncrasies and prototypical patterns of a certain style, genre, or composer.

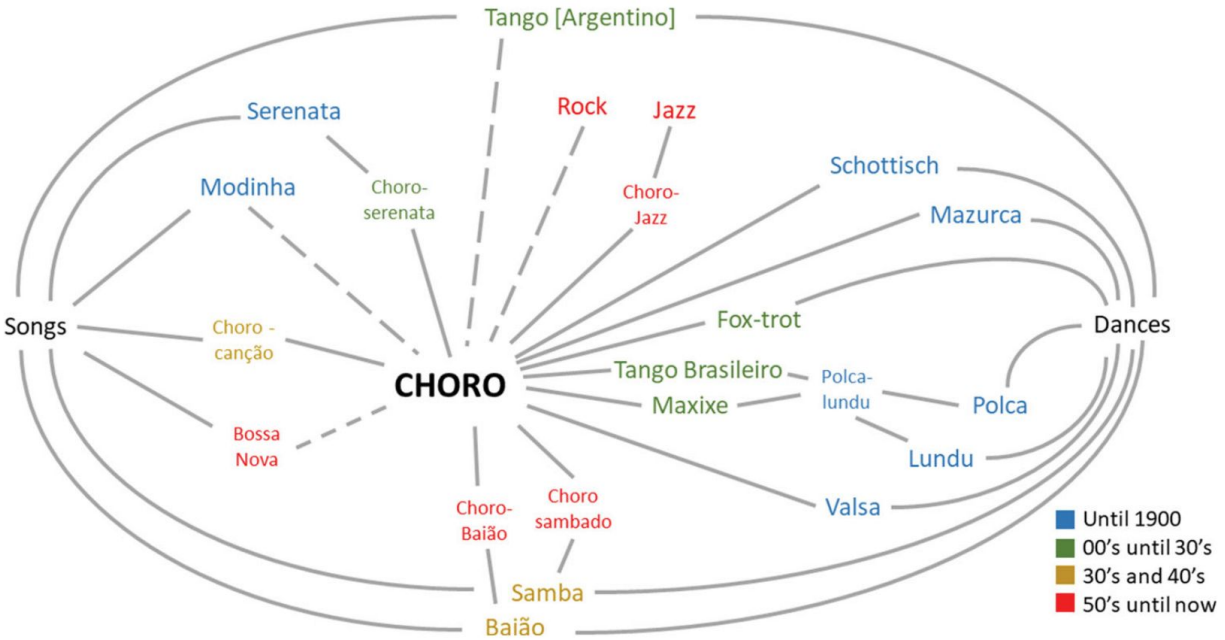
Although augmenting musical style analysis by applying statistical methods as well as concepts and measures from information theory has long been acknowledged by musicologists (Manzara et al., 1992; Meyer, 1957; Pearce & Wiggins, 2004; Weiß et al., 2018; Youngblood, 1958), the computational analysis of symbolic corpora has faced several difficulties due to the diversity of analytical methods (Basili et al., 2004; Brackett, 2016; Fabbri, 2014; McKay & Fujinaga, 2006) and the lack of standardised encoding and annotation formats (Neuwirth et al., 2018; Oramas et al., 2018).

In the words of Leonard Meyer, the goal of style analysis is 'to describe the patternings replicated in some group of works, to discover and formulate the rules and strategies that are the basis for such patternings, and to explain in the light of these constraints how the characteristics described are related to one another' (Meyer, 1989, p. 38). In this spirit, the present study provides an empirically-grounded style analysis of the musical genre of Choro, a primarily instrumental Brazilian music genre. Choro is a musical practice that lies outside of canonical datasets in music information retrieval (Panteli et al., 2018; Savage, forthcoming) on classical music, e.g. Bach, Haydn, Mozart, Beethoven (Jacoby et al., 2015; Moss et al., 2019; Rohrmeier & Cross, 2008; Sears et al., 2017), and popular music, e.g. Jazz, Beatles, Charts (Broze & Shanahan, 2013; Gauvin, 2015; Harte, 2010), and has thus not been extensively studied empirically so far. We take as our starting point the recent and comprehensive theoretical account *A estrutura do Choro* (Almada, 2006) and evaluate the descriptions therein against transcriptions of a collection of representative Choro pieces from the *Choro Songbook* (Chediak, 2009, 2011a, 2011b).

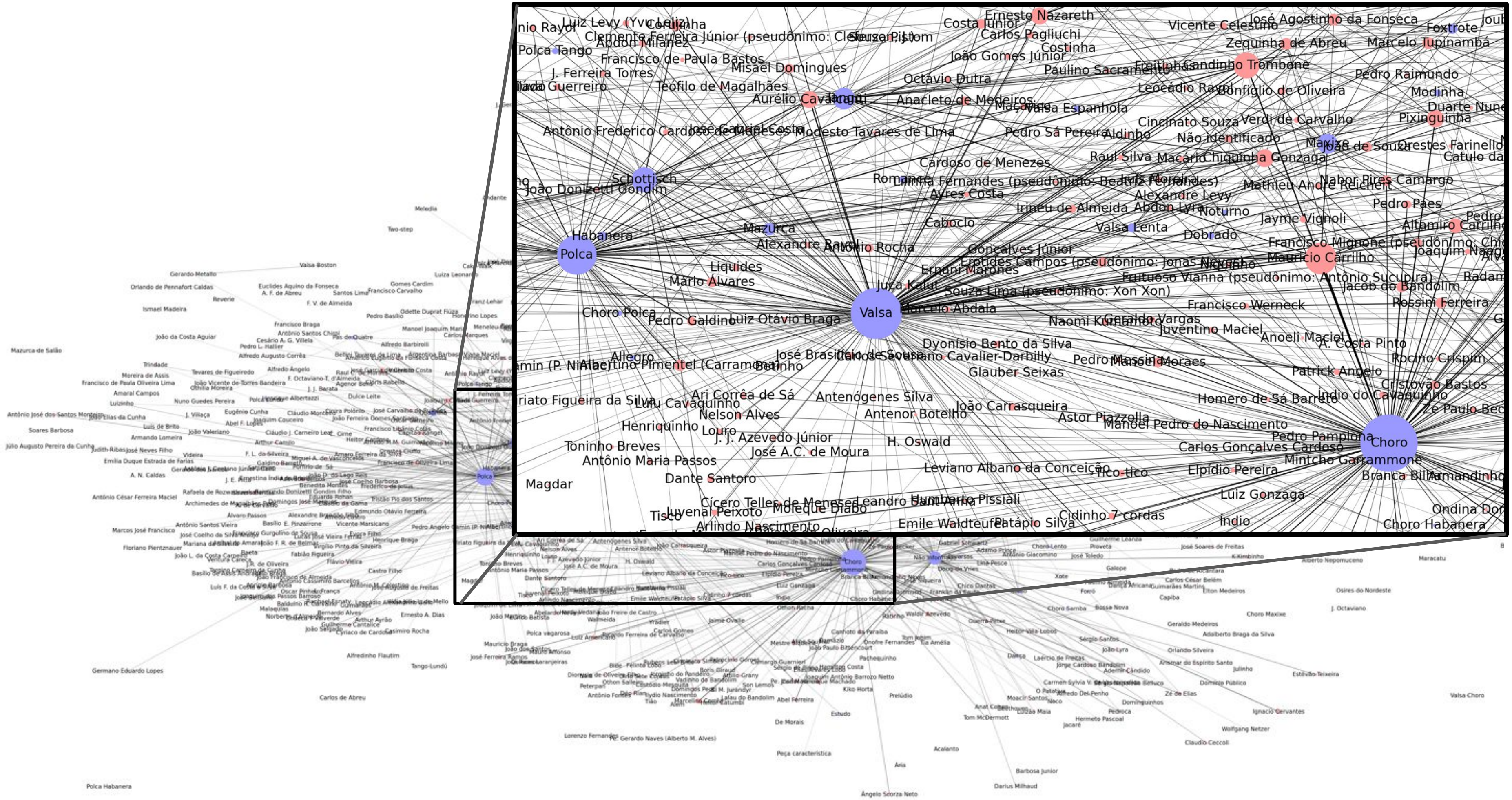
Our analyses consider the chord symbols and the formal structure of Choro pieces with computational

CONTACT Fabian C. Moss fabian.moss@epfl.ch Digital and Cognitive Musicology Lab, Digital Humanities Institute, École Polytechnique Fédérale de Lausanne, Lausanne, VD, Switzerland

[1] For discussions of this development see e.g. Neuwirth and Rohrmeier (2016); Temperley and VanHandel (2013).

