Hilbertian Statistical Models in Music Neuroscience

Dissertation submitted to obtain the titles of Doctor of Art Science from Ghent University and Doctor from the University of Granada within the PhD program in Mathematical and Applied Statistics

MARC VIDAL

SUPERVISORS
Prof. Ana M. Aguilera
Prof. Marc Leman

Ghent University · University of Granada, 2024
Institute for Psychoacoustics and Electronic Music
Department of Statistics and Operations Research | Institute of Mathematics

GHENT UNIVERSITY

UNIVERSIDAD DE GRANADA

Editor: Universidad de Granada. Tesis Doctorales
Autor: Marc Vidal Badía
ISBN: 978-84-1195-427-3
URI: https://digibug.ugr.es/handle/10481/94820

Show yourself more human than critical
and your pleasure will grow
– Domenico Scarlatti

Abstract (EN)

This dissertation addresses the analysis of data emerging in the field of music neuroscience, specifically data collected from neurophysiological monitoring techniques that can be modeled as random objects in spaces of smooth functions. Spaces equipped with a Hilbert structure offer a versatile and elegant framework for the generalization of various statistical techniques, ensuring adaptability and robustness in analyzing complex data structures. Within the context of functional data analysis, these spaces serve as essential tools for understanding and interpreting dynamic data trends over continuous domains. Given the relevance of independent component analysis (ICA) in neuroscience research, our investigation is directed towards its functional counterpart, a technique whose potential still remains relatively overlooked. Functional ICA can be considered a refinement of functional principal component analysis, aimed at identifying low-dimensional structures "as independent as possible" by exploiting the underlying topological features of the data. We provide a comprehensive account of the theoretical foundations of functional ICA and extend the method to Sobolev spaces of smoother functions. Some theoretical properties regarding functional data classification are also presented. Additionally, we develop a repertoire of related functional data techniques tailored for pre-processing and analyzing data in the emerging field of embodied music neuroscience, which investigates the neurological basis of how the body influences musical experience. Two methods based on nonlinear wavelet and polynomial approximations are developed for pre-processing artifactual activity in EEG signals and pupillometry. These methods yield excellent outcomes for neuromotor research, particularly considering the suboptimal condition of the recorded data due to locomotor activity. We further introduce a set of neural descriptors derived from data collected through the aforementioned non-invasive methods, aiming to uncover brain behavior during embodied musical interactions. More specifically, we focus on methodologies for modeling neurotransmitter activity, a critical aspect shown to be essential in shaping motor functionality and other proprioceptive sensations. Our experimental research is portrayed by the concept of emotion transferred into a neurological domain, providing a unique framework to define and capture the neural essence of embodiment in music.

Abstract (NL)

Deze thesis behandelt de analyse van data die ontstaan op het gebied van muziekneurowetenschap, in het bijzonder data verzameld met neurofysiologische meettechnieken die gemodelleerd kunnen worden als willekeurige objecten in ruimtes van gladde functies. Ruimtes uitgerust met een Hilbert-structuur bieden een veelzijdig en elegant kader voor de veralgemening van verschillende statistische technieken en garanderen zo het aanpassingsvermogen en de robuustheid bij de analyse van complexe datastructuren. Binnen de context van functionele data-analyse dienen deze ruimtes als essentiële instrumenten voor het begrijpen en interpreteren van dynamische datatrends over continue domeinen. Gezien de relevantie van onafhankelijke componentenanalyse (ICA) in neurowetenschappelijke studie, is ons onderzoek gericht op de functionele tegenhanger ervan, een techniek waarvan het potentieel nog steeds enigszins over het hoofd wordt gezien. Functionele ICA kan worden beschouwd als een verfijning van functionele principale componentenanalyse, gericht op de identificatie van "zo onafhankelijk mogelijk" laagdimensionale structuren door gebruik te maken van de onderliggende topologische kenmerken van de data. We geven een uitgebreide beschrijving van de theoretische grondslagen van functionele ICA en breiden de methode uit tot Sobolev-ruimtes van gladdere functies. Enkele theoretische eigenschappen met betrekking tot functionele dataclassificatie worden ook voorgesteld. Bovendien ontwikkelden we een repertoire van verwante functionele datatechnieken op maat voor het voorbewerken en analyseren van data in het opkomende gebied van de belichaamde muziekneurowetenschap, die de neurologische basis onderzoekt van hoe het lichaam muzikale ervaringen beïnvloedt. Er werden twee methodes gebaseerd op niet-lineaire wavelet- en polynomiale benaderingen ontwikkeld voor het voorbewerken van artefactuele activiteit in EEG-signalen en pupillometrie. Deze methodes leveren uitstekende resultaten op voor neuromotorisch onderzoek, vooral gezien de suboptimale conditie van de geregistreerde data als gevolg van bewegingsactiviteit. Verder introduceren we een reeks neurale descriptoren afgeleid van data die zijn verzameld met de eerdergenoemde niet-invasieve methodes, met als doel het gedrag van de hersenen tijdens belichaamde muzikale interacties bloot te leggen. Meer specifiek richten we ons op methodologieën voor het modelleren van neurotransmitteractiviteit, een kritisch aspect dat essentieel is bij de vormgeving van motorische interacties en andere proprioceptieve sensaties. Ons experimenteel onderzoek is gebaseerd op het concept van emotie in een neurologisch domein, wat een uniek kader biedt om de neurale essentie van belichaamdheid in muziek te definiëren en vast te leggen.

Abstract (ES)

En esta tesis se aborda el análisis de datos emergentes en el campo de la neurociencia de la música, más concretamente de datos grabados mediante técnicas de monitoreo neurofisiológico que pueden ser modelados como objetos aleatorios en espacios de funciones suaves. Los espacios equipados con estructura de Hilbert ofrecen un marco versátil y elegante para la generalización de un ámplio abanico de técnicas estadísticas, asegurando adaptabilidad y robustez en el análisis de estructuras de datos complejas. En el contexto del análisis de datos funcionales, estos espacios sirven como herramientas esenciales para comprender e interpretar tendencias dinámicas de datos sobre dominios continuos. Dada la relevancia del análisis en componentes independientes (ICA) para el análisis de datos neurocientíficos, nuestra investigación se dirige hacia su versión funcional, una técnica cuyo potencial aún permanece relativamente poco explorado. El ICA funcional puede considerarse una extensión del análisis en componentes principales funcional, orientado a identificar componentes "lo más independientes posible" mediante la explotación de las características topológicas subyacentes de los datos. Se proporciona un análisis exhaustivo de los fundamentos teóricos del ICA funcional y se extiende el método a espacios de Sobolev de funciones más suaves. También se presentan algunas propiedades teóricas sobre la clasificación de datos funcionales en relación al ICA functional. Asimismo, desarrollamos un repertorio de técnicas relacionadas de datos funcionales diseñadas para el preprocesamiento y análisis de datos en el campo emergente de la neurociencia musical encarnada, cuyo objetivo es investigar la base neurológica de cómo el cuerpo influye en las experiencias musicales. En particular, se desarrollan dos métodos basados en aproximaciones no lineales de wavelets y polinomios para el preprocesamiento de actividad artefactual en señales EEG y pupilometría. Estos métodos producen resultados excelentes para la investigación neuromotora, a pesar de la condición subómptima de los datos registrados durante la actividad locomotora. Además, presentamos un conjunto de descriptores neurales derivados de datos recopilados a través de los mencionados métodos no invasivos, con el objetivo de desvelar el comportamiento cerebral durante interacciones musicales encarnadas. Más específicamente, nos centramos en metodologías para modelar la actividad neurotransmisora, un aspecto crítico demostrado como esencial en la funcionalidad motora y otras sensaciones propioceptivas. Nuestra investigación experimental se presenta mediante el concepto de emoción transferido al dominio neurológico, proporcionando un marco único para definir y capturar la esencia neural de la encarnación en la música.

Acknowledgements

I would like to extend my heartfelt gratitude to my supervisors and Prof. T. H. Fritz for their invaluable guidance and unwavering support. Their commitment to excellence, rigor, wealth of expertise, and vast experience has been instrumental in shaping my academic career. However, it is not just their professional prowess that I value; their exceptional humanistic qualities have also left an indelible mark on me. I firmly believe that this unique combination of attributes has truly enriched and defined the essence of the work presented here. I am also deeply thankful to all colleagues and friends at the Institute for Psychoacoustics and Electronic Music (UGhent), the Institute of Mathematics (UGR), and the Max Planck Institute for Human Cognitive and Brain Sciences, who have actively contributed to this research experience. To my (biological) family, my Belgian family, and my partner, I owe an immeasurable debt of gratitude for their patience and unconditional support throughout my studies. Thank you.

Meticulous reading, helpful insights and constructive comments of the members of the Examination Committee are greatly appreciated. Special thanks are extended to Nádia Moura for her review of the text and contributions to the current dissertation.

This work was partially supported by the Methusalem Funding from the Flemish Government, the project FQM-307 of the Government of Andalusia (Spain) and by the project PID2020-113961GB-I00 of the Spanish Ministry of Science and Innovation (also supported by the FEDER programme). I further acknowledge the financial support of the Consejería de Conocimiento, Investigación y Universidad, Junta de Andalucía (Spain) and the FEDER programme for the project A-FQM-66-UGR20 and the IMAG-María de Maeztu grant CEX2020-001105-M/AEI/10.13039/501100011033. This dissertation was also supported by the Max Planck Institute for Human Cognitive and Brain Sciences (Leipzig), with which I am associated as a guest researcher, conducting my investigations under the supervision of Prof. T. H. Fritz.

Research output

List of original publications in chronological order:

This is an article-based dissertation. The included published articles and manuscripts are indicated with an asterisk (*).

A1 peer-reviewed publications:

- *Vidal, M., Rosso, M., and Aguilera, A. M. (2021). Bi-smoothed functional independent component analysis for EEG artifact removal. *Mathematics*, 9(11):1243. DOI: 10.3390/math9111243.
  [Category: Mathematics, JCR year: 2020, IF: 2.258, Rank: 24/330, Quartile: Q1]

- *Vidal, M. and Aguilera, A. M. (2022). Novel whitening approaches in functional settings. *Stat*, 12(1):e516. DOI: 10.1002/sta4.516.
  [Category: Statistics & Probability, JCR year: 2021, IF: 2.451, Rank: 28/125, Quartile: Q1]

- *Vidal, M., Onderdijk, E., K., Aguilera, A. M., Six, J., Maes, P.-J., Fritz, T. H., and Leman, M. (2023). Cholinergic-related pupil activity reflects level of emotionality during motor performance. *European Journal of Neuroscience*, n/a:1–14. DOI: 10.1111/ejn.15998.
  [Category: Neurosciences, JCR year: 2022, IF: 3.4, Rank: 136/272, Quartile: Q2]

Publications under review:

- *Vidal, M., Moura, N., Aguilera, A. M., Fritz, T. H., and Leman, M. (2024). Geometric-based turbulence analysis of EEG signals for modeling emotional arousal during active immersive virtual interactions. *Under review*.

- *Vidal, M., Leman, M., and Aguilera, A. M. (2024). Functional independent component analysis by choice of norm: a framework for near-perfect classification. *Under review*.

- *Vidal, M. and Aguilera, A. M. (2024). Wavelet thresholding on independent subspace factorizations of spatially indexed wide functional data for robust estimation of cortical activity. *Under review*.

Other publications:

- Moura, N., **Vidal, M.**, Aguilera, A. M., Vilas-Boas, J. P., Serra, S., and Leman, M., (2023). Knee flexion of saxophone players anticipates tonal context of music. *npj Science of Learning*. Nature portfolio, 8(22). Shared first author. DOI: 10.1038/s41539-023-00172-z.
  [Category: Neurosciences, JCR year: 2022, IF: 4.2, Rank: 96/272, Quartile: Q2]

Computational packages:

- **Vidal, M.** and Aguilera, A. M. (2023). pfica: Independent components analysis techniques for functional data. R package version 0.1.3. Repository: https://cran.r-project.org/web/packages/pfica/

- **Vidal, M.**, Leman M., and Six, J. (2023). eaR: Perception-based auditory and music analysis. R package version 0.2.1. Repository: https://github.com/m-vidal/eaR

List of scientific activities in chronological order:

- **Vidal, M.** and Aguilera, A. M. *Independent Component Analysis Techniques for Functional Data.* Presented at the 5th International Workshop on Functional and Operatorial Statistics, Brno, Czech Republic, June 23–25, 2021. [Poster presentation]

- Besada, J.L. and **Vidal, M.** *Confronting Composers' Visual Imagery with Listeners' Experience of Music.* Presented at the 8th International Conference on Spatial Cognition, Rome, Italy, September 13–17, 2021. [Talk]

- **Vidal, M.**, Aguilera, A. M., Leman, M., and Vindas, J. *Estimation Process in Hilbertian Independent Component Analysis Models*. Presented at the 28th Annual meeting of the Royal Statistical Society of Belgium, Liège, Belgium, October 21–22, 2021. [Poster presentation]
- **Vidal, M.**, Leman, M., Aguilera, A. M., and Fritz, T. *Neuromodulatory Effects of Movement During Auditory-Motor Tasks.* Presented at the 17th Annual Neuromusic VConference, Hamilton, Ontario L8S 4L8, Canada, November 20, 2021. [Poster presentation]
- **Vidal, M.** and Aguilera, A. M. *On the independent components model for functional data.* Presented at the 39th Annual Congress of the Spanish Statistical Society, Granada, Spain, June 7–10, 2022. [Talk]
- **Vidal, M.**, Onderdijk, K.E., Aguilera, A. M., Six, J., Maes, P.-J., Fritz, T., and Leman, M. *Measuring arousal with pupil dilation during experience of musical agency*. Presented at the 2nd Conference on Music & Eye-Tracking, Frankfurt, Germany, July 7–8, 2022. [Talk]
- **Vidal, M.** and Aguilera, A. M. *Theoretical properties of isotropic functional random variables*. Presented at the 1st Joint Workshop on Functional Data Analysis and Nonparametric Statistics, Madrid, Spain, June 6–9, 2023. [Talk]
- **Vidal, M.** and Aguilera, A. M. *Smoothed functional principal/independent components: computational and theoretical considerations*. Presented at the 64th ISI World Statistics Congress, Ottawa, Canada, July 16–20, 2023. [Talk]
- **Vidal, M.** and Aguilera, A. M. *Wavelet-based Sparse Optimization via Fixed-point Iteration Scheme in High-dimensional Data Analysis*. Presented at the 21st IMACS World Congress (University of Rome "La Sapienza"), Rome, Italy, September 11–15, 2023. [Invited talk]
- **Vidal, M.** and Moura, N. *Relationships of prediction and alignment between musicians' movement and rhythmical and tonal contexts of music*. Presented at the 2nd International Conference MÚSICA ANALÍTICA, Coimbra, Portugal, October 12–14, 2023. [Talk]
- **Vidal, M.** and Aguilera, A. M. *Some properties of whitening transformations in function spaces*. Presented at the 40th Annual Congress of the Spanish Statistical Society, Elche, Spain, November 7–10, 2023. [Invited talk]
- **Vidal, M.** *The near-perfect classification phenomenon: an overview of functional classification techniques applied to data coming from digital humanities*. Presented at the 1st specialist course in "Functional data analysis and applications in digital and augmented humanities", Ghent, Belgium, November 28–30, 2023. [Talk]
- **Vidal, M.** and Aguilera, A. M. *Multivariate functional ICA for spatially indexed data*. Presented at the fda-lille: Functional Data Analysis Workshop, Lille, France, March 15–16, 2024. [Invited talk]

Related academic activities:

- Organization of the UGent specialist course "Functional data analysis and applications in digital and augmented humanities". November 28–30, 2023. Lecture contents: probability principles, curve approximation (Fourier, B-spline, Wavelets), functional PCA and inference with functional data. Invited researchers: A. M. Aguilera, J. Beran, A. Caponera, I. Daubechies, M. Leman, A. Menafoglio and A. Michałko.