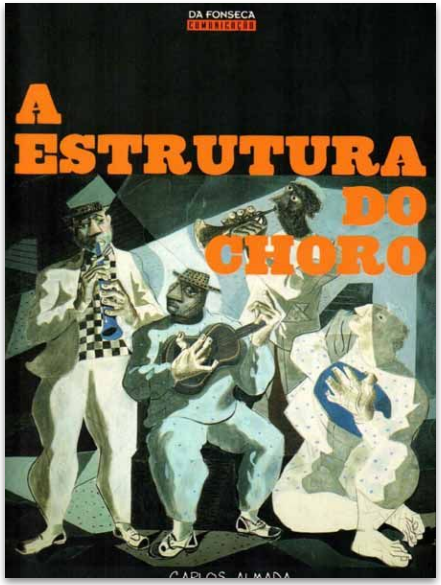


Moss, F. C., Souza, W. F., & Rohrmeier, M. (2020). Harmony and form in Brazilian Choro: A corpus-driven approach to musical style analysis. *Journal of New Music Research*, 49(5), 416–437.

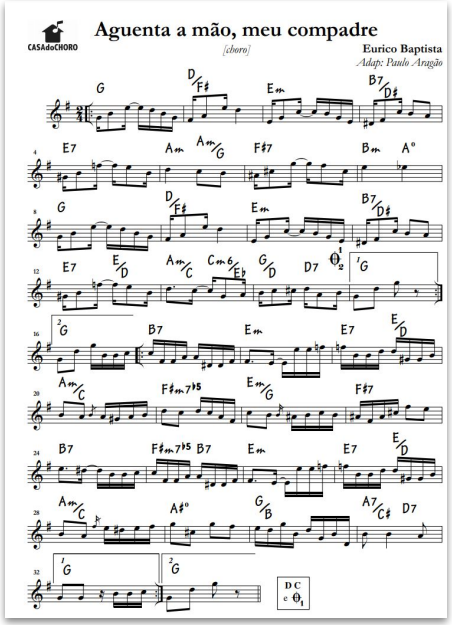


Polca Tango
Francisco de Paula Bastos
Misael Domingues
J. Ferreira Torres
Teófilo de Magalhães
Octávio Dutra
Anacleto de Medeiros
Valsa Espanhola
Ernesto Nazareth
Costa Júnior
Carlos Pagliuchi
Costinha
João Gomes Júnior
Paulino Sacramento
Vicente Celestino
José Agostinho da Fonseca
Foxtrote
Zequinha de Abreu
Marcelo Tupinambá
Pedro Raimundo
Modinha
Pixinguinha
Leocádio Rayol
Bonfiglio de Oliveira
Cincinato Souza
Verdi de Carvalho
Não identificado
Aldinho
Pedro Sá Pereira
Modesto Tavares de Lima
Cardoso de Menezes
Raul Silva
Chiquinha Gonzaga
Maxixe
Schottisch
João Donizetti Gondim
Romance
Ayres Costa
Alexandre Levy
Irineu de Almeida
Noturno
Mathieu André Reichert
Pedro Paes
Jayme Vignoli
Habanera
Polca
Mazurca
Caboclo
Valsa Lenta
Dobrado
Alexandre Rayol
Antônio Rocha
Gonçalves Júnior
Liquides
Mário Álvares
Juca Kalut
Mauricio Carrilho
Jacob do Bandolim
Souza Lima (pseudônimo: Xon Xon)
Frutuoso Vianna (pseudônimo: Antônio Sucupira)
Valsa
Choro Polca
Pedro Galdino
Luiz Otávio Braga
Francisco Werneck
Rossini Ferreira
Juventino Maciel
Anoeli Maciel
A. Costa Pinto
Dyonísio Bento da Silva
Allegro
Glauber Seixas
Patrick Angelo
Índio do Cavaquinho
Cristóvão Bastos
Ari Corrêa de Sá
Antenógenes Silva
Antenor Botelho
João Carrasqueira
Homero de Sá Barreto
Astor Piazzolla
Manoel Pedro do Nascimento
Nelson Alves
Henriquinho
Louro
J. J. Azevedo Júnior
H. Oswald
Toninho Breves
José A.C. de Moura
Pedro Pamplona
Choro
Carlos Gonçalves Cardoso
Mintcho Garrammone
Antônio Maria Passos
Leviano Albano da Conceição
Tico-tico
Elpídio Pereira
Dante Santoro
Magdar
Luiz Gonzaga
Cidinho 7 cordas
Tisco
Índio
Emile Waldteufel
Patápio Silva
Choro Habanera

Research questions and data

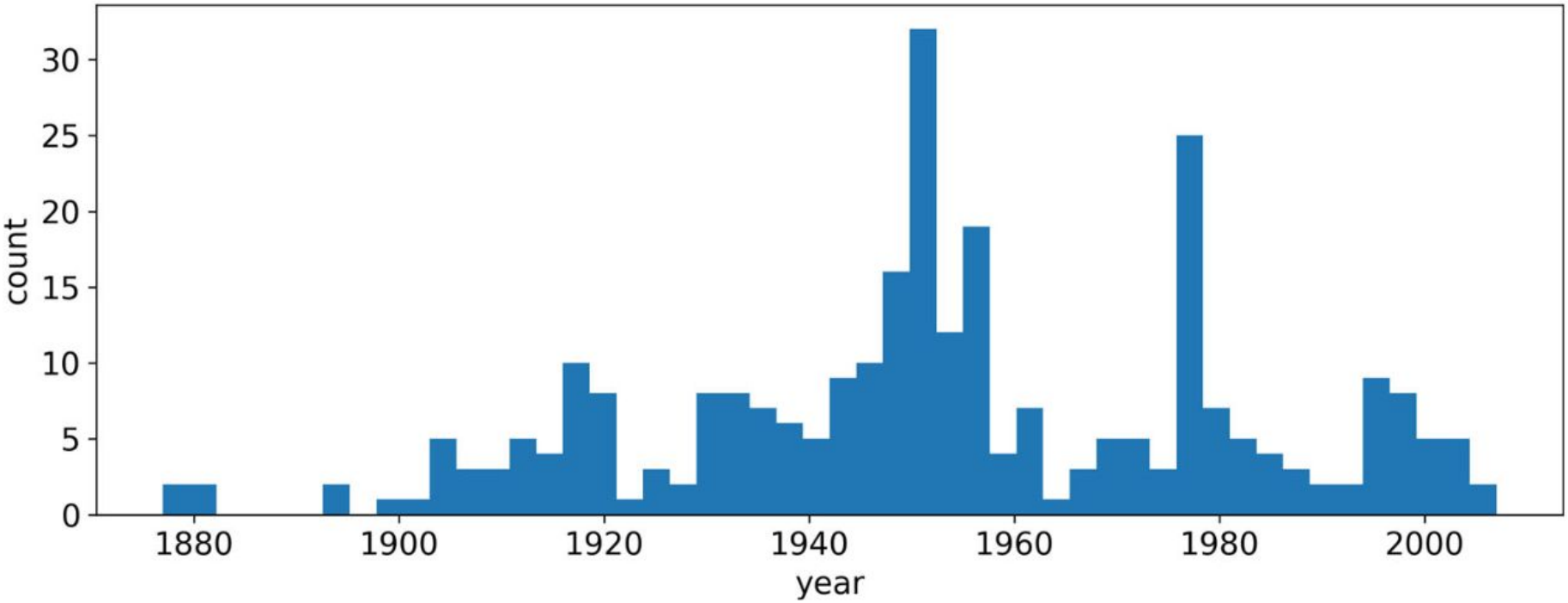


Research questions and “hypotheses”

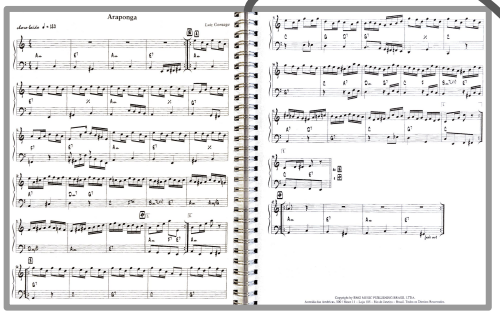


3 Songbooks / 295 Choros / different genres / representative sample?

A representative sample?



Songbooks
3 Volumes
296 Pieces
20,996 chord symbols
723 unique



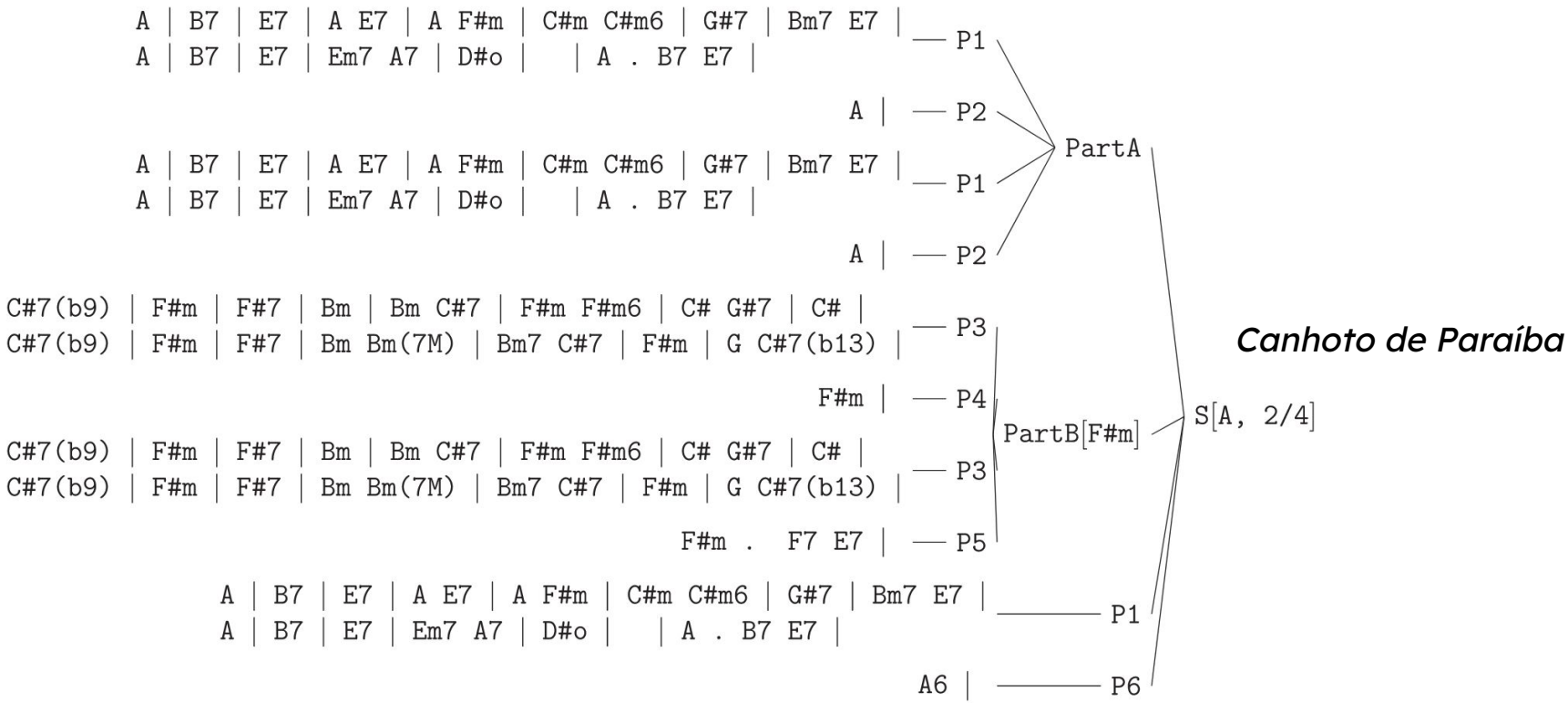
Transcriptions
Chord symbols
Key, Meter

```
P0: Am | E7 | Am | E7 |
P1: Am | E7 | | Am | | G7 | F7 | E7 |
Am | E7 | | A7 | Dm7 G7 | C F7M |
Bm7(b5) E7 | Am Am/G | Dm/F | Am | E7 |
P2: Am . F7  E7 |
P3: Am |

P4: G7 | C | G7 | C | G7 | C | G D7 |
G7 | | C | G7 | C A7 | Dm G7 | C F |
Bm7(b5) E7 | A7 D7 | G7 | C | G7 |
P5: C |
P6: C E7 |

PartA: $P1 $P2 $P1 $P3
PartB[C]: $P4 $P5 $P4 $P6

S[Am, 2/4]: $P0 $PartA $PartB $P1 $P0
```

Form
Hierarchical Representation

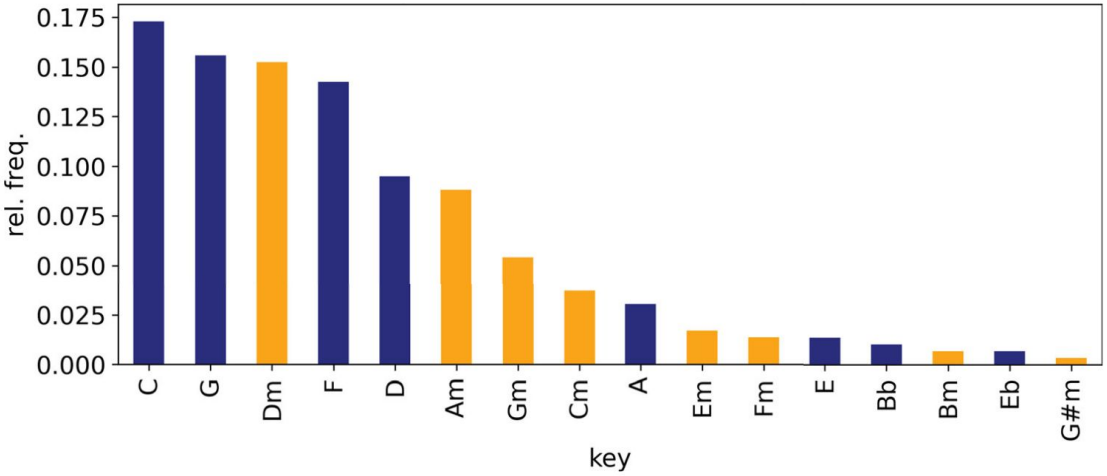

```
{'S': [{'P0': {1: {'Am': 1.0},
    2: {'E7': 1.0},
    3: {'Am': 1.0},
    4: {'E7': 1.0}}},
  {'PartA': [{'P1': {5: {'Am': 1.0},
      6: {'E7': 1.0},
      7: {'Am': 1.0},
      8: {'G7': 1.0},
      9: {'F7': 1.0},
      10: {'E7': 1.0},
      11: {'Am': 1.0},
      12: {'E7': 1.0},
      13: {'A7': 1.0},
      14: {'Dm7': 0.5, 'G7': 0.5},
      15: {'Dm7': 0.5, 'G7': 0.5},
      16: {'C': 0.5, 'F7M': 0.5},
      17: {'C': 0.5, 'F7M': 0.5},
      18: {'Bm7(b5)': 0.5, 'E7': 0.5},
      19: {'Bm7(b5)': 0.5, 'E7': 0.5},
      20: {'Am': 0.5, 'Am/G': 0.5},
      21: {'Am': 0.5, 'Am/G': 0.5},
      22: {'Dm/F': 1.0},
      23: {'Am': 1.0},
      24: {'E7': 1.0}}},
    {'P2': {25: {'.': 0.25, 'Am': 0.25, 'E7': 0.25, 'F7': 0.25},
      26: {'.': 0.25, 'Am': 0.25, 'E7': 0.25, 'F7': 0.25},
      27: {'.': 0.25, 'Am': 0.25, 'E7': 0.25, 'F7': 0.25},
      28: {'.': 0.25, 'Am': 0.25, 'E7': 0.25, 'F7': 0.25}}},
```

Recursive encoding of harmony and form

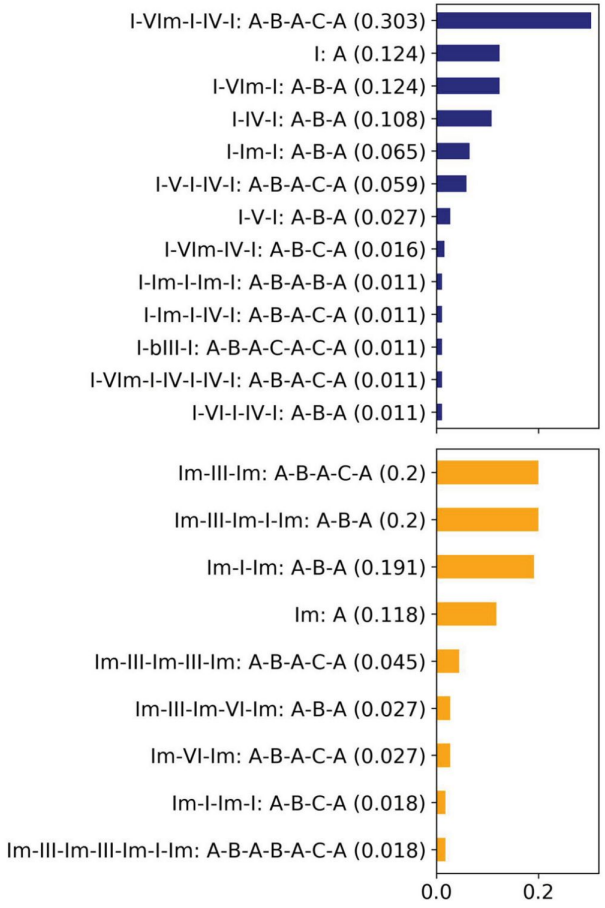


Canhoto de Paraíba

Counting



Parts	A	B	A	C	A
Major 1	I	V	I	IV	I
Major 2	I	VIm	I	IV	I
Minor	Im	III	Im	I	Im



Qualitative comparison

theoretical

Table 2. Theoretical prototypical harmonic progressions (Almada, 2006, pp. 20–23).