Contents

Notation[I]

CNS	Central nervous system
EEG	Electroencephalography
FDA	Functional data analysis
GA	Generalized arousal
ICA	Independent component analysis
KL	Kosambi-Karhunen-Loève [expansion]
MRA	Multiresolution analysis
PCA	Principal component analysis
RKHS	Reproducing kernel Hilbert space
X	a Hilbertian random variable
$\mathbb{X}$	a whitened Hilbertian random variable
$\boldsymbol{X}$	a vector of Hilbertian random variables
f, g	deterministic functions
u, v	univariate vectors
A, C, D	matrices
$A^\top$	the transpose of A
I_q	$q \times q$ identity matrix
$\mathrm{diag}(A)$	diagonal matrix with same diagonal elements as A
$\mathrm{vec}(\cdot)$	vectorization operator
$\mathbb{N}$	positive integers
$\mathbb{N}_0$	$\mathbb{N} \cup 0$
$\mathbb{R}$	real numbers
$\mathbb{R}_+$	positive real numbers
$\mathbb{C}$	complex numbers
$\mathbb{E}$	mathematical expectation
$\mathbb{R}^q$	q-dimensional Euclidean space
$\perp\!\!\!\perp$	stochastic (or statistical) independence
$\perp$	orthogonality
H	a real separable Hilbert space
I_H	the identity of H
L^2	space of square integrable functions
$\mathscr{B}(\cdot)$	the Banach space of all bounded operators
$\mathscr{B}(\cdot)_{\mathrm{HS}}$	the space of Hilbert-Schmidt operators
$\mathrm{tr}(\cdot)$	trace operator
$(\Omega, \mathscr{A}, P)$	a basic probability space
$\mathbb{B}$	Borel σ-algebra
B	a Borel set
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
a.e.	almost everywhere
a.s.	almost sure
iff	if and only if
s.t.	subject to
w.r.t.	with respect to
$\mathbb{I}$	indicator function
$\equiv$	equality by definition
■	end of proof
§	section/subsection

[1] Additional abbreviations used in this dissertation can be found at the beginning of some chapters. Any modifications to the existing notation are meticulously indicated within their respective chapters.

List of Figures

List of Tables

Introduction

This thesis straddles two worlds: that of high/infinite dimensional statistics and that of music neuroscience. While the convergence between these two fields may not be immediately apparent, here we endeavor to explore their joint underpinnings while preserving certain degree of independence in their development.

We concentrate on the study of infinite-dimensional statistical models designed for analyzing high-dimensional objects such as functions, images, surfaces, among others. This area of research requires a strong knowledge in functional analysis, operator theory, as well as measure and probability theory. We may think of our data as a random object with some topological properties to exploit, such as smoothness. A way to do that is by assuming our random object lives in some function space with a predefined metric or geometry (inner product), allowing us to analyze their topological properties in relation to some probability axioms one may build. The kind of analysis we are currently discussing, known as "Functional Data Analysis" (FDA), was featured by J. O. Ramsay and colleagues in the late 1980s (or even earlier) [245], with a primary focus on data represented as functions. Since then, FDA has become a substantial and multifaceted field of statistics that has found applications in a number of disciplines. We refer to [93, 123, 158, 247, 309] for a comprehensive overview on FDA. The evolution of FDA appears to be transitioning toward object-oriented data analysis [187], a field encompassing more complex objects as molecules of the statistical analysis – say, a FDA++.

The primary aim of employing these advanced methods is to study the neural underpinnings of embodiment –i.e, the incorporation of bodily experiences– during musical interactions. Our approach aligns with the principles of naturalistic neuroscience, seeking to investigate brain function and behavior in ecologically valid contexts that mimic real-world scenarios. However, studying embodied interactions in such circumstances poses challenges with current non-invasive neuroimaging techniques, as data quality may be compromised by artifact-related activity. Addressing this issue has been a fundamental aspect in this investigation. Conversely, our approach acknowledges the multisensory nature of these processes by integrating behavioral, neuroscientific, and other relevant data into statistical models that consider information across multiple probabilistic dimensions.

Research objectives in mathematical statistics

The works presented herein, especially those of a statistical nature, were motivated by a thesis entitled "Alternativas geométricas en el ACP de una V.A Hilbertiana" [Geometric alternatives in principal component analysis (PCA) of a Hilbertian random variable] written by Prof. F .A. Ocaña (UGR) and supervised by Prof. A. M. Aguilera [224]. Our primary objective was to extend the results of the aforementioned thesis by delving deeper into the concept of functional independent component analysis (ICA), a dimension reduction technique that can be considered a refinement of the functional PCA.

Despite ICA has gained widespread popularity across diverse research domains since its foundational theory was developed around the 1990s, its functional counterpart has, in contrast, received relatively limited attention. Methods based on reproducing kernel Hilbert spaces were among the earliest to exploit ICA within an infinite-dimensional feature space [19, 273]. Not until early the 2010s, however, a theoretical framework for functional ICA was first presented in a conference proceedings [111]. While no posterior contributions on this topic are attributed to the authors, they introduced pioneering concepts critical to functional ICA. These encompass the notions of statistical independence, irreducibility and IC separability within the context of infinite-dimensional spaces. Some early attempts to materialize these principles using functional data can be found in [87, 197]. Subsequent to these initial works, the most notable contribution in addressing the estimation of functional ICs can be attributed to D. Peña and colleagues [231]. Although the authors did not delve into the specifics of the functional ICA model, they introduced the kurtosis operator, proposed a computational approach for its implementation while showed some of its theoretical properties with regard to the classification of Gaussian processes. The functional ICA model as such, was later introduced in the context of univariate and multivariate functional data in [176, 305]. While these papers are primarily geared towards practical applications in functional classification, neither of them has extended the theoretical results in [231]. Additionally, there are numerous potential applications of functional ICA beyond functional classification that have yet to be thoroughly explored.

Here, our aim is to provide a comprehensive study on functional ICA based on smoothing and other nonlinear estimators, demonstrating their relevance and applicability in analyzing neuroscientific data. As described below, we focus on studying brain behavior during embodied music interactions.

Research objectives in music neuroscience

Why do we feel compelled to move when listening to music? How is it possible that pianists can play without looking at the keyboard, or why can violinists accurately tune complex passages without even glancing at the fingerboard? To what extent does action contribute to shaping the experience of music? Or vice versa, does music perception influence our motor behavior? These questions are usually addressed within the embodied music cognition (EMC) hypothesis [169, 171]. It is undeniable that the brain and body form an inseparable tandem, shaping the way we interact with the environment. Nevertheless, the neurological mechanisms underlying motor behavior able to modulate and probably facilitate or enhance perception during the experience of music (i.e. embodiment) remain somewhat poorly understood; see [23, 156, 166] as earlier pioneering contributions to the field or more recent research aligned with this trend [101, 227, 257, 280, 281]. In this dissertation, we investigate the neurology of embodiment during musical interactions, an area of study that we refer to as *embodied music neuroscience* (EMN). One major objective of this thesis is to provide support for the concept of EMN based on empirical evidence.

Experimental investigations outlined here were initially motivated to uncover how movement could influence the perception of tonal harmony. Some insights into this question can be found in Moura et al. [204], where we demonstrated, using causal analysis methods and Leman's auditory model [136, 168], that bodily movement during music performance anticipated tonal context of music (surprisingly, knee flexion evidences this aspect of embodiment!). See also [202–204] for further details. Nevertheless, the present investigation focuses on other aspects of EMC. We aimed at identifying suitable descriptors for characterizing the dynamics of brain activity across different time scales and spatial dimensions during embodied music interactions. Therefore, we hypothesize that there might be a neural signature characterizing these processes of embodiment. To address this, we devised two experimental studies focusing on singing performance and emotion, as emotion plays a crucial role in configuring embodied music interactions, while singing can serve as a powerful medium for expressing these emotions [169]. Specific hypotheses and further discussion can be found in Chapter 4, Chapter 5, and §6.3. In both studies, we develop advanced FDA models to derive neural descriptors critical for understanding embodiment. Our research primarily relies on non-invasive neuroimaging techniques, specifically EEG and pupillometry.

Thesis overview and structure

In Chapter 1, we review the **main concepts and backgrounds of FDA** essential for the subsequent chapters. This includes elementary operator theory, methods for data representation using basis functions, and the fundamental principles of functional PCA. In Chapter 2, we present three papers [296, 300, 303] that focus on the **theory and applications of functional ICA**. The foundational background of ICA is covered in the initial subsections, followed by a rational integration of the aforementioned papers. Proofs for certain propositions and theorems are excluded in the two first chapters since they can be found in the associated published papers as detailed therein. Exceptions to this rule include cases where proofs are unpublished or are provided for the sake of clarity in our developments.

In Chapter 3, we introduce a methodology that bridges some of the theoretical underpinnings established Chapter 2 with our experimental investigations. There, we propose a **functional ICA approach based on wavelets for artifact removal** particularly tailored for EEG data acquired during experimental conditions involving body movement. The chapter discusses some issues regarding typical practices for EEG artifact removal, and illustrates the performance of the proposed method with thorough simulation studies and analysis of selected EEG datasets. Current pre-processing approaches will be applied to the EEG data recorded in one of our experiments.

We conducted two experimental studies to **investigate brain behavior during embodied music interactions**. In Study I (Chapter 4), we have investigated **how motor task related emotionality corresponds to pupillometry signatures** in a singing performance paradigm. We present a **multivariate functional PCA based on Pfaff's generalized arousal** principles [235] to study how pupil dynamics across different subbands relate to level of emotionality during the different motor tasks. The model is built upon prior investigations that relate pupil dilation to neurotransmitter activity. We also introduce an **unsupervised method to pre-process responses to ocular events** during motor tasks. This method is flexible, easy to apply, and has a low computational cost, yielding robust estimates of cognitive-related pupil activity. The experimental paradigm introduced in Study I is elaborated upon in a subsequent study that incorporates a more complex design involving immersive virtual reality interactions (Chapter 5). We introduce a **model for the analysis of turbulence in EEG recordings based on a multivariate functional ICA** for spatially indexed data that is derived from our prior research in Chapter 2 and Chapter 3.

Our aim is to analyze turbulent flows on dominant EEG components that are assumed to reflect generalized arousal function in a similar vein as in Study I. With this, we **investigate how the level of emotionality during embodied musical interactions in immersive virtual reality relates to turbulence activity in alpha and gamma ranges**.

1 Mathematical and statistical framework

This chapter provides definitions, notations, and basic concepts used throughout the dissertation. For an extensive account of the material discussed here, see [125, 224, 321].

§ 1.1. Basic probabilistic principles

Consider a sample space Ω and assume for convenience that $\mathcal{A}$ is a σ-algebra (an algebra closed under complementation and countable unions). Define a mapping $P : \mathcal{A} \to \mathbb{R}$, called probability, such that it adheres to the following axioms after Kolmogorov:

1. For each $E \in \mathcal{A}$, $P(E) \geq 0$,
2. $P(\Omega) = 1$ (the entire sample space has probability equal to 1),
3. $P(E_1 \cup E_2 \cup E_3 \cdots) = P(E_1) + P(E_2) + P(E_3) + \cdots$ if $E_i \cap E_j = \emptyset$ for all $i \neq j$ (all E's are disjoint).

Under these considerations, the triple $(\Omega, \mathcal{A}, P)$ is then called *probability space*. A probability space is usually build upon an abstract point set Ω, devoid of specific mathematical properties such as algebra or topology. To better understand and apply mathematical concepts to such spaces, we can explore different ways of connecting the elements of $\mathcal{A}$ to more structured mathematical spaces. In this dissertation, we deal with elements presumed to reside in a normed or Banach space (a vector space equipped with a norm function) with geometrical structure (inner product). This particular space, commonly referred to as *Hilbert space* and denoted here by H, is a mathematical construct that can encompass various objects, including functions, images, shapes, etc. A critical property of spaces with such structure is that every Cauchy sequence (of elements) converges to a limit within the space itself, which means the space is complete. This property ensures that geometrical operations can be consistently performed. In what follows, we can consider a Hilbert space H over the field of the reals $\mathbb{R}$, and $\langle \cdot, \cdot \rangle : H \times H \to \mathbb{R}$ an inner product on H inducing the norm $\| \cdot \| = \langle \cdot, \cdot \rangle^{1/2}$.

A common scenario in probability theory is to deal with mappings called random variables defined by $X : \Omega \to \mathbb{R}$, i.e., X is a function that assigns to each element of the sample space a real number. Instead, here we consider $X : \Omega \to H$, where H is a separable Hilbert space, and therefore each element of the sample space has now an associated element from H. We call X a *Hilbertian* or *H-valued random variable*. More formally, X is a mesurable function from $(\Omega, \mathcal{A}, P)$ to $(H, \mathbb{B}_H)$ where $\mathbb{B}_H$ is the Borel σ–field generated by the class of all open subsets of H. The probability of an event $E \in \mathbb{B}_H$ is determined by the measure induced on $(H, \mathbb{B}_H)$, defined as $P_X(E) = P(X^{-1}(E)) = P(\omega \in \Omega : X(\omega) \in E)$[1]. The induced measure of X is commonly referred to as its *distribution* or *law*.

Definition 1. *Let X be H-valued random variable. The element of H denoted by $\mathbb{E}X \in H$ verifying*

$$\langle \mathbb{E}X, x \rangle = \int_\Omega \langle X(\omega), x \rangle \mathrm{d}P(\omega) \quad \forall x \in H \tag{1.1}$$

is called the expectation of X.

Note that in 1.1 we are integrating the projection of X in the direction of x over all possible outcomes in the sample space with respect to the probability measure P.

The random variable X is said to be *Bochner* integrable if $\mathbb{E}\|X\| = \int_\Omega \|X(\omega)\| \mathrm{d}P(\omega) < \infty$, which implies the existence of the expectation or first moment of X. If exists, this expectation is unique. We can further consider the existence of the pth moment (for an integer p), i.e., $\int_\Omega \|X(\omega)\|^p \mathrm{d}P(\omega) < \infty$, leading us to extend the concept of common moment functions within the framework of Hilbert spaces.

[1] Note that $X^{-1}(\cdot)$ denotes the pre-image or inverse image function associated with X. Therefore, $X^{-1}(E)$ consists of all outcomes in the sample space that get mapped to the set E under X

§ 1.2. Moment operators

Preliminaries

An operator $\mathcal{T} : H \to H$ is said to be continuous if for every $\epsilon \in \mathbb{R}$ there is some $\delta \in \mathbb{R}$ such that $\|\mathcal{T}(x) - \mathcal{T}(y)\| < \epsilon$ whenever $\|x - y\| < \delta$, which ensures the operator behaves smoothly and predictably for all $x, y \in H$. If furthermore, $\mathcal{T}(ax + y) = a\mathcal{T}(x) + \mathcal{T}(y)$ for all $a \in \mathbb{R}$, the operator $\mathcal{T}$ satisfies the properties of linearity. Every linear operator is continuous in the finite dimensional case, but in infinite dimensions a linear operator $\mathcal{T}$ is only continuous iff it is bounded, i.e., $\|\mathcal{T}\| \equiv \sup\{\|\mathcal{T}(x)\| : \|x\| \leq 1\} < \infty$ where $\sup\{\cdot\}$ is the supremum function. The kernel of a linear operator $\mathcal{T}$, denoted $\ker(\mathcal{T})$, comprises all elements in the domain of $\mathcal{T}$ that map to the null space in the codomain. The range (or image) of $\mathcal{T}$, denoted $\mathrm{ran}(\mathcal{T})$, encompasses all possible output elements in the codomain obtained by applying $\mathcal{T}$ to elements in the domain.