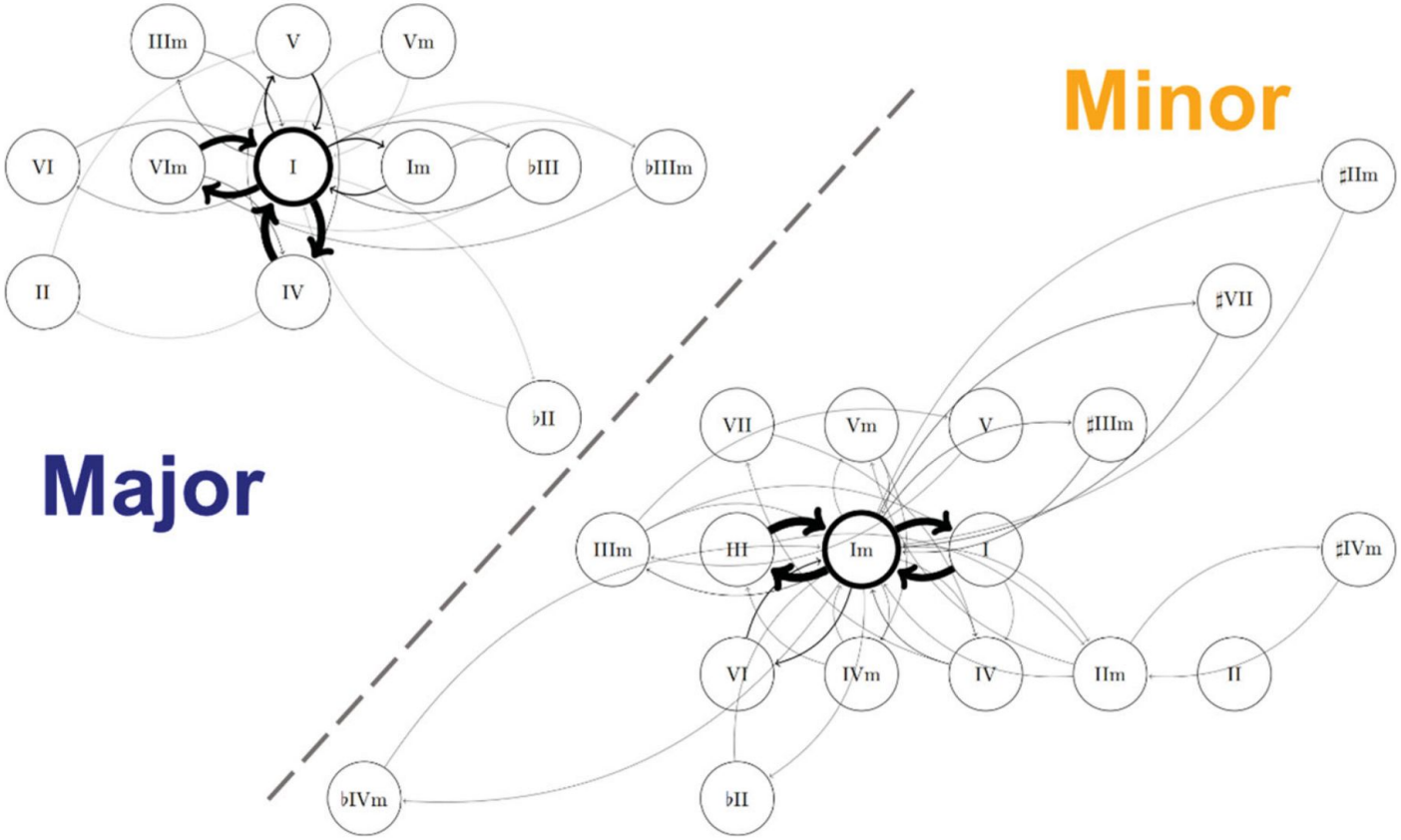
Major	1	2	3	4	5	6	7	8
A	I-V°/II	II	II-V	I	IV	I	V/V	V
B	V	I-V/II	V/V-V	I	V/VI	VI	V/V	V
C	I6-♭III°	II	V	I	V/VI	VI	V/III	III-V
D	I	V	VI	V/VI	V/VI	V/II	V/V	V
E	I-V	I	VI-V/VI	VI	V°/III	I6	V/V	V
F	I-V/VI	VI-V/II	II-V	I	V/III	III	V/III	III-V
	9	10	11	12	13	14	15	16
A	I-V°/II	II	V/VI	VI	V°/III	I6-V/II	II-V	I
B	V	I-V/II	V/IV	IV	IV	I	V/V-V	I
C	I6-♭III°	II	V	I	V/II	V/V	V	I
D	I	V	VI	V/VI	II-IVm	I6-♭III°	II-V	I
E	I-V/VI	V/II	II-V/II	II	IV-V°/V	I4-V/II	II-V	I
F	I-V/VI	VI-V/II	II-V	V/II	IV-IVm	I-V/II	V/V-V	I

empirical

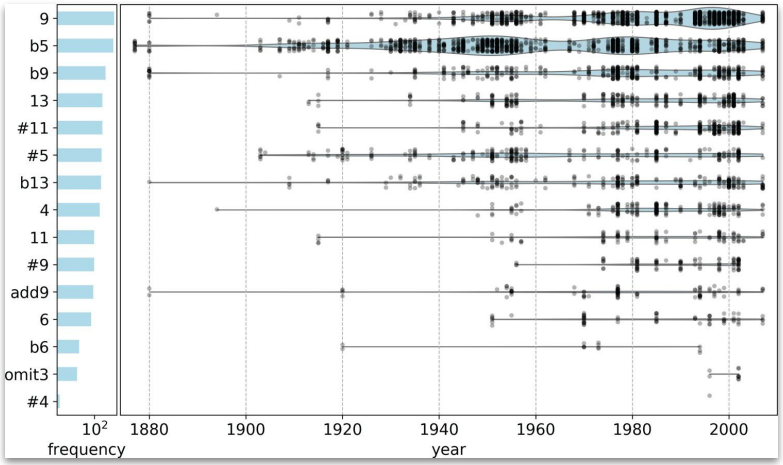
Table 3. Top 5 empirical most likely 16-bar phrases.

Major	1	2	3	4	5	6	7	8
1	V7	I	V7	I	V7	I	V7	I
2	I	IIm	V7	I	IIm	V7	I	I
3	I	IIm7-V7	IV7	III7	II7-V7	I	V-II7	V7
4	I	IIm7-V7	I/3-bIIIo	IIm7-V7	I	V-IIIm7	II7/5-II7	V7
5	V7	I	III7/3#	VIm	IIm/3	I	V7	I
	9	10	11	12	13	14	15	16
1	V7	I	V7	I	V7	I	V7	I
2	I	IIm	V7	I	IIm	V7	I	I
3	I	IIm7-V7	IV7-IV7/3	VI/7	IIm/3b-#IVo	I/5	IIm7-V7	I
4	I-VIm7	IIm7-V7	I/3-bIIIo	IIm7-V7	VI7	IIm-IVm	IIm7-V7	I-III7
5	V7	I	III7/3#	VIm	IIm/3	I	V7	I-V7-III7