If $F : H \to \mathbb{R}$ is a linear and bounded functional, there exists a unique element $y \in H$ such that $F(x) = \langle x, y\rangle$ for every $x \in H$ (Riesz' representation theorem [321, §1.2]). It follows that if $\mathcal{T}$ is bounded, its adjoint $\mathcal{T}^*$ is also bounded and are related by the equation $\langle \mathcal{T}x, y\rangle = \langle x, \mathcal{T}^* y\rangle$ for all $x, y \in H$. If $\mathcal{T}^* = \mathcal{T}$ and $\langle \mathcal{T}x, x\rangle \geq 0$ the operator is self-adjoint (Hermitian) and positive definite. Note that a positive operator can be raised to any positive power; therefore, $(\mathcal{T}^{1/n})^n = \mathcal{T}, n \in \mathbb{N}$. Operators that commute with the adjoint are normal. As an example, consider the isometry $\|\mathcal{T}x\| = \|x\|$, then $\mathcal{T}$ is a unitary operator with the property $\mathcal{T}\mathcal{T}^* = \mathcal{T}^*\mathcal{T} = I_H$, where I_H is the identity operator.

A fundamental property of a Hilbert space is the existence of the so-called orthonormal basis. Two elements $x, y \in H$ are said to be orthogonal if $\langle x, y\rangle = 0$. A sequence of elements $\varphi \equiv (\varphi_j)_{j\in\mathbb{N}}$ on H is called orthonormal basis if all elements of φ are mutually orthogonal, have unit norm and the series

$$x = \sum_{j=1}^{\infty} \langle x, \varphi_j\rangle \varphi_j, \tag{1.2}$$

is convergent in the norm topology of H [321, pp.1-6], for all $x \in H$. Then, $\|x\|^2 = \sum_{j=1}^{\infty} |\langle x, \varphi_j\rangle|^2$, which follows from Parseval's Theorem. Equation 1.2 can alternatively be expressed as $x = \mathcal{P}_\varphi(x)$, where $\mathcal{P}_\varphi(\cdot)$ denotes the orthogonal projection on to the subspace generated by φ. The operator $\mathcal{P}_\varphi$ is idempotent ($\mathcal{P}_\varphi^2 = \mathcal{P}_\varphi$), and self-adjoint ($\mathcal{P}_\varphi = \mathcal{P}_\varphi^*$). The orthogonal complement of the subspace generated by φ is $(I_H - \mathcal{P}_\varphi)$ and the rank of the projection operator is equal to the dimension of the subspace on to which it projects.

A bounded linear operator $\mathcal{T} : H \to H$ is said to be compact if it maps bounded sets in H to sets that have all their limit points contained within a closed set with respect to the norm topology on H (see [321, §1.3]). A critical property of a compact operator in H is the existence of a canonical decomposition. For two orthonormal sets $(\varphi_j, \psi_j)_{j\in\mathbb{N}}$ in H one has that $\mathcal{T}x = \sum_{j=1}^{\infty} \theta_j \langle x, \varphi_j\rangle \psi_j$ (or equivalently, $\mathcal{T}\varphi_j = \theta_j\psi_j$), where θ_j is a sequence of real numbers decreasing to 0 called the eigenvalues of $\mathcal{T}$. If furthermore $\mathcal{T}$ is compact self-adjoint, the same result is expressed in terms of a unique basis, i.e., $\mathcal{T}\varphi_j = \theta_j\varphi_j$. A compact operator $\mathcal{T}$ is said to be trace class if $\mathrm{tr}(\mathcal{T}) = \sum_{j=1}^{\infty} \langle \mathcal{T}\varphi_j, \varphi_j\rangle$ converges independently of the chosen orthonormal basis. We note that having finite trace as defined by the above quantity does not guarantee the invertibility of $\mathcal{T}$, since zero can be an accumulation point in the spectrum of $\mathcal{T}$. A trace-class (or nuclear) operator $\mathcal{T}$ is of the Hilbert-Schmidt (HS) class, often called Schatten-2 [321, §1.4], if $\mathrm{tr}(\mathcal{T}^*\mathcal{T}) < +\infty$. The class of all HS operators forms a separable Hilbert space itself, and will be denoted as by $\mathcal{B}_{HS}(H)$. The associated operator norm, denoted as $\|\mathcal{T}\|_{\mathrm{HS}}^2 = \mathrm{tr}(\mathcal{T}^*\mathcal{T})$ for all $\mathcal{T} \in \mathcal{B}_{\mathrm{HS}}(H)$, is a generalisation of the Frobenius norm for finite-dimensional matrices.

Covariance, skewness and kurtosis operators

Here, the term *moment operators* refers to linear mappings that capture information about the moments of an H-valued random variable. Laha and Rohatgi [165, pp.473] identified operators of this type to possess certain properties such as compactness, self-adjointness, positive definiteness and finite trace in the HS sense. In the following, we consider probability measures P on $(H, \mathbb{B}_H)$ such that $\int_\Omega \|X(\omega)\|^4 \mathrm{d}P(\omega) < \infty$. Unless otherwise stated, the mean value $\mathbb{E}X$ will be assumed 0. Under these assumptions, the covariance (or second moment) operator $\mathcal{C}_X \in \mathcal{B}_{HS}(H)$ is uniquely determined by the relation

$$\langle \mathcal{C}_X(x), y\rangle = \int_\Omega \langle X(\omega), x\rangle\langle X(\omega), y\rangle \mathrm{d}P(\omega). \tag{1.3}$$

It is well-known that 1.3 is positive definite, self-adjoint (therefore symmetric) with finite HS norm or trace; see [165, pp.471–73].

Now consider the linear operator $x \otimes y : H \to H$, called the tensor product operator, and defined by $(x \otimes y)(\cdot) = \langle \cdot, x \rangle y$ for all $x, y \in H$. This operator has rank 1 and is HS.

Lemma 1 (Ocaña [224], Lemmas 1.3.2, 1.3.4)**.** *The covariance operator defined in 1.3 can be expressed as* $\mathbb{E}(X \otimes X)$.

In accordance with the given definition of the tensor product, we now enumerate some of its properties, as they will be utilized throughout this dissertation. Let $\mathcal{B}(H)$ denote the algebra of all bounded linear operators on H.

Lemma 2. *For* $x, y, z \in H$ *and* $\mathcal{T}_1, \mathcal{T}_2 \in \mathcal{B}(H)$

1. $(x \otimes y)^* = y \otimes x$;
2. $(\mathcal{T}_1 \otimes \mathcal{T}_2)(x \otimes y) = (\mathcal{T}_1 x) \otimes (\mathcal{T}_2 y) = \mathcal{T}_1 (x \otimes y) \mathcal{T}_2^*$;
3. $(z \otimes y)(y \otimes x) = \|y\|^2 (z \otimes x)$.

Peña et al. [231], based on [201], defined a kurtosis (fourth-moment) operator $\mathcal{K}_X \in \mathcal{B}_{HS}(H)$ which is uniquely determined by the relation

$$\begin{aligned} \langle \mathcal{K}_X(x), y \rangle &= \int_\Omega \langle X(\omega), X(\omega) \rangle \langle X(\omega), x \rangle \langle X(\omega), y \rangle \mathrm{d}P(\omega) \\ &= \int_\Omega \|X(\omega)\|^2 \langle X(\omega), x \rangle \langle X(\omega), y \rangle \mathrm{d}P(\omega) \end{aligned} \tag{1.4}$$

The operator $\mathcal{K}_X$ shares the same properties as $\mathcal{C}_X$ (see §2.4), and following the same reasoning provided in Lemma 1, $\mathcal{K}_X$ can be further expressed as $\mathbb{E}\{(X \otimes X)^2\} = \mathbb{E}\{\|X\|^2 (X \otimes X)\}$. It is a common procedure to work with whitening representations of X in order to remove scaling effects and ensure that the kurtosis operator accurately reflects the shape of the distribution that characterizes X. Whitened functional random variables are discussed in Chapter 2.

To our knowledge, a third moment operator of an H-valued random variable has not yet been established in the literature. Extending the skewness matrix of a multivariate random variable (e.g., $YY^\top Y$ as in [201]) into the form of a normal operator is not feasible, somehow limiting the application of the Spectral Theorem. In finite dimensions, the diagonalization of non-normal operators can be addressed using the Jordan decomposition; however, in infinite dimensions, alternative methods must be pursued. One potential approach could be to define an skewness operator as follows:

$$\mathcal{S}_X = \mathbb{E}\left(\sqrt{(X \otimes X)^3}\right), \tag{1.5}$$

which can be shown to be related to the power two of the skewness matrix in [201]. It is easy to prove that the operator $\mathcal{S}_X \in \mathcal{B}_{HS}(H)$ shares the same properties of the moment operators $\mathcal{C}_X$ and $\mathcal{K}_X$. The study of this operator, however, lies beyond the scope of the current dissertation.

§ 1.3. Hilbertian independence and irreducibility

In this subsection, we elaborate on the notion of independence in Hilbert spaces, a concept that lacks a straightforward intuition compared to its definition in conventional probability theory. Independence is a central and somewhat elusive concept in probability, often ambiguously used and conflated with other properties such as uncorrelatedness and sparsity.

Consider a basic probability space $(\Omega, \mathcal{A}, P)$ and a probability measure $P : \mathcal{A} \to [0, 1]$. Two events $E_1, E_2 \in \mathcal{A}$ are said to be statistically independent iff $P(E_1 \cap E_2) = P(E_1) \cdot P(E_2)$, i.e., their joint probability equals the product of their probabilities. Further, let $\mathcal{I}$ to be an index set, with the index i varying on $\mathcal{I}$. A collection of events $(E_i)_{i \in \mathcal{I}}$ are said to be independent iff for any finite subset $\mathcal{I}_n = \{i_1, i_2, \ldots, i_n\}$ of $\mathcal{I}$

$$P\left(\bigcap_{i \in \mathcal{I}_n} E_i\right) = \prod_{i \in \mathcal{I}_n} P\left(E_i\right). \tag{1.6}$$

Note that the above definition implies more than just asking $P\left(\bigcap_{i \in \mathcal{I}} E_i\right) = \prod_{i \in \mathcal{I}} P\left(E_i\right)$, and still further more than pairwise independence. In other words, we have that subclasses of independent classes are necessarily independent. Stepniak [284] gives a didactic illustration of statistical independence, exemplifying cases of joint and pairwise independence. For a more advanced treatment of the subject, see [174].

We now introduce some concepts to show how independence is conceived in an infinite dimensional Hilbert space H. Consider H^*, the dual space of H, i.e., H^* is the separable Hilbert space consisting of all bounded linear functionals from H to $\mathbb{R}$, and X a H-valued random variable. Then for every $l \in H^*$, $l(X)$ is a real valued random variable. Two H-valued random variables X, Y defined on $(\Omega, \mathcal{A}, P)$ are said to be identically distributed if $P_X = P_Y$. Let $(X_i)_{i\in\mathbb{N}}$ be a collection of H-valued random variables. If $l(X_i)$ is a collection of identically distributed random variables, then X_i are identically distributed. See [165, pp.447-8] for further details.

Definition 2. *A sequence of random variables $X_1, X_2, \cdots : \Omega \to H$ is said to be independent iff for every $n \in \mathbb{N}, n \geq 2$, any (finite) subset $\mathcal{I}_n \subseteq \mathbb{N}$ of cardinality n, and all mesurable sets $E_i \in \mathbb{B}_H$ $(i \in \mathcal{I}_n)$, the relation*

$$P\left(\bigcap_{i\in\mathcal{I}_n} \{X_i \in E_i\}\right) = \prod_{i\in\mathcal{I}_n} P\left(\omega \in \Omega : X_i(\omega) \in E_i\right), \tag{1.7}$$

holds.

Note that independence of an infinite number of events is defined by demanding that every finite subset is independent.

Remark 1. *If $\{l_i(X_i), i \in \mathcal{I}_n, l_i \in H^*\}$ is a collection of independent random variables, then $\{X_i, i \in \mathcal{I}_n\}$ is also independent.*

So far, we have defined independence through finite sequences of variables residing in an infinite-dimensional separable Hilbert space. However, determining when X (a single random variable) is independent becomes a more nuanced task, and understanding the implications of this concept is crucial in refining Definition 2.

Definition 3 (Independent random variable, Gutch and Theis [111], Definition 2)**.** *A random variable $X : \Omega \to H$ is said to be independent if for an orthonormal basis $(\varphi_j)_{j\in\mathbb{N}}$ of H and for $q \in \mathbb{N}, q \geq 2$, any subset $\mathcal{I}_q \subseteq \mathbb{N}$ of cardinality q,*

1. *The random variables $\langle X, \varphi_{j_1}\rangle, \ldots, \langle X, \varphi_{j_q}\rangle$ are independent;*
2. *$\mathcal{P}_{\phi_{\mathcal{I}}}(X)$ and $\mathcal{P}_{\varphi_{\overline{\mathcal{I}}}}(X)$ are independent, where $\mathcal{P}_{\varphi_{\overline{\mathcal{I}}}}(X) = (I_H - \mathcal{P}_{\varphi_{\mathcal{I}}})(X)$.*

Assumption 2 in Definition 3 reads as "the orthogonal projection to the subspace spanned by a subset of the basis is independent to the projection to the complement of the subspace" and, while difficult to prove analytically due to the ranks of such projections, it bears critical implications for the study of functional ICA. A desirable property of a projection is that $\mathcal{P}_{\varphi_J}(X) \perp \mathcal{P}_{\varphi_{\overline{J}}}(X)$, which usually holds when $\mathcal{P}_{\varphi_{\overline{J}}}(X)$ is null or a negligible white noise, but not necessarily when $\mathcal{P}_{\varphi_J}(X)$ is reduced to a few components. Assumption 2 goes beyond linear independence and rather entails that both $\mathcal{P}_{\varphi_J}(X)$ and $\mathcal{P}_{\varphi_{\overline{J}}}(X)$ are mutually independent to define X as independent. This occurs in very particular and restrictive scenarios, which eventually reduces to an orthogonality constraint in Gaussian cases; see [40, Lemma 1.2.6]. From now on, we will use the symbol $\perp\!\!\!\perp$ to denote statistical independence[2] in our developments.

We also note that if assumption 1 Definition 3 in holds, then $S_{j_1} \perp\!\!\!\perp \cdots \perp\!\!\!\perp S_{j_q}$, where $S_j = \langle X, \varphi_j\rangle\varphi_j$ are Hilbert-valued functional independent components. Additionally, the independent variables $S_{j_1}, \ldots, S_{j_q}$ are called irreducible if further decomposition of any of the components is not possible. Current notion of independence will facilitate our understanding of the functional ICA model proposed in Chapter 2.

In the context of Hilbert spaces, we have seen that independence is defined through an isometric isomorphism between the dual space H^* and the original space H. This isomorphism allows the identification of Borel measures on H^* with those on H when an orthonormal basis of H is chosen. Through this identification, the statistical properties of X can be analyzed, as otherwise there is no explicit parametric form assumed for the distribution of elements in an abstract Hilbert space. Going back to Definition 2, we note that joint independence is likely to be true if any of the X_i's is not independent in the sense of Definition 3.

§ 1.4. Featuring the Hilbert space with basis functions

Functional statisticians work in spaces with structure that allows them to reflect the underlying regularity of the processes they study. Although the (separable) Hilbert space of square integrable functions $H \equiv L^2$ is

[2] It is worth mentioning that the symbol $\perp\!\!\!\perp$, denoting stochastic independence, can be credited to Gustav Elfving, a Finnish statistician, probabilist, and mathematician, who first used such symbol in some lecture notes around 1949–1950; see [219].

certainly the most common starting point, Hilbert spaces of smoother functions such as kernel or Sobolev spaces are typically preferred. In real-world applications, it is however impossible to observe a process continuously and data is commonly sampled as a finite set of discrete observations. Here, the approach for analyzing these data involves utilizing basis of L^2 functions for their approximation. In the end, this allows working with coefficients obtained through a projection on to a predefined basis (whether or not it is orthonormal) that better captures the characteristics of our data.

Now, we formulate the problem of approximation of a single realization of an arbitrary functional random variable. We study the classic non-parametric problem of recovering the values of an unknown function $\{f \in L_T^2 : T \subseteq \mathbb{R}\}$ from noisy observations $y_i \in \mathbb{R}$. This corresponds to

$$y_i = f(t_i) + \varepsilon_i \quad (i = 1, \dots, n), \tag{1.8}$$

where $\varepsilon_i \in \mathbb{R}$ is an error term, and $t_1, t_2, \dots, t_n$ are sampling points here assumed in an equidistant grid within T. We are concerned with estimators of f based on different systems of basis functions, including Fourier, B-splines, and Wavelet bases. Because basis functions form a vector space, similarly to 1.2, we can represent f as

$$f(t) = \sum_{j=1}^{q} a_j \phi_j(t), \tag{1.9}$$

where $\phi_j(t)$ are the set of q basis functions defining the space and a_j their respective coefficients.

A classical orthonormal basis system is the Fourier basis. This basis consists of a series of periodic trigonometric functions on $T = [0, 1]$ given by

$$\{\sqrt{2}\sin(2\pi jt); j \in \mathbb{N}\} \cup \{\sqrt{2}\cos(2\pi jt); j \in \mathbb{N}\} \cup \{1\}. \tag{1.10}$$

The Fourier basis is a typical choice for the approximation of stable and periodic functions that have a roughly constant curvature [292].

To represent curves that are not uniformly smooth, splines are a more suitable choice due to their good local support. Splines are curves made of piecewise polynomial functions that smoothly join at prefixed knots. A B-spline, short for "basis spline", is a curve defined by a basis system of degree p that generates a space of splines of the same degree. Let $\mathtt{U} = \{u_1, u_2 \dots, u_k, \dots, u_m\}$, be a sequence of knots over $T \subset \mathbb{R}$, i.e an increasing and uniformly-spaced sequence of $m \geq 1$ real numbers. For each $k = 1, \dots, m$, a B-spline basis function of degree p (denoted by $\mathfrak{B}_{k,\mathtt{U}}^p(x)$) is defined recursively as

$$\mathfrak{B}_{k,\mathtt{U}}^p(x) = \frac{x - u_k}{u_{k+p} - u_k}\mathfrak{B}_{k,\mathtt{U}}^{p-1}(x) + \frac{u_{k+p+1} - x}{u_{k+p+1} - u_{k+1}}\mathfrak{B}_{k+1,\mathtt{U}}^{p-1}(x), \tag{1.11}$$

where

$$\mathfrak{B}_{k,\mathtt{U}}^0(x) = \begin{cases} 1 & \text{if } x \in [u_k, u_{k+1}), \\ 0 & \text{otherwise.} \end{cases} \tag{1.12}$$

The above formula when $p = 3$ (meaning that cubic polynomials are used) produce the popular "Cubic B-splines". These are smooth curves with continuous first and second derivatives. For a knot sequence with m unique elements, $q = m - p - 1$ B-splines of degree p can be generated.

In this dissertation, we will also work with wavelet expansions. Wavelet basis functions are known for their versatility in handling non-uniform curves, particularly to approximate abrupt changes and complex varying local behaviors. A wavelet is an oscillatory function $\psi \in L_{\mathbb{R}}^2$ satisfying $\int_{-\infty}^{+\infty} \psi(x)\mathrm{d}x = 0$, $\|\psi\| = 1$, centered in the neighborhood of $x = 0$ and decaying to 0 as $x \to \pm\infty$. The dyadic dilations and translations of any wavelet design convoluted on $f(t) \in L_T^2$ produce the so-called Multiresolution Analysis (MRA) [71, 186], which is an orthogonal decomposition of a function in a set of wavelet basis coefficients at different levels of resolution corresponding to the dimension of the wavelet basis.

How do we choose the most suitable basis for fitting the data, along with determining the appropriate basis dimension? While providing a comprehensive answer would be rather extensive, a common approach involves using an appropriate loss function, such as ordinary least squares, combined with a cross-validation procedure. Then, choosing the optimal finite-dimensional representation of $f(t)$ corresponds to finding the vector of coefficients that minimize the sum of squared errors at each t_i. Thus, the objective is to minimize

$$\mathrm{MSE}(y \mid a) = (y - \Phi a)^\top (y - \Phi a), \tag{1.13}$$

where $\Phi = (\phi_j(t_i))_{n\times q}$ is a matrix containing the discretized basis functions. Then, the estimate of a that minimizes the mean squared error is $\hat{a} = (\Phi^\top\Phi)^{-1}\Phi^\top y$.

To enhance the accuracy of the basis approximation, refinement is sometimes necessary to mitigate possible overfitting effects. This can be achieved through the application of a roughness penalty. Consider the integrated squared d-order derivative

$$\int_T \left[R^{[d]}f(t)\right]^2 \mathrm{d}t = a^\top \mathfrak{P}_d a, \tag{1.14}$$

where the matrix $\mathfrak{P}_d$ is defined by $\mathfrak{P}_d = \int_T R^{[d]}\phi(t)R^{[d]}\phi(t)^\top \mathrm{d}t$ with $R^{[d]}\phi(t) = \left(R^{[d]}\phi_1(t),\ldots,R^{[d]}\phi_q(t)\right)^\top$. In the penalized least squares problem, the coefficients of the smoother are determined by minimizing

$$\mathrm{CPMSE}_d(y \mid a) = (y-\Phi a)^\top (y-\Phi a) + \theta a^\top \mathfrak{P}_d a. \tag{1.15}$$

where $\theta \in \mathbb{R}_{\geq 0}$ is a penalty parameter. Then, the coefficients are estimated as $\hat{a} = (\Phi^\top\Phi + \theta\mathfrak{P}_d)^{-1}\Phi^\top y$.

When the above regression problem is formulated with B-splines, we refer to them as smoothing B-splines. We can readily convert this estimator into a more computationally efficient one by using a discrete penalty, which corresponds to what is known as penalized spline (P-splines) regression. Then, $\mathfrak{P}_d$ is calculated as a matrix representation of the d-order difference operator $R^{[d]}$; for additional details and advanced approaches, see [4, 67, 85, 212]. In the case of wavelets, a variety of strategies can be employed depending on the MRA approach. These generally include the use of shrinkage techniques, as discussed in Chapter 3, as well as other similar smoothing schemes like those outlined here; see [13] and references therein.

§ 1.5. Functional principal component analysis

The gestation of functional principal component analysis (functional PCA) owes significantly to Mercer's Theorem (1909) and the Kosambi- (1943 [161]) Karhunen- (1946-7 [144, 145]) Loève (1948 [184]) Theorem (KKLT, or KLT for short). Jointly interpreted, Mercer's theorem states that any positive semi-definite symmetric function can be represented as a convergent sum of products of basis functions in $H \equiv L^2_T$, while KLT uses such decomposition for the expansion of a random function in H. Related works initially emerged for inferential purposes on stochastic processes [110, 249], while the consolidation of functional PCA as reduction technique started after Dauxois' asymptotic study [73], which investigated whether the eigenelements of the empirical covariance operator converged to the eigenelements of its theoretical counterpart as well as their asymptotic distributions; see also [42, 123] for greater generality of the results. Early works on functional PCA include [36, 245, 263], which anticipated preliminary stages of an incipient computational development [7, 8, 246]. In the 90s, FDA and functional PCA underwent significant progress. Remarkably, for what this dissertation concerns, two smoothed functional PCA approaches [254, 276] were introduced, the latter incorporating the roughness penalty into the orthonormality constraint of the covariance eigenfunctions via Sobolev norms and inner products; see also [167, 242] for extended asymptotic results on Silverman's functional PCA method. Ocaña et al. [223] generalized this procedure a few years later and established the principles to compute functional PCA estimates under general settings [222]. By that time, Ramsay and Silverman's popular book on FDA [247] was published, which, along with Bosq's book [42], constitute two fundamental state-of-the-art contributions to the field. In parallel, KLT was also studied beyond common function spaces, for example, in the context of random fields [14] and on Riemannian manifold structures; see [89] for the general case.

Theory

Suppose $(H, \langle\cdot,\cdot\rangle)$ is a separable Hilbert space of real valued continuous functions, and that one aims to find an optimal approximation, say $\tilde{X}$, of the random variable X taking values H. This problem corresponds to finding the linear span that minimizes $\mathbb{E}\|X-\tilde{X}\|^2$, and it is commonly referred to as functional PCA.

Since we are dealing with countable bases, it is conceivable to define a procedure to identify coordinates/coefficients with the highest variability at each step, thereby simultaneously minimizing the mean square error mentioned above. We therefore aim to find a sequence of real-valued random variables, $(\xi_j)_{j\in\mathbb{N}}$, which are generalized linear combinations of X with maximum variance. These variables are commonly referred to as principal components (PCs) or KL expansion coefficients. An iterative method can be formulated to find a sequence of orthonormal functions $(\gamma_j)_{j\in\mathbb{N}}$ whose elements verify

- $\gamma_1 = \underset{\|h\|^2=1}{\arg\max}\, \langle \mathscr{C}_X(h), h\rangle$,

- if $j > 1$, then $\gamma_j = \underset{\|h\|^2=1}{\arg\max} \langle \mathscr{C}_X(h), h\rangle$ and $\left\langle \mathscr{C}_X\left(\gamma_{j'}\right), h\right\rangle = 0, \quad \forall j' < j,$

such that the variance of the random variables $\xi_j = \langle \gamma_j, X\rangle$ is maximized. This problem is well-known to be equivalent to the decomposition of the covariance operator, therefore the functional PCA is obtained from the eigensystem

$$\mathscr{C}_X\left(\gamma_j\right) = \lambda_j \gamma_j, \tag{1.16}$$

where the eigenvalues are positive satisfying $\lambda_j \geq \lambda_{j+1}, \forall j \in \mathbb{N}$. The solution to 1.16 is uniquely defined when the λ_j's have multiplicity one. Assuming $\left\langle \gamma_j, \gamma_{j'}\right\rangle = \delta_{j,j'}$, where $\delta_{j,j'}$ is the Kronecker delta, this orthonormal property is translated to the ξ_j's as following orthogonal property

$$\mathbb{E}(\xi_j \xi_{j'}) = \left\langle \mathscr{C}_X\left(\gamma_j\right), \gamma_{j'}\right\rangle = \lambda_j \delta_{j,j'}. \tag{1.17}$$

Additionally, observe that if $\lambda_j = 0$, then $\gamma_j = 0$ (a.e.). Because of 1.16, the current definition of functional PCA can be modified by substituting the constraint which defines the jth eigenfunction by $\langle \gamma_{j'}, h\rangle = 0, \forall j' < j$, which allows for other geometrical modifications as discussed in [223].

Proposition 1. *Let $X : \Omega \to H$ be a random variable and $(\lambda_j, \gamma_j)_{j\in\mathbb{N}}$ the eigensystem of $\mathscr{C}_X$ such that the γ_j's form an orthornormal family and $(\lambda_j) \subseteq \mathbb{R}$ is a sequence of values sorted in a decreasing order. Then,*

- *$(\gamma_j)_{j\in\mathbb{N}}$ determine the functional PCA via $\langle \gamma_j, X(\omega)\rangle$;*
- *$(\xi_j)_{j\in\mathbb{N}}$ are uncorrelated random variables and $\mathbb{E}\xi_j^2 = \lambda_j$;*
- *$(\lambda_j)_{j\in\mathbb{N}}$ is a decreasing sequence of distinct nonnegative real values that converge to zero.*

Therefore, if $\mathscr{C}_X$ admits spectral decomposition, the expression for the KL expansion is

$$X(\omega) = \sum_{j=1}^{\infty} \xi_j(\omega)\gamma_j, \tag{1.18}$$

where $\xi_j(\omega) = \langle \gamma_j, X(\omega)\rangle$.

Theorem 1 (Ocaña et al. [223], Theorem 4.10)**.** *Let H be defined by any continuous inner product $\langle\cdot,\cdot\rangle_\tau$ such that for a symmetric positive definite operator $\mathcal{T} : (H, \langle\cdot,\cdot\rangle_\tau) \mapsto (H, \langle\cdot,\cdot\rangle_\tau)$ one can define*

$$\langle f, g\rangle_\rho = \langle \mathcal{T}(f), g\rangle_\tau = \langle \mathcal{T}^{1/2}(f), \mathcal{T}^{1/2}(g)\rangle_\tau. \tag{1.19}$$

Then, the functional PCA of X with $\langle\cdot,\cdot\rangle_\rho$ is equivalent to the functional PCA of $\mathcal{T}^{1/2}(X)$ with $\langle\cdot,\cdot\rangle_\tau$, in the sense that their KL expansions with both inner products are related as follows:

- *$X(\omega) = \sum_j \xi_j(\omega)\gamma_j$ (functional PCA of X with $\langle\cdot,\cdot\rangle_\rho$);*
- *$\mathcal{T}^{1/2}(X)(\omega) = \sum_j \xi_j(\omega)\mathcal{T}^{1/2}\left(\gamma_j\right)$ (functional PCA of $\mathcal{T}^{1/2}(X)$ with $\langle\cdot,\cdot\rangle_\tau$).*

As observed, the PCs remain unchanged, and the eigenfunctions are related by $\mathcal{T}^{1/2}$.

Theorem 1 establishes the existence of a bijective linear map between two Hilbert spaces that preserves the inner product structure. This result is critical for the empirical computation of the functional PCA and will be taken into account in Chapter 2.

The first major consequence of KL expansion is that it leads to an explicit characterization of the reproducing kernel Hilbert space (RKHS) of X, in particular for all the elements in H which are, under certain conditions, in the closure of the span of the γ_j's. This aspect is thoroughly discussed in Chapter 2. Furthermore, the current theory can be extended to spaces of the type $H \equiv H_1 \times H_2 \times \cdots \times H_n$ (cartesian product), each of those corresponding to an H-valued random variable; see, for example, [134, 159, 247]. We have developed methods for analyzing groups of H-valued random variables (multivariate functional data) in our experimental studies.

2 Functional independent component analysis

This chapter presents a summarized compendium of the following papers/manuscripts:

- [303] Vidal, M., Rosso, M., and Aguilera, A. M. (2021). Bi-smoothed functional independent component analysis for EEG artifact removal. *Mathematics*, **9**(11):1243.
- [296] Vidal, M. and Aguilera, A. M. (2022). Novel whitening approaches in functional settings. *Stat*, **12**(1):e516.
- [300] Vidal, M., Leman, M., and Aguilera, A. M. (2024a). Functional independent component analysis by choice of norm: a framework for near-perfect classification. *Under review*.