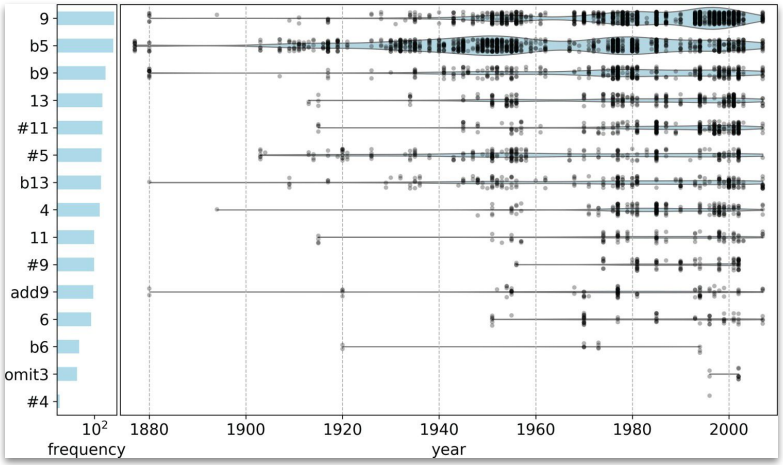
Modulation plans

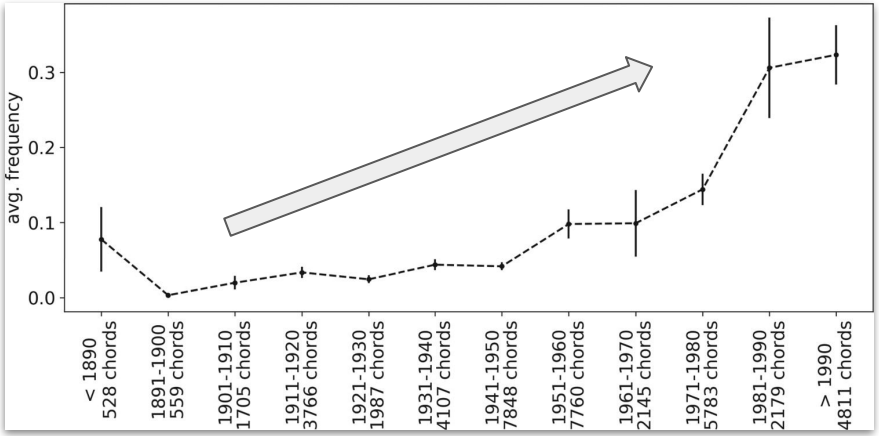


Diachronic chromatization



Diachronic chromatization

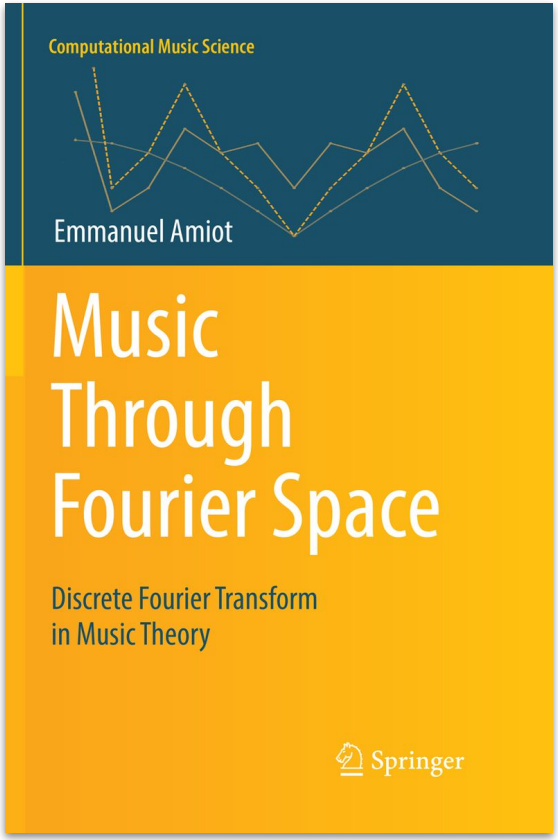


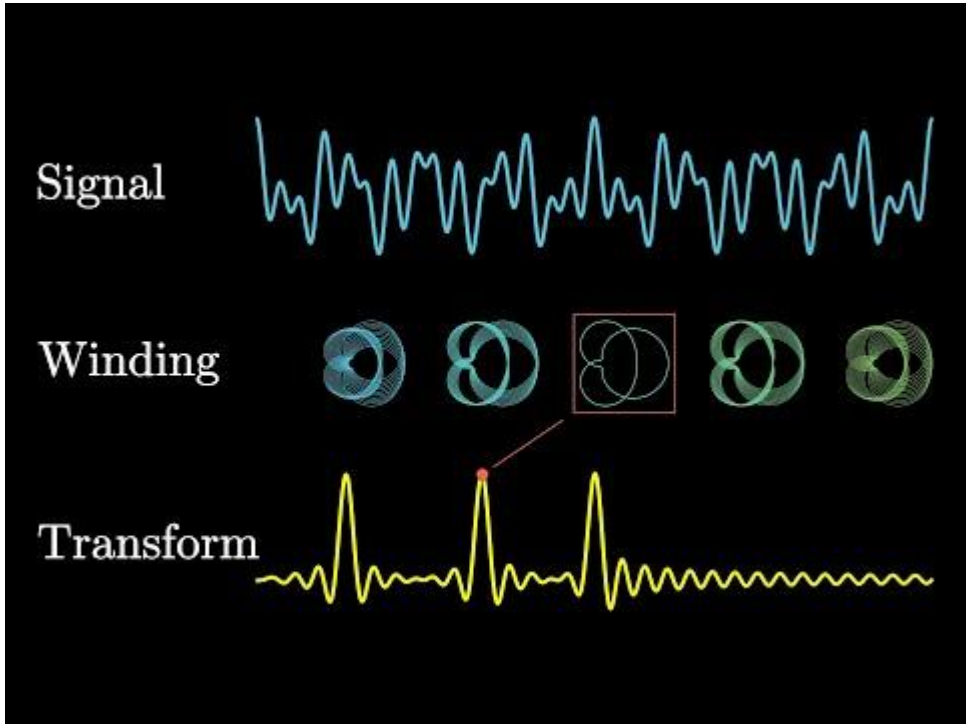


But: only what is notated, *not* what was played → include recordings

Music Analysis and the DFT

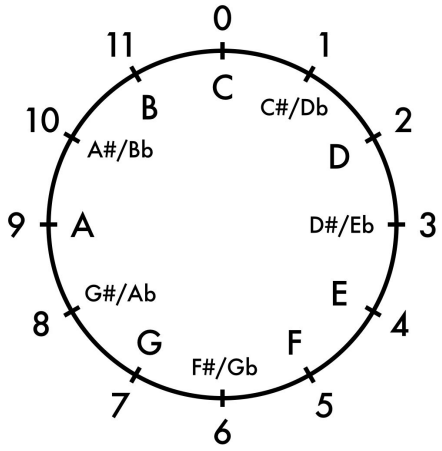
Music Theory and the Discrete Fourier Transform

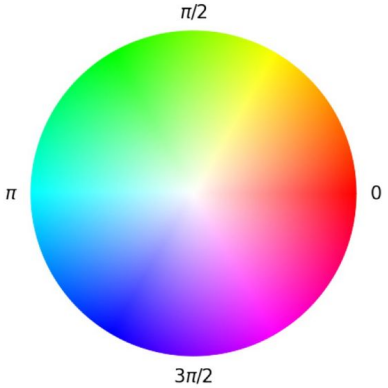




https://www.youtube.com/watch?v=spUNpyF58BY

Discrete Fourier Transform (DFT)





Def.: A *pitch-class vector* $x = P(s, w)$ contains counts all pitch classes for a segment starting at s with width w.

Definition 3.1 (Discrete Fourier Transform). The *discrete Fourier transform* (DFT) of any pitch-class vector x (i.e., any choice of $x = P(s, w)$) corresponds to the mapping

$$F\colon \mathbb{R}^{12}_{\geq 0} \to \mathbb{C}^{12}, \qquad F(x)[k] = \sum_{j=0}^{11} x[j] e^{i2\pi j \frac{k}{12}},$$

and $F(x)[k]$ is called the k-th *Fourier coefficient* of x.