The above papers have been synthesized in one single chapter to maintain a coherent discourse in the dissertation. Here, we begin with a concise introduction to multivariate independent component (IC) models and discuss the concept of separability, which is critical to ICA and functional ICA. In §2.2, we establish the conditions for the existence of a whitening transformation in infinite-dimensional spaces, a pre-processing step often used in the majority of ICA algorithms. Further, in §2.3, we introduce the notion of whitening operator, study the properties of various whitening transformations, and derive computational algorithms for the estimation of the proposed transformations in terms of basis expansions. Both sections contain the main theoretical results published in [296]. Subsequently, in §2.3, the functional IC model is introduced, and we provide a formal definition of the kurtosis operator. In the next sections, two novel smoothed functional ICA models based on kurtosis are presented. The former (§2.5) is based on the notion of penalized kurtosis, a concept that follows from Silverman's method [276] for smoothing principal components we introduce in [300]. There, the penalties are directly imposed to the eigenfunctions of the kurtosis operator. The latter (§2.6), uses the original Silverman's method to estimate functional ICA on a suitable smoothed KL expansion via the KL coefficients (see [303]). Furthermore, in §2.7 we present the theoretical properties of the kurtosis operator in relation to a generalized Fisher discriminant function and the relationship it entails with the Feldman-Hájek dichotomy for Gaussian measures. Current theoretical results can be found in [300]. Finally, we illustrate the performance of our methods through simulations and the analysis of various real datasets in binary classification problems.

§ 2.1. Independent component model and separability: from finite to infinite

ICA was motivated by neurophysiological challenges in the 1980-90s, as PCA encountered limitations to accurately separate mixed signals into their underlying independent components (ICs). See e.g., [141], or [128] for instructive examples. Since then, there has been a substantial body of research on various types of ICA procedures and their interpretations. In general, these procedures can be categorized into two main classes: one involves specifying a particular parametric model for the distributions of the ICs, while the other adopts a semiparametric or non-parametric approach that makes no or weak assumptions about their underlying distributions.

In the multivariate ICA model, one assumes that a centered p-variate random vector X is a linear mixture of a p-variate vector of mutually ICs. The model is commonly expressed as $X = AS$, where $A \in \mathbb{R}^{p\times p}$ is a nonsingular matrix and S a random vector of ICs. The goal is therefore to estimate A and S given only X, which is challenging due to inherent indeterminacies of the model, such as A being unidentifiable. Nevertheless, if S has non-Gaussian components and AS is again independent, it is possible to show that A can be represented as, at most, the product of a permutation and scaling within the components. This result was initially validated in [64] through the following theorem:

Theorem 2 (Darmois–Skitovitch [129, 279, 291]). *Suppose $S_1, \ldots, S_n$ are independent random variables, and let*

$$L_1 = \sum_{j=1}^{n} a_j S_j \quad \text{and} \quad L_2 = \sum_{j=1}^{n} b_j S_j \tag{2.1}$$

be two linear forms with $(a_j, b_j) \in \mathbb{R}$. If we assume $L_1 \perp\!\!\!\perp L_2$, then all S_j for which $a_j b_j \neq 0$ are normally distributed.

Darmois-Skitovich's Theorem establishes that those S_j in 2.1 must be Gaussian if L_1 and L_2 are independent. The theorem indirectly supports the that if the components of S can be assumed non-Gaussian, they can be uniquely recovered (up to scaling and permutation); see [64], also [288] for results in the multivariate and complex setting. Valderrama and Aguilera [291] further used Theorem 2 to establish a sufficient condition for Gaussianity of a random vector via its PCA.

Under mild assumptions on S, Theis [289], following [64], demonstrated the identifiability[1] of the ICA model under general settings without resorting to Theorem 2. This leads to the following theorem, which is derived within the framework of a noiseless IC model, with as many sources (or ICs) as observations.

Theorem 3 (ICA separability, Theis [289], Theorem 2)**.** *Let S be a p-variate independent random vector and $A \in \mathbb{R}^{p\times p}$ a nonsingular matrix. Assume one of the following hypothesis:*

1. *S has, at most, one Gaussian or deterministic component[2] and admits second order moments.*
2. *S has non-Gaussian components, and its probability density function P_S exists and is twice continuously differentiable.*

Then, if AS is again independent, the matrix A is equivalent to the identity.

Proof. First, we provide some definitions. For clarity, we will follow the notation used by the author.

A function $f : \mathbb{R}^p \to \mathbb{C}$ (note that $\mathbb{R} \subset \mathbb{C}$) is called separated/linearly separated if there exists a set of one dimensional functions $g_1, \dots, g_p : \mathbb{R} \to \mathbb{C}$ such that $f(x) = g_1(x_1) \cdots g_p(x_p)$, for all $x \in \mathbb{R}^p$. For separated f, we use the tensor product, i.e., $f \equiv g_1 \otimes \cdots \otimes g_p$ in the same vein as in [289]. Further, denote by $\mathrm{C}^r(U; V)$ the ring of continuously differentiable functions of order r, from $U \subset \mathbb{R}^p$ to $V \subset \mathbb{C}$, with U open.

For the first hypothesis, the author starts showing that due to a pre-whitening step, A can be assumed without loss of generality orthonormal. Therefore A is the product of a permutation and scaling matrix. Then, the proof builds upon the fact that among all densities and characteristic functions, the Gaussians satisfy the differential equation

$$af^2 - ff'' + f'^2 \equiv 0, \qquad f \in \mathrm{C}^2(\mathbb{R}; \mathbb{C}), a \in \mathbb{C}, \tag{2.2}$$

whose solutions are $f \equiv 0$ or $f(u) = \exp\left(\frac{a}{2}u^2 + bu + c\right)$, with $u \in \mathbb{R}$ and $b, c \in \mathbb{C}$ (see [289, Lemma 3]). Furthermore, the author presents the following Lemma to complete the proof.

Lemma 3 (Theis [289], Lemma 4)**.** *Let $g_i \in \mathrm{C}^2(\mathbb{R}; \mathbb{C})$ and B a $p \times p$ invertible matrix such that for $x \in \mathbb{R}^p$ the factorization $f(x) \equiv g_1 \otimes \cdots \otimes g_p(Bx)$ holds. Then for all indices ℓ and $i \neq j$ with $B_{\ell i}B_{\ell j} \neq 0$, g_ℓ satisfies 2.2 for some constant a.*

Now, set $X = AS$, and let $x, s \in \mathbb{R}^p$. Further, let $\widehat{S}(s) \equiv \mathbb{E}_S(\exp is^\top S)$ be the characteristic function of S. If S is independent, then $\widehat{S} \equiv g_1 \otimes \cdots \otimes g_p \in \mathrm{C}^2(\mathbb{R}^p; \mathbb{C})$, where $g_i \equiv \widehat{S}_i$. The characteristic function of AS can be calculated as

$$\widehat{AS}(x) = \mathbb{E}_S(\exp ix^\top AS) = \widehat{S}(A^\top x) = g_1 \otimes \cdots \otimes g_p(A^\top x). \tag{2.3}$$

If we let $B \equiv (B_{ij}) = A^\top$, given that AS is also independent, we have $f(x) \equiv \widehat{AS}(x) = g_1 \otimes \cdots \otimes g_p(Bx)$. By assuming $A \not\sim I_p$ and noting that $B = A^\top$, there exists $k \neq \ell$ and $i \neq j$ with $B_{ki}B_{kj} \neq 0$ and $B_{\ell i}B_{\ell j} \neq 0$, such that both g_k and g_ℓ satisfy 2.2 according to Lemma 3. This contradicts the first hypothesis, as it shows that S_k and S_ℓ are both Gaussian (see also [289, Corollary 2]).

For the second hypothesis, let $P_S \in \mathrm{C}^2(\mathbb{R}^p; \mathbb{R})$ be the density of S. Since S is assumed independent (with no Gaussian component), $P_S \equiv g_1 \otimes \cdots \otimes g_p$ with $g_i \equiv P_{S_i}$. Then, the density of AS is

$$P_{AS}(x) = |\det A|^{-1} P_S\left(A^{-1}x\right) = |\det A|^{-1} g_1 \otimes \cdots \otimes g_p(A^{-1}x), \tag{2.4}$$

where $|\det A|^{-1}$ is a scaling factor which ensures that the total probability integrates to 1. Given that AS is also independent, we can define $f(x) \equiv |\det A| P_{AS}(x) = g_1 \otimes \cdots \otimes g_p(Bx)$ by letting $B \equiv (B_{ij}) = A^{-1}$.

If, once more, we assume that $A \not\sim I_p$, then $B = A^{-1} \not\sim I_p$. Hence, there exist indices ℓ and $i \neq j$ such that $B_{\ell i}B_{\ell j} \neq 0$. According to Lemma 3, g_ℓ satisfies the differential equation 2.2. This means that the ℓth component of S is Gaussian, leading to a contradiction (see also [289, Corollary 1]). ∎

[1] That is, the possibility of determining the mixing matrix A.

[2] S is said to have a Gaussian component if one S_j is Gaussian, i.e., $P_{S_j}(x) = d\exp(-ax^2 + bx + c)$ with $(a, b, c, d \in \mathbb{R}, a > 0)$, while deterministic means that the component is constant.

Indeed, Theorem 3 confirms the separability of the linear IC model: if $X = AS$ and W is a demixing matrix such that WX is independent, then $WA \sim I_p$, implying $W^{-1} \sim A$. In view of the current results, Nordhausen and Oja [217] proposed a more precise characterization of the ICA model, which intuitively draws some analogies with the definition of independence provided in §1.3.

Definition 4 (ICA model, Nordhausen and Oja [217], §3.1)**.** *Let X be a centered p-variate random vector. Then, X follows an IC model if*

$$X = AS = A_1 S_1 + A_2 S_2, \tag{2.5}$$

where $A = (A_1, A_2) \in \mathbb{R}^{p\times p}$ is nonsingular with $A_1 \in \mathbb{R}^{p\times q}, A_2 \in \mathbb{R}^{p\times(p-q)}$. The random vector $S = (S_1^\top, S_2^\top)^\top$ verifies (i) $\operatorname{var}(S_i) = 1$ *$(i = 1, \dots, p)$, (ii) $S_1 \perp\!\!\!\perp S_2$ (iii) the components of S_1 are independent non-Gaussian and of S_2, independent Gaussian.*

As noted in [217], it is often presumed that $q = p-1$, which implies at most one component is Gaussian, as discussed in Theorem 3. Meanwhile, the current model formulation appears to be more flexible, as it does not strictly require S_2 to have a specific dimension.

Now, let's shift our focus back to the Hilbert space setting discussed in Chapter 1. Extending the notion of separability to infinite dimensions brings forth several inherent challenges. Gutch and Theis [111] based on Theorem 3, provided the following result in this respect:

Theorem 4 (Gutch and Theis [111], Theorem 1)**.** *Let $(H, \langle\cdot,\cdot\rangle)$ be a separable Hilbert space and S a H-valued independent random variable whose characteristic function $\hat{S}(u) \equiv \mathbb{E}_S[\exp(i\langle S, u\rangle)], u \in H$, is twice differentiable. Suppose $A : H \to H$ is a linear operator with bounded inverse, and let $X = AS$ be again independent. If one considers an arbitrary orthonormal basis $(\varphi_j)_{j\in\mathbb{N}}$ of H and there exist indices $\ell \in \mathbb{N}$ and $j \neq k \in \mathbb{N}$ such that $\langle\varphi_j, A\varphi_\ell\rangle \neq 0 \neq \langle\varphi_k, A\varphi_\ell\rangle$, then the ℓth component of S has to be Gaussian.*

If the inverse of the operator A exists, then it is continuous and defined for all elements of H. This implies that A is both injective (one-to-one) and surjective (onto), meaning it establishes a bijective mapping between elements of the domain and codomain. This isomorphism is interpreted in the following sense: if $\langle\varphi_j, A\varphi_\ell\rangle \neq 0 \neq \langle\varphi_k, A\varphi_\ell\rangle$ (this operation describes specific non-orthogonal relationships with regard to an arbitrary orthonormal basis of H), then S must be Gaussian, and any mixing of the components of S into more than one component of X will be Gaussian too. This also follows from extensions of the Darmois–Skitovitch and Cramér's theorems; see, e.g., [129]. If none of the components of S are Gaussian but independent, then A maps each component of S to a single component of X, therefore $A = \mathrm{Id}$. For more details, see proof in [111].

Although general grounds for separability have been provided, those may be deemed somewhat artificial as the law of X may not allow for a linear transformation resulting in independent components. Therefore, a more explicit characterization of the problem is necessary since the model outlined becomes too restrictive. In the presence of numerous Gaussian sources, the non-Gaussian ones may be barely distinguishable. Conversely, when the number of Gaussian sources is limited to a few or one, their separation becomes optimal (assuming they are independent, although this is not always the case [266]). In a broad sense (take e.g. Definition 4), the objective is to maximize the non-Gaussianity of the sources by separating the ones that do not contribute to make the model fully identifiable. Although yet unexplored in functional settings, these issues have been previously studied, for example, in [131, 198]. Current notion of decomposition in ICs based on non-Gaussianity has also found a reasoning in the central limit theorem and in the framework of information geometry [57].

Taking into account the preceding factors, pre-whitening the data —essentially, ensuring linear independence by eliminating second-order correlations— is a common and useful step to enhance the identification of the ICs. However, extending this procedure to infinite-dimensional spaces carries certain problems due to the covariance operator having unbounded inverse. This is discussed in the following sections, and the functional ICA model will be presented afterwards.

§ 2.2. The Cameron-Martin space geometry

Next, we assume a common scenario in functional data where H is a separable space of real-valued functions on a closed interval $\mathcal{I} = [0, T], T > 0$ with inner product operator $\langle\cdot,\cdot\rangle : H \times H \to \mathbb{R}$ and norm $\|\cdot\| : H \to [0, \infty)$. Given a probability space $(\Omega, \mathcal{A}, P)$, a H-valued random functional variable is the mapping $X : \Omega \to H$ that is $\mathbb{B}_H$-measurable, where $\mathbb{B}_H$ is the σ-field generated by the class of all open subsets of H.

Suppose $\mathbb{E}\|X\|^2 < \infty$. Then, X has mean function $\mu = \mathbb{E}X$ and covariance operator $\mathscr{C}_X = \mathbb{E}\{(X-\mu)\otimes(X-\mu)\}$ admitting the spectral representation

$$\mathscr{C}_X = \sum_{j=1}^{\infty} \lambda_j (\gamma_j \otimes \gamma_j) = \sum_{j=1}^{\infty} \lambda_j \mathscr{P}_{\gamma_j}, \tag{2.6}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ is its set of non-negative eigenvalues converging to zero and $(\gamma_j)_{j\in\mathbb{N}}$ an orthonormal basis of corresponding eigenfunctions. We employ the notation $\mathscr{P}_{\gamma_j} = \gamma_j \otimes \gamma_j$ for the projection of H on to the one dimensional eigenspace spanned by γ_j. Unless otherwise stated, we assume that $\mathscr{C}_X$ has strictly positive eigenvalues, hence $\mathscr{C}_X$ is injective. Furthermore, as $\mathscr{C}_X$ is a self-adjoint positive operator, consider there exists the operator $\mathscr{C}_X^{1/2}$ such that $(\mathscr{C}_X^{1/2})^2 = \mathscr{C}_X$.

Definition 5 (Vidal and Aguilera [296], Definition 1)**.** *The whitening operator Ψ transforms a functional variable X into a new element $\mathbb{X} = \Psi(X-\mu)$ with zero mean and covariance operator being exactly the identity in H.*

We note that there is no convention in how mean-centering should be performed, if before or after the transformation. In the sequel it will be assumed that $\mu = 0$.

A natural way to produce a whitening operator is via factorization of the precision (the inverse of $\mathscr{C}_X$), which suggests the expression $\Psi^*\Psi = \mathscr{C}_X^{-1}$, where Ψ^* is the adjoint of Ψ. A priori, major drawbacks might arise in this context, as the precision operator turns out to be unbounded and, in general, X does not belong to its domain (see, for example, [191, §2.2]). One should therefore proceed with care, as even considering certain kinds of regularization, $\Psi(X)$ may not exist in the sense of Definition 5. We now discuss how Ψ can also be correctly validated to lead to a number of whitening operators in the functional setting through a particular space geometry.

The Cameron-Martin (CM) space can be defined as a certain completion of H with respect to the norm induced by a Gaussian measure. Originally, this space was associated with a Wiener measure due to the investigations conducted by Cameron and Martin [53] on its kernel. However, the name CM space has since been extended to encompass affine transformations of Wiener measures as well as other Gaussian distributions.

Following the factorization Theorem [179, §4.2], the usual definition of a CM space is as follows:

$$\mathbb{M} = \left\{ h \in H : \sum_{j=1}^{\infty} \frac{\langle h, \gamma_j \rangle^2}{\lambda_j} < \infty \right\} \quad \text{with} \quad \langle g, h \rangle_{\mathbb{M}} = \sum_{j=1}^{\infty} \frac{\langle g, \gamma_j \rangle \langle h, \gamma_j \rangle}{\lambda_j}. \tag{2.7}$$

Note the CM space is not only a Hilbert space but also a RKHS. The necessary and sufficient condition for an element in H to be in $\mathbb{M}$ is that satisfies $\sum_j \lambda_j^{-1} \langle h, \gamma_j \rangle^2 < \infty$. Casting the current law, often referred to as Picard's criterion [86, §2.2], is not simply a procedural step since $\mathbb{M}$ only becomes closed under the norm induced by the RKHS topology or finite dimensional constraints. Under Picard's criterion, we can use 2.7 to reparametrize the random variable X to have the identity as covariance operator in $\mathbb{M}$, and through it define a family of whitening transformations. For all $f, g \in \mathbb{M}$, note that the inner product in 2.7 can be written as $\langle g, h \rangle_{\mathbb{M}} = \langle \mathscr{C}_X^{1/2\dagger} f, \mathscr{C}_X^{1/2\dagger} g \rangle$, where $\mathscr{C}_X^{1/2\dagger}$ is the the Moore-Penrose inverse of $\mathscr{C}_X^{1/2}$ [125, §3.5.7].

§ 2.3. Optimal whitening transformations

Whitening operators

A whitening operator can be defined as a two-step transformation, with representation $\Psi : \mathbb{M} \to \mathbb{M}$, restricted to map elements in the range of $\mathscr{C}_X$. Therefore, the other part of the mapping comprises a projection on to the space generated by $(\gamma_j)_{j\in\mathbb{N}}$, so that X becomes entirely determined by the covariance operator $\mathscr{C}_X$ before whitening. Unless otherwise stated, in the following sections we reset X to the range space of $\mathscr{C}_X$, hence the proposed whitening operators map elements of the kind $\sum_{j=1}^{\infty} \langle \cdot, \gamma_j \rangle \gamma_j$.

The most typical whitening transformation is obtained via square root factorization of the precision operator $\mathscr{C}_X^{\dagger}$, which stands as a direct extension of the popular zero phase component analysis multivariate whitening [26]. The aforementioned operator is here denoted by $\Psi_{\gamma\otimes\gamma} = \mathscr{C}_X^{1/2\dagger}$, and its spectral decomposition is straightforwardly written as

$$\Psi_{\gamma\otimes\gamma} = \mathscr{C}_X^{1/2\dagger} = \sum_{j=1}^{\infty} \lambda_j^{-1/2} (\gamma_j \otimes \gamma_j). \tag{2.8}$$

Now, we describe some properties of this mapping.

Proposition 2 (Vidal and Aguilera [296], Proposition 1). *The covariance operator of* $\mathbb{X} = \mathcal{U}\Psi_{\gamma\otimes\gamma}(X)$ *satisfies the identity in* $\mathbb{M}$ *for any unitary transformation* $\mathcal{U}$ *in* $\overline{\text{ran}}(\mathscr{C}_X^{1/2})$*, the closure of the range space of* $\mathscr{C}_X^{1/2}$.

According to the above proposition, rotational freedom becomes apparent, which leads us to the possibility of defining a family of whitening operators, as described below.

A slight modification in 2.8 alternatively yields the non-symmetric whitening operator

$$\Psi_{\gamma\otimes e} = \sum_{j=1}^{\infty} \lambda_j^{-1/2}(\gamma_j \otimes e_j), \tag{2.9}$$

where $(e_j)_{j\in\mathbb{N}}$ is a fixed orthonormal basis of H. In fact, here we see the role of operator $\mathcal{U}$ as the agent of sending γ_j to e_j. Current operator follows the principles of [98], which only considers a single rotation of the covariance matrix eigenvectors.

To further extend the class of whitening operators, one can consider a succinct form of decorrelation by defining the diagonal operator $\mathcal{V} = \text{diag}(\mathscr{C}_X) = \sum_{k=1}^{\infty} \mathcal{P}_{e_k}\mathscr{C}_X\mathcal{P}_{e_k}$, where $\mathcal{P}_{e_k} = (e_k \otimes e_k)$. The operator $\mathcal{V}$ is not unique as it depends on an arbitrary orthonormal basis of H. In other words, there is no privileged orthonormal basis on H to define $\mathcal{V}$, and for each one of them different operators can be obtained. This way, the notion of standardization in multivariate analysis can be extended to the functional case by the operator $\mathcal{V}^{1/2\dagger}$, where $\mathcal{V}^{1/2\dagger}$ is the Moore Penrose inverse of $\mathcal{V}^{1/2}$. Further usefulness of this operator will become clear in the following sections.

Two whitening procedures with appealing properties were defined in [147] by constraining the arbitrariness of the transformation to inherent autocorrelations. Next, we will suppose that $\mathcal{R} : \mathbb{M} \to \mathbb{M}$ defined as $\mathcal{R} = \mathcal{V}^{1/2\dagger}\mathscr{C}_X\mathcal{V}^{1/2\dagger}$ is a compact operator, boundedly invertible, with associated singular system $(\rho_j, \varphi_j)_{j\in\mathbb{N}}$. The operator $\mathcal{R}$ closely resembles the usual correlation matrix and its spectral decomposition leads to an eigenspace that will be of use in combination to $\mathcal{V}^{1/2\dagger}$. Thus, there is no loss of generality in assuming that

$$\Psi_{\varphi\otimes\varphi} = \mathcal{R}^{1/2\dagger}\mathcal{V}^{1/2\dagger} = \left\{\sum_{j=1}^{\infty} \rho_j^{-1/2}(\varphi_j \otimes \varphi_j)\right\}\mathcal{V}^{1/2\dagger}, \tag{2.10}$$

or analogously to 2.9,

$$\Psi_{\varphi\otimes e} = \left\{\sum_{j=1}^{\infty} \rho_j^{-1/2}(\varphi_j \otimes e_j)\right\}\mathcal{V}^{1/2\dagger} \tag{2.11}$$

satisfies the usual properties of a whitening transformation in the sense of Definition 5. Both operators decline the span of $\mathscr{C}_X$ by its diagonal, merging it with the spectral decomposition of $\mathcal{R}$, which leads to a non-symmetric operator. In turn, the operator $\Psi_{\varphi\otimes\varphi}$ is up to permutation or sign changes but also invariant under unitary transformations within the subspace spanned by the eigenvectors of $\mathcal{R}$. Again, we can write $\mathcal{U}\Psi_{\varphi\otimes\varphi}$, where $\mathcal{U}$ denotes a unitary operator in the range of $\mathcal{R}^{1/2\dagger}\mathcal{V}^{1/2\dagger}$.

Triangular factorization of self-adjoint and positive operators, might provide us another whitening procedure closely related to the Cholesky decomposition of the precision matrix. Due to a result of Krein (see Theorem 3.4.5 in [20]), the usual precision operator can be factored as

$$(I_H + \mathscr{C}_X)^{-1} = (I_H - \Delta^*)(I_H - \Delta), \tag{2.12}$$

where Δ is a triangular Volterra operator on $H = L^2_{\mathcal{J}}$ and Δ^* its adjoint. The factorization in 2.12 leads to the whitening operator $\Psi_\Delta = (I_H - \Delta)$. Nevertheless, we will restrict the analyses to a common Cholesky decomposition of the precision operator in the finite dimensional setting. Optimal whitening with triangular operators presents further technical difficulties requiring a separate study, in part, because of the great variety of them.

Optimal functional whitening

Similarly to the multivariate case, optimality in a functional whitening transformation can be identified in two different ways. The first corresponds to a problem aiming to find a component wise transformation that is closer to the original functional variable using some measure of adjustment or resemblance. The second is related to the ability of the whitening operator to compress the original functional variable and retain the maximum information content.

The adjustment between the original projected and whitened functional variable is measured by the minimization of

$$\mathbb{E}\|X - \mathbb{X}\|^2 = \operatorname{tr}(\mathscr{C}_X) - 2\operatorname{tr}(\mathscr{C}_{X\mathbb{X}}) + \operatorname{tr}(\mathscr{C}_{\mathbb{X}}), \tag{2.13}$$

where $\mathscr{C}_{X\mathbb{X}} = \mathbb{E}(X \otimes \mathbb{X})$ is the cross-covariance operator between X and $\mathbb{X}$. As $2\operatorname{tr}(\mathscr{C}_{X\mathbb{X}})$ is the only dependent between the original and the whitened variable, the minimization problem can be reduced to the maximization of $\operatorname{tr}(\mathscr{C}_{X\mathbb{X}})$. We implicitly assume that $\mathbb{X}$ falls in a space with the usual inner product and that X lies in the closure of $\mathbb{M}$.

Proposition 3 (Vidal and Aguilera [296], Proposition 2)**.** *The whitening operator $\Psi_{\gamma\otimes\gamma}$ is the unique transformation that minimizes $\mathbb{E}\|X - \mathbb{X}\|^2$.*

The least-squares problem in Proposition 3, however, is restrictive in the sense that it only allows to quantify the goodness of fit of the whitening transformation without not being further explanatory of correlations or level of compression. For a correlation-based similarity objective, a scale-invariant measure is usually required. Here, we consider a functional extension of the criteria used in [147], consisting of the minimization of the mean squared error between the standardized functional variable and the whitened one. The operator $\mathscr{V}^{1/2^\dagger}$ scales the original variable without removing correlations, allowing to construct a scale invariant measure without being necessary to compute derivatives. The optimality objective is then expressed as

$$\mathbb{E}\|\mathscr{V}^{1/2^\dagger}(X) - \mathbb{X}\|^2 = \operatorname{tr}\{\mathscr{C}_{\mathscr{V}^{1/2^\dagger}(X)}\} - 2\operatorname{tr}\{\mathscr{C}_{\mathscr{V}^{1/2^\dagger}(X)\mathbb{X}}\} + \operatorname{tr}(\mathscr{C}_{\mathbb{X}}), \tag{2.14}$$

which corresponds to the maximization of $\operatorname{tr}\{\mathscr{C}_{\mathscr{V}^{1/2^\dagger}(X)\mathbb{X}}\}$.

Proposition 4 (Vidal and Aguilera [296], Proposition 3)**.** *The whitening operator $\Psi_{\varphi\otimes\varphi}$ is the unique transformation that minimizes $\mathbb{E}\|\mathscr{V}^{1/2^\dagger}(X) - \mathbb{X}\|^2$.*

Robustness in small local changes is not necessarily guaranteed when a whitening transformation is based on minimal least squared adjustment. To measure the degree of compression of a whitening transformation, Kessy et al. [147] used the row sum of squared cross-covariance and cross-correlations between the components of the whitened and the original random vector. Then, a monotonically decreasing condition on the resultant variance is established to be maximized. As one might suspect, a similar approximation can be developed in the functional data setting.

First, note that the operators $\mathscr{C}_{X\mathbb{X}}$ and $\mathscr{C}_{\mathbb{X}X}$ are not self-adjoint, whereas $\mathscr{C}_{X\mathbb{X}}$ is the adjoint of $\mathscr{C}_{\mathbb{X}X}$. Define then the compound operator $\mathscr{C}_{X\mathbb{X}} \circ \mathscr{C}_{X\mathbb{X}}$, which is self-adjoint and compact.

Formally, a way to measure how the whitening operator effectively compresses the original functional variable in terms of a cross-covariance relation is by

$$\sigma_{\text{cov}} = \sup_{\|e_k\|=1} \langle e_k, \mathscr{C}_{\mathbb{X}X} \circ \mathscr{C}_{X\mathbb{X}}(e_k)\rangle\,. \tag{2.15}$$

We can similarly proceed for the cross-correlation operator, now defined in the same sense of [163] as $\mathscr{R}_{X\mathbb{X}} = \mathscr{C}_X^{1/2^\dagger}\mathscr{C}_{X\mathbb{X}}\tilde{\mathscr{C}}_X^{1/2^\dagger}$, where $\tilde{\mathscr{C}}_X = \mathbb{E}(\mathbb{X} \otimes \mathbb{X})$. Thus, if the aim is to maximize the compression under a cross-correlation measure, we look for the maximization of the rate

$$\sigma_{\text{corr}} = \sup_{\|e_k\|=1} \langle e_k, \mathscr{R}_{\mathbb{X}X} \circ \mathscr{R}_{X\mathbb{X}}(e_k)\rangle\,. \tag{2.16}$$

Simulations conducted in [296] show that the whitening operators $\Psi_{\gamma\otimes e}$ and $\Psi_{\varphi\otimes e}$ maximize the proposed compression rates.

Finite dimensional approximation

Let X_i $(i = 1, \ldots, n)$ be n independent copies of X not directly observable. The curves X_i are reconstructed from a vector of measurements collected in a finite set of equidistant time points $t_{i0}, t_{i1}, \ldots, t_{im_i}$, contaminated with additive independent errors, i.e., $X_{ik} = X_i(t_{ik}) + \epsilon_{ik}$ $(k = 0, \ldots, m_i)$. Suppose the observations are assumed in a q-dimensional space $H^{[q]}$ of H spanned by a collection of basis functions $\phi = (\phi_1, \ldots, \phi_q)^\top$ not necessarily orthonormal in the usual sense. For two functions $f = \phi^\top \mathtt{f}$ and $g = \phi^\top \mathtt{g}$, the inner product is defined by $\langle f, g\rangle = \mathtt{f}^\top \mathscr{G} \mathtt{g}$, where $\mathscr{G} \in \mathbb{R}^{q\times q}$ is the Gram matrix of inner products between pairs of basis functions. Then, X_i can be expressed as the vector valued function $X(t) = A\phi(t)$, where $A \in \mathbb{R}^{n\times q}$ is a matrix of coefficients and $\phi(t) = (\phi_1(t), \ldots \phi_q(t))^\top$.