Def.: Dividing a *pitch-class vector* by its sum returns a pitch-class distribution:

$$\tilde{P}(s, a) = P(s, a) / \textstyle\sum_j P(s, a)[j]$$

Visualize Fourier mapping of pitch-class vectors

Jennifer Harding: Jenn's Visual Pitch Class Vector Calculator

http://www.jenndharding.com/vectorCalculator

Keyscapes

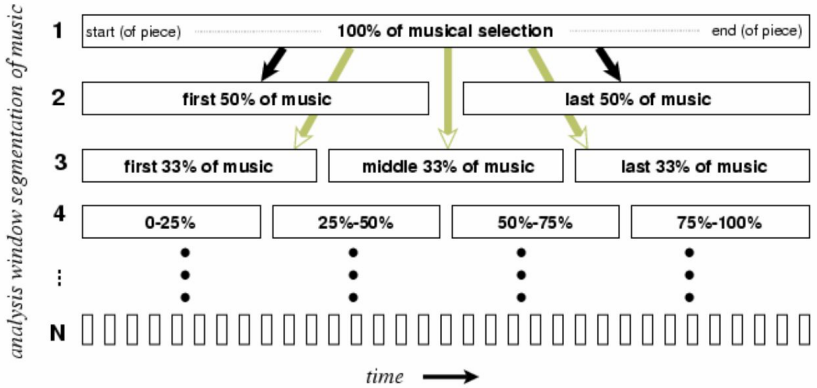


Figure 1: Type 1 analysis window configuration.

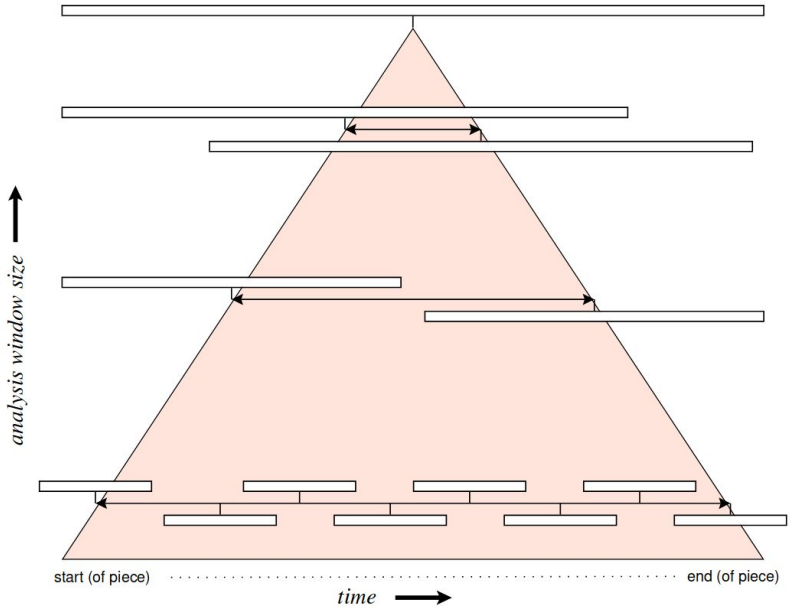


Figure 4: Type 2 analysis window arrangement.

Sapp, C. S. (2001). Harmonic Visualizations of Tonal Music. *International Computer Music Conference Proceedings*, 1–8. http://hdl.handle.net/2027/spo.bbp2372.2001.029

Wavescapes =

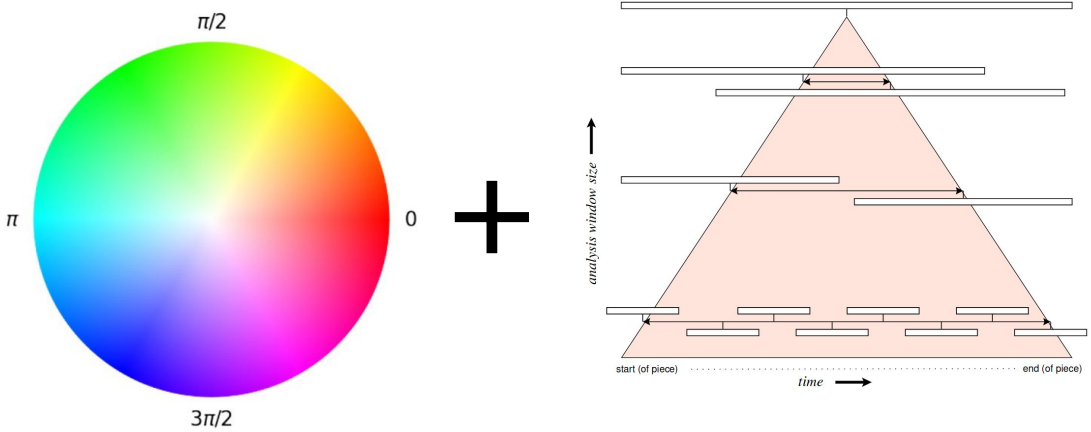


Figure 4: Type 2 analysis window arrangement.

Viaccoz, C., Harasim, D., Moss, F. C., & Rohrmeier, M. (2022). Wavescapes: A visual hierarchical analysis of tonality using the discrete Fourier transform. *Musicae Scientiae*, 10298649211034906. https://doi.org/10.1177/10298649211034906

Article

MUSICÆ SCIENTIÆ

Musicae Scientiae
1–38

DOI: 10.1177/10298649211034906
journals.sagepub.com/home/msx
SAGE

Wavescapes: A visual hierarchical analysis of tonality using the discrete Fourier transform

Cédric Viaccoz, Daniel Harasim, Fabian C. Moss and Martin Rohrmeier
Digital and Cognitive Musicology Lab, École Polytechnique Fédérale de Lausanne, Switzerland

Abstract
Many structural aspects of music, such as tonality, can be expressed using hierarchical representations. In music analysis, so-called keyscapes can be used to map a key estimate (e.g., C major, F minor) to each subsection of a piece of music, thus providing an intuitive visual representation of its tonality, in particular of the hierarchical organization of local and global keys. However, that approach is limited in that the mapping relies on assumptions that are specific to common-practice tonality, such as the existence of 24 major and minor keys. This limitation can be circumvented by applying the discrete Fourier transform (DFT) to the tonal space. The DFT does not rely on style-specific theoretical assumptions but only presupposes an encoding of the music as pitch classes in 12-tone equal temperament. We introduce *wavescapes*, a novel visualization method for tonal hierarchies that combines the visual representation of keyscapes with music analysis based on the DFT. Since wavescapes produce visual analyses deterministically, a number of potential subjective biases are removed. By concentrating on one or more Fourier coefficients, the role of the analyst is thus focused on the interpretation and contextualization of the results. We illustrate the usefulness of this method for computational music theory by analyzing eight compositions from different historical epochs and composers (Josquin, Bach, Liszt, Chopin, Scriabin, Webern, Coltrane, Ligeti) in terms of the phase and magnitude of several Fourier coefficients. We also provide a Python library that allows such visualizations to be easily generated for any piece of music for which a symbolic score or audio recording is available.

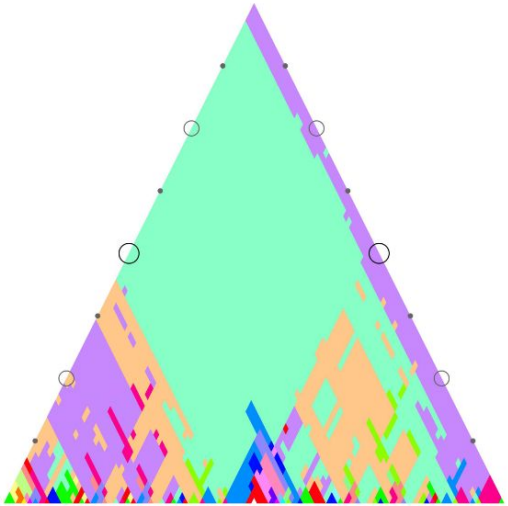
Keywords
Discrete Fourier transform, music analysis, keyscapes, tonal hierarchy, visualization

Many domains of human cognition, such as music, language and action planning, exhibit hierarchical structure (Arbib, 2013; Rebuschat et al., 2012). In the case of music, several structural features are organized hierarchically, for instance formal arrangement, rhythm, melody

Corresponding author:
Daniel Harasim, Digital and Cognitive Musicology Lab, École Polytechnique Fédérale de Lausanne, Lausanne, CH-1015, Switzerland.
Email: daniel.harasim@epfl.ch

Hierarchical tonal structures / historical changes

(A) Keyscape of Bach's Prelude in C major (BWV 846)