The q-dimensional sample covariance operator $\mathscr{C}_X^{[q]}$ is defined for any $f \in H^{[q]}$ as

$$\mathscr{C}_X^{[q]}(f) = \langle C^{[q]}(s, \cdot), f\rangle \tag{2.17}$$

where $C^{[q]}$ is the covariance kernel function of X admitting the following matrix representation

$$\begin{aligned} C^{[q]}(s,t) &= n^{-1}X(s)^\top X(t) \\ &= \phi(s)^\top \mathcal{G}^{-1/2}(n^{-1}\mathcal{G}^{1/2}A^\top A\mathcal{G}^{1/2})\mathcal{G}^{-1/2}\phi(t) \\ &= \hat{\phi}(s)^\top \Sigma_{A\mathcal{G}^{1/2}}\hat{\phi}(t). \end{aligned} \tag{2.18}$$

Then, the coordinates of $C^{[q]}$ can be expressed in terms of an orthonormalized basis $\hat{\phi}(t) = \mathcal{G}^{-1/2}\phi(t)$ as $n^{-1}\mathcal{G}^{1/2}A^\top A\mathcal{G}^{1/2} = \Sigma_{A\mathcal{G}^{1/2}}$. The matrix $\Sigma_{A\mathcal{G}^{1/2}} \in \mathbb{R}^{q\times q}$ has the eigendecomposition $\Sigma_{A\mathcal{G}^{1/2}} = U\Lambda U^\top$ where U are eigenvectors and Λ is a diagonal matrix with entries the eigenvalues of $\Sigma_{A\mathcal{G}^{1/2}}$. The eigenfunctions of $\mathcal{C}_X^{[q]}$ are then defined as $\gamma(t) = B\phi(t)$ with $B = \mathcal{G}^{-1/2}U$. Further, consider the decomposition $\Sigma_{A\mathcal{G}^{1/2}} = D^{1/2}R_{A\mathcal{G}^{1/2}}D^{1/2}$, where $D = \operatorname{diag}(\Sigma_{A\mathcal{G}^{1/2}})$ and $R_{A\mathcal{G}^{1/2}}$ is a matrix capturing the correlations of $\Sigma_{A\mathcal{G}^{1/2}}$ with eigendecomposition $R_{A\mathcal{G}^{1/2}} = V\Theta V^\top$. With this, we next derive the whitening procedures described in the last subsection. Normalization is omitted for the sake of clarity.

Proposition 5 (Vidal and Aguilera [296], Proposition 4)**.** *Let us consider the orthonormalized basis $e(t) = \hat{\phi}(t)$ of $H^{[q]}$. Using the functional representation $X(t) = A\mathcal{G}^{1/2}\hat{\phi}_j(t)$, the coefficients of each functional whitening operator are obtained by their respective multivariate whitening procedures of the orthonormalized coefficient matrix $A\mathcal{G}^{1/2}$ as follows:*

$$\begin{aligned} \Psi_{\gamma\otimes\gamma}\{X(t)\} &= (A\mathcal{G}^{1/2})\Sigma^{-1/2}_{A\mathcal{G}^{1/2}}\hat{\phi}(t), \\ \Psi_{\gamma\otimes e}\{X(t)\} &= (A\mathcal{G}^{1/2})U\Lambda^{-1/2}\hat{\phi}(t), \\ \Psi_{\varphi\otimes\varphi}\{X(t)\} &= (A\mathcal{G}^{1/2})D^{-1/2}R^{-1/2}_{A\mathcal{G}^{1/2}}\hat{\phi}(t), \\ \Psi_{\varphi\otimes e}\{X(t)\} &= (A\mathcal{G}^{1/2})D^{-1/2}V\Theta^{-1/2}\hat{\phi}(t), \\ \Psi_{\Delta}\{X(t)\} &= (A\mathcal{G}^{1/2})L\hat{\phi}(t), \end{aligned} \tag{2.19}$$

where L is the solution to the Cholesky factorization $LL^\top = \Sigma^{-1}_{A\mathcal{G}^{1/2}}$.

§ 2.4. Functional ICA model based on kurtosis

Kurtosis extremization is a common problem addressed in the context of ICA, which is often referred to by the name of fourth-order blind identification (FOBI); see, for example, [218]. While multivariate ICA methods are based on various kinds of non-Gaussian optimization, functional ICA has primarily been developed using FOBI or related families of estimators. This limitation is probably due to the fact that the concept of a density function is not straightforwardly defined in functional data spaces. Furthermore, in the absence of a parametric form of the component distribution, one strongly relies on the topological features of the data. Current functional ICA models are then defined through functional PC reduction [176, 197, 305], or penalized versions of it to exploit and regulate the level of smoothness [303] (see, §2.6). These approaches assume that the interesting ICs lie in the subspace spanned by the first functional principal factors. However, some ICs could be mixed with the discarded PCs, and there is also little room remaining to investigate other non-linear structures underlying the original data because the aforementioned models are constrained by the linearity of the PCs. To address this, a more general model is proposed in §2.5. It should be stressed that current functional IC models exploit temporal independence, akin to early developments in ICA for time series analysis [60, 225].

As has been previously outlined in [303], functional ICA can be summarized in the following steps:

- Whitening the functional random variable;
- space rotation via scatter operators and spectral decomposition;
- projection on to the space generated by the operator's eigenfunctions and expansion.

Typically, this procedure is described through a demixing and mixing transformation, the latter mapping the Hilbert-valued functional independent components into the original space of functions.

The model

The functional ICA aims to enhance the estimation of independent components via orthogonal rotations of $\mathbb{X}$. Let $U(\mathbb{M})$ denote the class of all unitary operators in $\mathcal{B}(\mathbb{M})$. The functional IC model can be expressed as

$$\Gamma(X) = \mathcal{U}\Psi(\mathcal{P}_\gamma X) = Z, \tag{2.20}$$

where Γ is commonly known as the demixing operator, $\mathcal{U} \in U(\mathbb{M})$ and Z is a H-valued element with independent component functions satisfying $\mathscr{C}_Z = \mathscr{P}_{\overline{\text{ran}}(\Psi)}$. Another way to see model 2.20 is by means of a mixing operator $A \in U(\mathbb{M})$ which corresponds to

$$X = AZ + (I_H - \mathscr{P}_\gamma)(X), \tag{2.21}$$

where $A = \Psi^{-1}\mathcal{U}$. Note that $\mathscr{P}_\gamma$ is the projection operator on to the span of the γ_j's. A key question in functional ICA is henceforth how to determine $\mathcal{U}$. We thus present the properties of the kurtosis operator, previously defined in [231], whose spectral decomposition will provide us the operator of interest.

Assumption 1. $\mathbb{E}\|X\|^4 < \infty$.

Under the above assumption, we can establish the existence of a mapping $\mathcal{K} \in \mathscr{B}_{HS}(\mathbb{M})$ with the action $\mathbb{X} \mapsto \mathcal{K}_\mathbb{X}$ defined as $\mathbb{E}\{(\mathbb{X} \otimes \mathbb{X})^2\}$. This mapping is referred to as the kurtosis operator of $\mathbb{X}$ and their properties are derived below.

Proposition 6. *The operator $\mathcal{K}_\mathbb{X}$ is self-adjoint, positive definite with finite trace. From the Hilbert-Schmidt boundeness, $\mathcal{K}_\mathbb{X}$ admits the spectral decomposition*

$$\mathcal{K}_\mathbb{X} = \sum_{j=1}^{\infty} \kappa_j \psi_j \otimes \psi_j, \tag{2.22}$$

where $(\kappa_j)_{j\in\mathbb{N}}$ is a sequence of strictly positive eigenvalues in decreasing order and $(\psi_j)_{j\in\mathbb{N}}$ the associated eigenfunctions. The operator $\mathcal{K}_\mathbb{X}$ is unitary equivariant with respect to an unitary operator $\mathcal{U} \in U(\mathbb{M})$, that is, $\mathcal{K}(\mathcal{U}\mathbb{X} + b) = \mathcal{U}\mathcal{K}_\mathbb{X}\mathcal{U}^*$, for each $b \in \mathbb{M}$.

Proof. For any $f \in \mathbb{M}$ one can verify that

$$\langle \mathcal{K}_\mathbb{X} f, f\rangle = \mathbb{E}\left\langle (\mathbb{X}\otimes\mathbb{X})^2 f, f\right\rangle = \mathbb{E}\left(\langle \mathbb{X}, \mathbb{X}\rangle \langle \mathbb{X}, f\rangle^2\right) = \mathbb{E}\left(\|\mathbb{X}\|^2 \langle \mathbb{X}, f\rangle^2\right) \geq 0, \tag{2.23}$$

so clearly, $\mathcal{K}(\mathbb{X})$ is nonnegative definite. Self-adjointness becomes apparent since, per definition, $\mathcal{K}_\mathbb{X} = \mathbb{E}\{\|\mathbb{X}\|^2(\mathbb{X}\otimes\mathbb{X})\}$. Therefore $\mathcal{K}_\mathbb{X} = \mathcal{K}_\mathbb{X}^*$, where $\mathcal{K}_\mathbb{X}^*$ is the adjoint of $\mathcal{K}_\mathbb{X}$, and consequently $\langle \mathcal{K}_\mathbb{X} f, g\rangle = \langle f, \mathcal{K}_\mathbb{X} g\rangle$ with $f, g \in \mathbb{M}$. Self-adjoint operators are necessarily normal. To show that $\mathcal{K}(\mathbb{X})$ is trace class, consider any orthonormal basis $\{e_1, e_2, \ldots\}$ of $\mathbb{M}$. Similarly as in (2.23) observe that

$$\text{tr}(\mathcal{K}_\mathbb{X}) = \sum_{j=1}^{\infty} \left\langle \mathcal{K}_\mathbb{X} e_j, e_j\right\rangle = \sum_{j=1}^{\infty} \mathbb{E}\left(\|\mathbb{X}\|^2 \left\langle \mathbb{X}, e_j\right\rangle^2\right) = \mathbb{E}\|\mathbb{X}\|^4.$$

Furthermore, by our assumptions $\mathbb{E}\|X\|^4$ is finite, and so it is $\mathbb{E}\|\mathbb{X}\|^4$ when Picard's condition holds. According to the spectral theorem, operators satisfying the above properties are diagonalizable.

To conclude, consider a unitary operator $\mathcal{U} \in \mathscr{B}(\mathbb{M})$ and a location function $b \in \mathbb{M}$. Then, we can write

$$\mathcal{K}(\mathcal{U}\mathbb{X} + b) = \mathbb{E}\{(\mathcal{U}\mathbb{X}\otimes\mathcal{U}\mathbb{X})^2\} = \mathbb{E}\{\|\mathcal{U}\mathbb{X}\|^2(\mathcal{U}\mathbb{X}\otimes\mathcal{U}\mathbb{X})\}.$$

Taking separately both factors of the last expression, we have that $\|\mathcal{U}\mathbb{X}\|^2 = \langle \mathbb{X}, \mathcal{U}^*\mathcal{U}\mathbb{X}\rangle = \|\mathbb{X}\|^2$ and $\mathcal{U}\mathbb{X}\otimes\mathcal{U}\mathbb{X} = \mathcal{U}(\mathbb{X}\otimes\mathbb{X})\mathcal{U}^*$, which follows from the tensor product properties. As a result, $\mathcal{K}(\mathcal{U}\mathbb{X}) = \mathcal{U}\mathcal{K}(\mathbb{X})\mathcal{U}^*$, showing that unitary equivariance holds. ∎

Note that the kurtosis operator naturally fosters the properties of the covariance operator. This operator is uniquely determined by the relation

$$\mathbb{E}\left(\langle f, \mathbb{X}\rangle \langle \mathbb{X}, f\rangle \langle \mathbb{X}, \mathbb{X}\rangle\right) = \langle f, \mathcal{K}_\mathbb{X} f\rangle, \tag{2.24}$$

which holds for all $f \in \mathbb{M}$.

The independent components are of the form $\langle \mathbb{X}, \psi_j\rangle$, where the orthonormal family $(\psi_j)_{j\in\mathbb{N}}$, called independent component weight functions, are obtained by solving $\psi_j = \text{argmax}_f \text{kurt}(\langle \mathbb{X}, f\rangle)$ subject to $\|f\|^2 = 1, \langle f, \psi_j\rangle = 0$. This way, the kurtosis based functional ICA is determined by the solutions to the eigenproblem

$$\mathcal{K}_\mathbb{X}(\psi_j) = \kappa_j \psi_j. \tag{2.25}$$

The independent component scores $\tilde{\xi}_j = \langle \mathbb{X}, \psi_j \rangle$ are then generalized linear combinations of $\mathbb{X}$ with maximum kurtosis satisfying $\operatorname{kurt}(\tilde{\xi}_j) = \langle \mathcal{K}_{\mathbb{X}} \psi_j, \psi_j \rangle = \kappa_j$. It is easy to see that the model 2.20 is completed by defining $\mathcal{U} = \sum_{j=1}^{\infty} (\psi_j \otimes \psi_j)$.

Now, assume that κ_j and ψ_j satisfy equation 2.25. Then, $\lim_{M \to \infty} \sum_{j=1}^{M} \tilde{\xi}_j \psi_j$ converges uniformly to $\mathbb{X}$, which is a direct consequence of Mercer's Theorem for a symmetric positive-definite operator. Since the correlation structure of X is removed in $\mathbb{X}$, note that the additive sum of projections will be rather changeable if compared with the usual KL expansion.

§ 2.5. Kurtosis based smoothed functional ICA

We next extend the method introduced by Silverman [276] for smoothing principal component estimates to functional ICA by introducing the notion of penalized kurtosis. In doing so we aim at controlling the smoothness of the IC weight functions by applying roughness penalties through a second order linear differential operator. The small perturbations produced by this type of penalty will presumably have a regularizing effect on the lower order eigenelements by adjusting possible distortions.

In what follows, we shall consider that $\mathbb{M}_\theta$ is a closed subspace of continuously differentiable functions with weighted Sobolev inner product and corresponding norm

$$\langle f, g \rangle_\theta = \langle f, g \rangle + \theta \langle Rf, Rg \rangle, \quad \|f\|_\theta^2 = \langle f, f \rangle_\theta, \tag{2.26}$$

where $\theta \in \mathbb{R}_{\geq 0}$ is a penalty parameter and R, a bounded self-adjoint differential operator on $\mathbb{M}_\theta$ with the action $f \mapsto f''$. Note that there can exist many smooth eigenspaces as values are given to θ and if $\theta = 0$, then $\mathbb{M}_\theta = \mathbb{M}$.

The spectral decomposition of $\mathcal{K}_{\mathbb{X}}$ prevails in our model but now imposing orthonormality in terms of $\langle \cdot, \cdot \rangle_\theta$. Analogously to Silverman's method, the novel penalized IC approach maximizes

$$\frac{\operatorname{kurt}\langle f, \mathbb{X} \rangle}{\langle f, f \rangle + \theta \langle Rf, Rf \rangle} = \frac{\langle f, \mathcal{K}_{\mathbb{X}} f \rangle}{\|f\|_\theta^2}, \tag{2.27}$$

for all $f \in \mathbb{M}_\theta, f \neq 0$. Note that θ controls the roughness of the function f as measured by the penalty $\langle Rf, Rf \rangle$. Consequently, one can find a collection of smoothed functions $\psi_{\theta,j} \in \mathbb{M}_\theta$ that maximize 2.27 which is equivalent to solve the following optimization problem:

$$\begin{aligned} &\psi_{\theta,1} = \operatorname{argmax}_f \langle f, \mathcal{K}_{\mathbb{X}} f \rangle \text{ s. t. } \|f\|_\theta^2 = 1, \\ &\psi_{\theta,k} = \operatorname{argmax}_f \langle f, \mathcal{K}_{\mathbb{X}} f \rangle \text{ s. t. } \|f\|_\theta^2 = 1, \langle f, \psi_{\theta,j} \rangle_\theta = 0, \text{ for all } j < k, \ (k = 2, 3, \ldots). \end{aligned} \tag{2.28}$$

The orthonormal condition over the smoothed IC weight functions is fixed by the inner product $\langle \cdot, \cdot \rangle_\theta$ whereas the kurtosis of the independent components is given by $\langle \cdot, \cdot \rangle$. In this sense, the smoothed IC weight functions form an orthonormal system of $\mathbb{M}_\theta$.