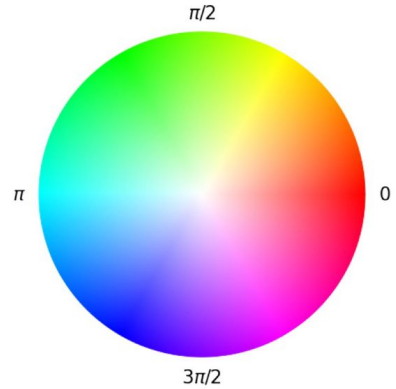


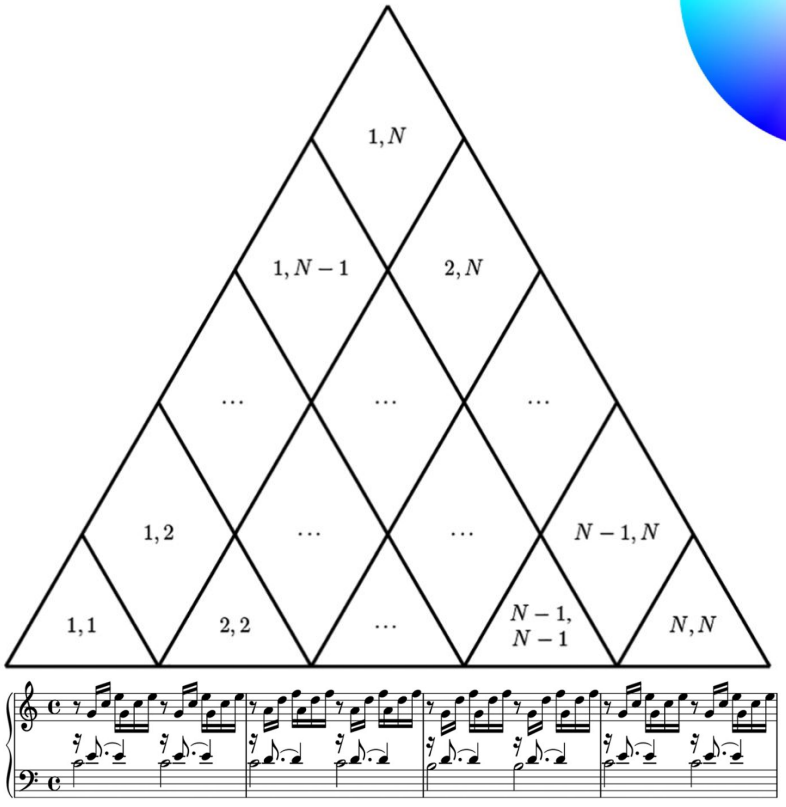
(B) Keyscape of the first 23 measures from the first movement of Liszt's Faust Symphony (S 108)

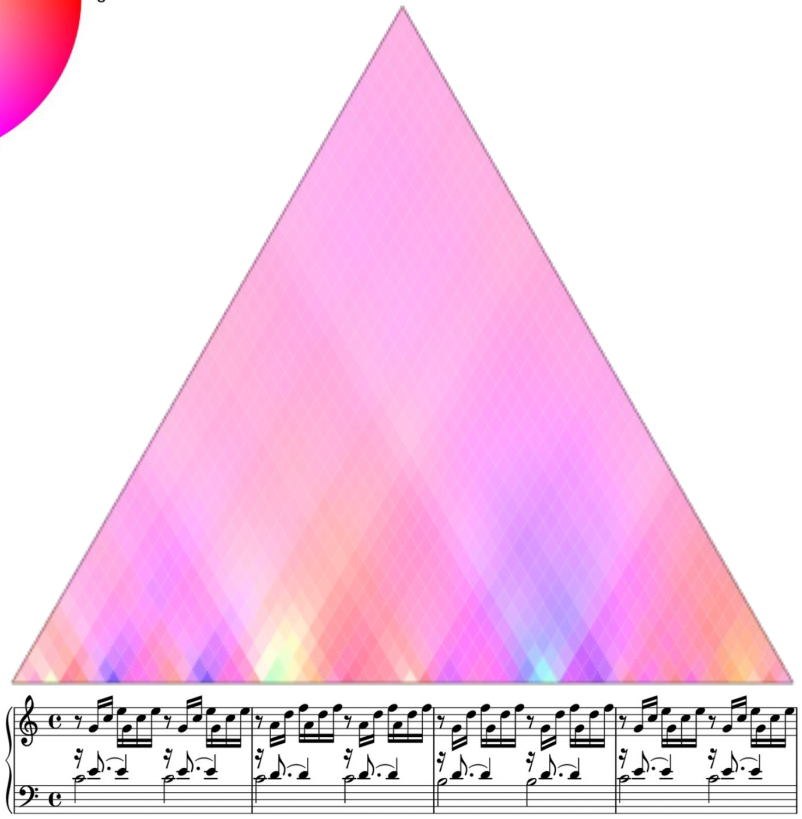
(C) Minor (top row) and major (bottom row) scales categorized into different colors

Viaccoz, C. (2020). *Visual Hierarchical Analysis of Tonality using the Discrete Fourier Transform* [Masters Thesis]. École Polytechnique Fédérale de Lausanne.

Wavescapes



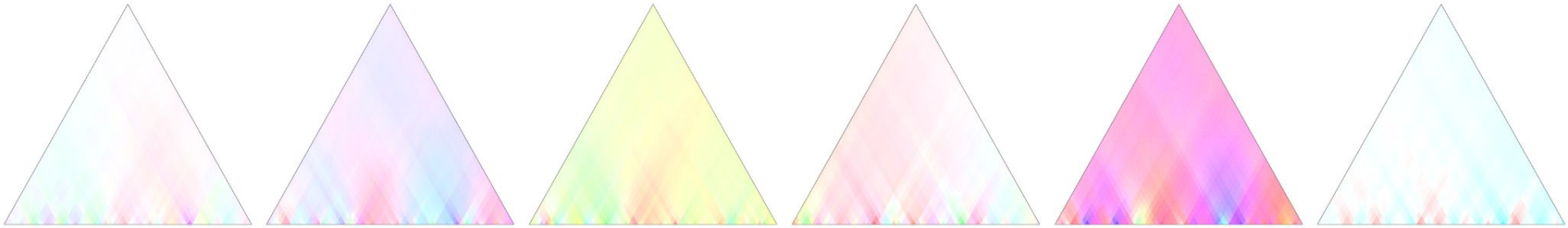




Wavescapes

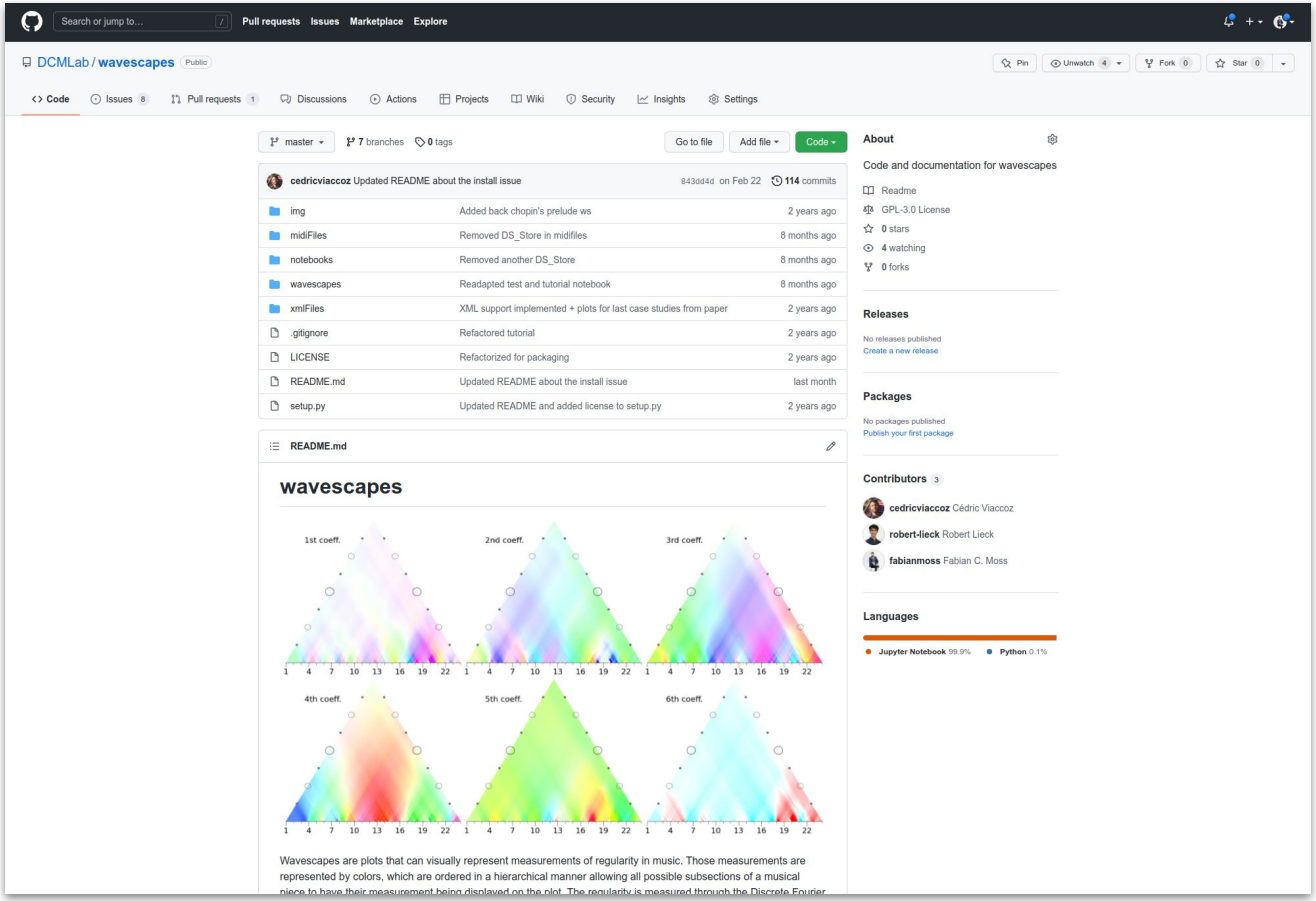
⇒ Some Fourier coefficients have higher activity (= brighter colors)

⇒ Some Fourier coefficients (may) show clear patterns



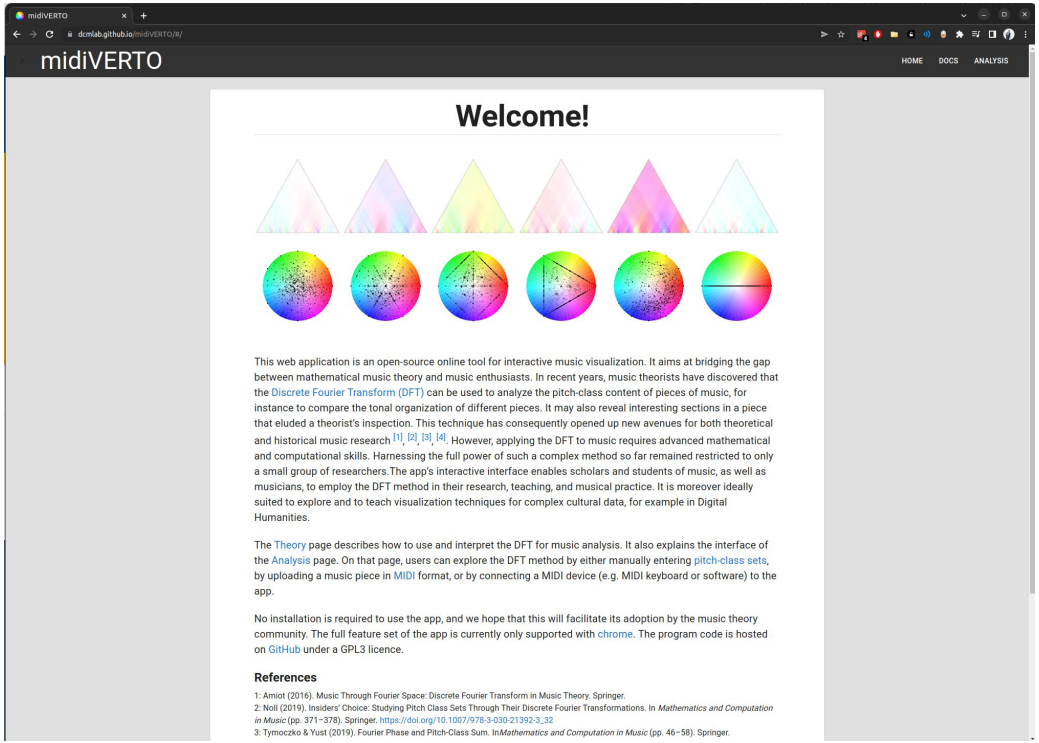
chromaticity | dyadicity | triadicity | tetradicity | diatonicity | whole-tone scale

Wavescapes



https://github.com/DCMLab/wavescapes

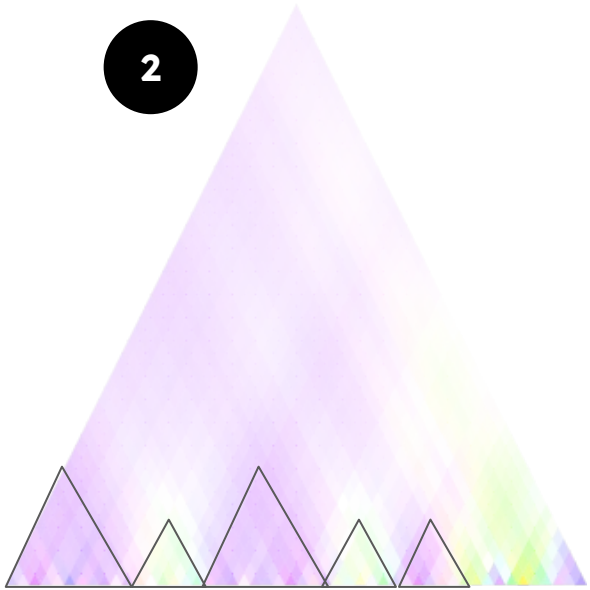
Enabling interactive music visualization for a wider community



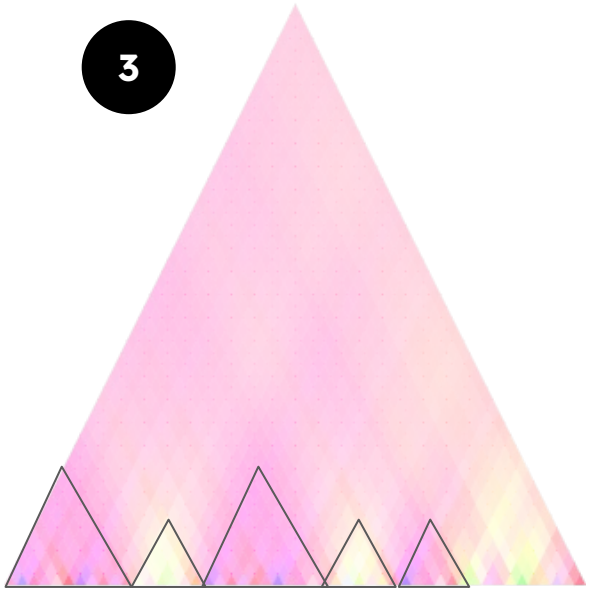
https://dcmlab.github.io/midiVERTO/

dhcenter
UNIL·EPFL

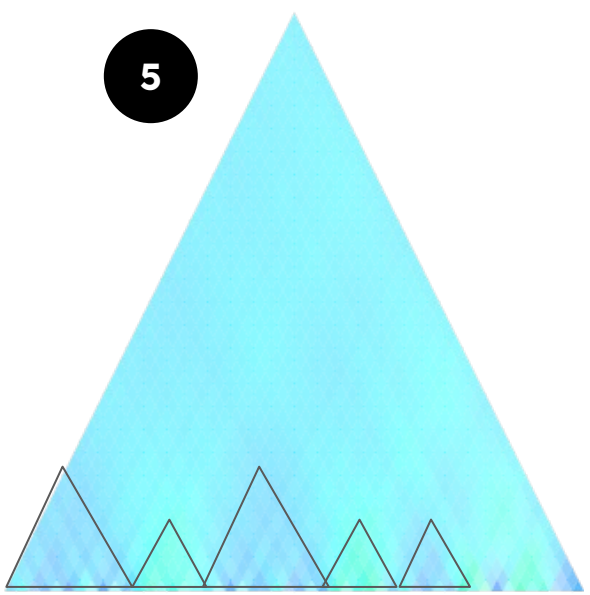
Formal Analysis using DFT: E. Nazaret, *Odeon* (1909)



“dyadicity”



“triadicity”



“diatonicity”

Let’s try it out

- Download your favorite song as a MIDI file (Google it)
- Got to https://dcmlab.github.io/midiVERTO
- Upload the file and wait until the wavescapes have loaded
- Try to understand the plots, find interesting areas
- Modify the parameters and see how the visualizations change
- Which aspects of the piece’s tonality can you find in the wavescapes or phantom curves?

Many open questions

- How to deal with diversity of sources?
- How to include sheet music and performance recordings?
- How to deal with {melody, rhythm, harmony, …} together?
- (How) can we relate Choro to historical precursors?
- …
- → How can technology and digital methods help us to study Choro?

Thank you for your attention!

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Fabian C. Moss (Würzburg, Germany)
2 April 2024 – Universidade Federal do Rio de Janeiro

Bavarian
Research Alliance