Under the assumption of finite fourth moments and if Picard's law holds, for any $f \in \mathbb{M}_\theta$, $\theta \geq 0$, the solutions to the optimization problem 2.28 are given by the equation

$$\langle f, \mathcal{K}_{\mathbb{X}} \psi_{\theta,j} \rangle = \kappa_{\theta,j} \langle \psi_{\theta,j}, f \rangle_\theta. \tag{2.29}$$

Proposition 3.1 establishes in same sense as Theorem 3.1 in [242] that the solutions to 2.28 exist almost surely; the details provided there can be succinctly deduced for the present case. In order to get the main results given in the next section, consider from Proposition 4.3 in [223] for the covariance operator (functional PCA) that, due to a continuous assumption on the usual inner product $\langle \cdot, \cdot \rangle$ for the new inner product $\langle \cdot, \cdot \rangle_\theta$, there exists a positive definite, symmetric, continuous and bounded operator S^2 such that $\langle f, g \rangle = \langle S^2(f), g \rangle_\theta$ (see [242] for a rigorous definition of the operator S).

Proposition 7. *The eigensystem of the smoothed functional ICA, obtained as the solutions to Equation 2.29 and denoted by $(\psi_{\theta,j}, \kappa_{\theta,j}) \in \mathbb{M}_\theta \times \mathbb{R}$, is equivalent to the eigensystem of $\mathcal{K}_{S^2(\mathbb{X})}$ with the inner product $\langle \cdot, \cdot \rangle_\theta$.*

Proof. This result is immediate from the relation between the two inner products $\langle X, f \rangle = \langle S^2(X), f \rangle_\theta$. As a consequence, the optimization problems 2.25 and 2.28 are the same. ∎

Equivalently to the results provided in [223] for the smoothed functional PCA, this algorithm can be regarded as an equivalence between the smoothed functional ICA and the spectral decomposition of the kurtosis operator of the half-smoothed process $S(\mathbb{X})$ with the usual inner product.

Proposition 8. *$(\psi_\theta, \kappa_\theta) \in \mathbb{M}_\theta \times \mathbb{R}$ is an eigenelement of $\mathcal{K}_{S^2(\mathbb{X})}$ with $\langle\cdot,\cdot\rangle_\theta$ iff $(S^{-1}(\psi_\theta), \kappa_\theta)$ is an eigenelement of $\mathcal{K}_{S(\mathbb{X})}$ with $\langle\cdot,\cdot\rangle$.*

Proof. Since the operator S^2 is a symmetric positive definite operator, then there exists its square root operator, S which is also a symmetric positive definite operator verifying $S^2 = SS$, so that

$$\langle f, g\rangle = \langle S^2(f), g\rangle_\theta = \langle S(f), S(g)\rangle_\theta\,. \tag{2.30}$$

Let us denote by $\mathcal{K}_{S^2(\mathbb{X})}$ the kurtosis operator of the random variable $S^2(\mathbb{X})$ with the inner product $\langle\cdot,\cdot\rangle_\theta$ and by $\mathcal{K}_{S(\mathbb{X})}$ the kurtosis operator of $S(\mathbb{X})$ with $\langle\cdot,\cdot\rangle$.

The equivalence between the two eigensystems is clearly deduced from the following relationship between both kurtosis operators:

$$\mathcal{K}_{S^2(\mathbb{X})} = S\mathcal{K}_{S(\mathbb{X})}S^{-1}. \tag{2.31}$$

In fact,

$$\begin{aligned}\mathcal{K}_{S^2(\mathbb{X})}(f) &= \mathbb{E}\{\langle f, S^2(\mathbb{X})\rangle_\theta \langle S^2(\mathbb{X}), S^2(\mathbb{X})\rangle_\theta S^2(\mathbb{X})\}\\ &= S[\mathbb{E}\{\langle S^{-1}(f), S(\mathbb{X})\rangle\langle S(\mathbb{X}), S(\mathbb{X})\rangle S(\mathbb{X})\}]\\ &= S\mathcal{K}_{S(\mathbb{X})}(S^{-1}(f)).\end{aligned}$$

■

Then, we can establish the following equivalences which follows immediately from Propositions 7 and 8.

Corollary 1. *The smoothed independent components $\tilde{\xi}_{\theta,j},\ j \in \mathbb{N}_+$ of X satisfying $\mathbb{E}(\tilde{\xi}_{\theta,j}\tilde{\xi}_{\theta,j'}) = \delta_{jj'}$ are equivalently obtained by the following projections:*

1. $\langle \mathbb{X}, \psi_{\theta,j}\rangle$
2. $\langle S^2(\mathbb{X}), \psi_{\theta,j}\rangle_\theta$
3. $\langle S(\mathbb{X}), S^{-1}(\psi_{\theta,j})\rangle$.

As a consequence, the orthogonal representation for the half-smoothed whitened variable $S(\mathbb{X})$ in terms of the independent components is obtained by the expansion

$$S(\mathbb{X}) = \sum_{j=1}^{\infty} \tilde{\xi}_{\theta,j} S^{-1}(\psi_{\theta,j}). \tag{2.32}$$

Basis expansion implementation

A general strategy for solving the continuous eigenproblem to an equivalent matrix eigenanalysis is to consider a representation of the empirical counterpart of X with a finite basis of functions. Let $X^{[q]}(t) = (X_1^{[q]}(t), \ldots, X_n^{[q]}(t))^\top$ be a vector-valued function containing n copies of X assumed in a q-dimensional Hilbert space. Each function of $X^{[q]}(t)$ admits the basis function representation

$$X^{[q]}(t) = A\phi(t), \tag{2.33}$$

where $A \in \mathbb{R}^{n\times q}$ is a matrix of coefficients and $\phi(t) = (\phi_1(t), \ldots, \phi_q(t))^\top$ their respective vector of basis functions. The linear span of $\phi(t)$ is denoted by $H^{[q]}$ with inner product defined as $\langle f, g\rangle = \mathtt{f}^\top \mathcal{G}\mathtt{g}$, where $\mathtt{f}, \mathtt{g}$ are the coefficient vectors of the functions $f, g \in H^{[q]}$ and $\mathcal{G} = \langle\phi_j, \phi_k\rangle \in \mathbb{R}^{q\times q}, j, (k = 1 \ldots q)$, that is, the inner products of each pair of basis functions, so that possibly $\mathcal{G} \neq I_q$ when $\phi(t)$ might be not orthonormal in the usual sense. Recalling the computational algorithms in §2.3, we henceforth assume that $\mathbb{X}^{[q]}(t) = \tilde{A}\phi(t)$ is a set of whitened functional data, i.e., a basis expansion with coefficient matrix satisfying $n^{-1}\tilde{A}^\top\tilde{A} = \mathcal{G}$ ($\tilde{A}$ have identity covariance matrix in the topology of the space). Then, from expression 2.24, we can define the sample kurtosis operator as

$$\begin{aligned}\mathcal{K}_{\mathbb{X}^{[q]}}(f)(s) &= n^{-1}\sum_{i=1}^{n}\left\langle \mathbb{X}_i^{[q]}, \mathbb{X}_i^{[q]}\right\rangle \left\langle \mathbb{X}_i^{[q]}, f\right\rangle \mathbb{X}_i^{[q]}(s) = \left\langle n^{-1}\sum_{i=1}^{n}\left\|\mathbb{X}_i^{[q]}\right\|^2 \mathbb{X}_i^{[q]}(s)\mathbb{X}_i^{[q]}, f\right\rangle\\ &= \langle K^{[q]}(s,\cdot), f\rangle,\end{aligned} \tag{2.34}$$

where $K^{[q]}(s,t)$ is a kurtosis kernel function admitting the following representation in terms of an orthonormalized basis

$$K^{[q]}(s,t) = \phi^\top(s)\mathcal{G}^{-1/2}(n^{-1}\mathcal{G}^{1/2}\tilde{A}^\top D\tilde{A}\mathcal{G}^{1/2})\mathcal{G}^{-1/2}\phi(t), \tag{2.35}$$

where $D = \operatorname{diag}(\tilde{A}\mathcal{G}\tilde{A}^\top)$, i.e. $D_{ii} = \|\mathbb{X}_i\|^2$.

Proposition 9. *Given the basis expansion in 2.33, the functional ICA of $X_i^{[q]}$ with respect to the inner product $\langle\cdot,\cdot\rangle$ is equivalent to the multivariate ICA of matrix $A\mathcal{G}^{1/2}$ with the usual metric in $\mathbb{R}^q$.*

Proof. As indicated beforhand, the weight functions of the independent components are obtained as the eigenfunctions of the sample kurtosis operator by solving the following eigenproblem:

$$\mathcal{K}_{\mathbb{X}}\psi(t) = \kappa\psi(t). \tag{2.36}$$

If we expand the independent component weight functions as $\psi(t) = \phi(t)^\top b$, then the problem turns in matrix form as

$$n^{-1}\tilde{A}^\top D\tilde{A}\mathcal{G}b = \kappa b, \tag{2.37}$$

which is equivalent to

$$n^{-1}\mathcal{G}^{1/2}\tilde{A}^\top D\tilde{A}\mathcal{G}^{1/2}u = \kappa u, \tag{2.38}$$

with $u = \mathcal{G}^{1/2}b$. That is, $\Sigma^{[4]}_{\tilde{A}\mathcal{G}^{1/2}}u = \kappa u$. Taking into account that the matrix $\tilde{A}\mathcal{G}^{1/2}$ is the whitening counterpart of the matrix $A\mathcal{G}^{1/2}$, we can conclude that the functional ICA is equivalent to ICA of matrix $A\mathcal{G}^{1/2}$. ∎

Proposition 10. *For any $\theta > 0$, the penalized functional ICA of $X_i^{[q]}$ defined by the successive optimization problem in 2.28 is equivalent to the multivariate ICA of the matrix $A\mathcal{G}^{1/2}$ using the metric $\mathcal{M} = (L^{-1}\mathcal{G}^{1/2})^\top(L^{-1}\mathcal{G}^{1/2})$ in $\mathbb{R}^q$, with L defined by the factorization $\mathcal{G}_\theta = \mathcal{G} + \theta\mathfrak{P}_2 = LL^\top$, and $\mathfrak{P}_2$ the matrix whose elements are $\langle R\phi_j, R\phi_k\rangle$.*

Proof. If we expand the weight functions in 2.27 as $f(t) = \sum_{j=1}^q b_j\phi_j(t) = \phi(t)^\top b$ where $b = (b_1,\ldots,b_q)^\top$, the coefficients of $\psi_{\theta,j}$, are obtained by solving the penalized kurtosis problem 2.27 expressed in matrix form as

$$\frac{n^{-1}b^\top\mathcal{G}\tilde{A}^\top D\tilde{A}\mathcal{G}b}{b^\top\mathcal{G}b + \theta b^\top\mathfrak{P}_2 b} = \frac{b^\top\mathcal{G}^{1/2}\Sigma^{[4]}_{\tilde{A}\mathcal{G}^{1/2}}\mathcal{G}^{1/2}b}{b^\top(\mathcal{G}+\theta\mathfrak{P}_2)b}. \tag{2.39}$$

The above developments can be used to transform the eigenequation 2.29 into the matrix eigenproblem

$$\mathcal{G}^{1/2}\Sigma^{[4]}_{\tilde{A}\mathcal{G}^{1/2}}\mathcal{G}^{1/2}b = \kappa_\theta\mathcal{G}_\theta b, \tag{2.40}$$

where $\mathcal{G}_\theta = \mathcal{G} + \theta\mathfrak{P}_2$. Then, by performing the factorization $\mathcal{G}_\theta = LL^\top$, the eigenequation 2.40 can be rewritten as

$$L^{-1}\mathcal{G}^{1/2}\Sigma^{[4]}_{\tilde{A}\mathcal{G}^{1/2}}\mathcal{G}^{1/2}(L^{-1})^\top v = \kappa_\theta v, \tag{2.41}$$

where $v = L^\top b$, with $v^\top v = 1$.

Now, defining $w = (L^{-1}\mathcal{G}^{1/2})^{-1}v$, the eigenproblem turns on

$$\Sigma^{[4]}_{\tilde{A}\mathcal{G}^{1/2}}(L^{-1}\mathcal{G}^{1/2})^\top(L^{-1}\mathcal{G}^{1/2})w = \kappa_\theta w, \tag{2.42}$$

with $w^\top(L^{-1}\mathcal{G}^{1/2})^\top(L^{-1}\mathcal{G}^{1/2})w = 1$. This means that the smoothed functional ICA is equivalent to the ICA of matrix $A\mathcal{G}^{1/2}$ with a new metric in $\mathbb{R}^q$ defined by $\langle x,y\rangle_{\mathcal{M}} = x^\top\mathcal{M}y$, for all $x,y\in\mathbb{R}^q$.

Therefore, solving 2.41 yields to $b_{\theta,j} = (L^{-1})^\top v_{\theta,j} = (L^{-1})^\top L^{-1}\mathcal{G}^{1/2}$ such that $\psi_{\theta,j}(t) = \phi(t)^\top b_{\theta,j}$ is the solution to the eigenequation 2.40. By computing the successive optimization problems in 2.28 we obtain a set of orthonormal eigenfunctions verifying

$$\|\psi_{\theta,j}\|^2_\theta = b_{\theta,j}^\top\mathcal{G}_\theta b_{\theta,j} = v_j^\top v_j = 1; \quad \langle\psi_{\theta,j},\psi_{\theta,k}\rangle_\theta = b_{\theta,j}^\top\mathcal{G}_\theta b_{\theta,k} = v_j^\top v_k = 0. \tag{2.43}$$

∎

From the relation between inner products given by 2.30, it can be deduced that the operator S^2 is defined as $S^2(f) = \phi(t)^\top(\mathcal{G}+\theta\mathfrak{P}_2)^{-1}\mathcal{G}\mathtt{f}$, with $f = \phi(t)^\top\mathtt{f}$. Then, for the smoothed whitened data $S(\mathbb{X})$, the independent component scores are obtained as $\xi_{\theta,j} = A^\top\mathcal{G}(L^{-1})^\top v_{\theta,j}$ and the kurtosis eigenfunctions as $\varphi_j = S^{-1}(\psi_{\theta,j})$.

§ 2.6. Kurtosis based functional ICA on smoothed KL expansions

Vidal et al. [303] introduced an alternative functional ICA build upon smoothed KL expansions. There, the covariance eigenfunctions are estimated using Silverman's method [276], which corresponds to the maximization problem

$$\gamma_{\theta,j} = \operatorname{argmax} \frac{\operatorname{var}(\langle \gamma, X \rangle)}{||\gamma||^2 + \theta \langle R\gamma, R\gamma \rangle} = \max \frac{\langle \gamma, \mathscr{C}_X \gamma \rangle}{||\gamma||^2_\theta}, \quad \text{s.t. } \langle \gamma, \gamma_{\theta,k} \rangle_\theta = 0, \text{ for all } k < j. \tag{2.44}$$

The functions $(\gamma_{\theta,j})_{j \in \mathbb{N}}$ form a complete orthonormal system in the subspace endowed by $\langle \cdot, \cdot \rangle_\theta$, making this basis non-compatible for an IC model in H. As previously mentioned, [223] developed a generalization of the Silverman's method providing the following equivalences:

Proposition 11 (Ocaña et al. [223], Proposition 5.4; Vidal et al. [303], Proposition 1)**.** *Assume for any $f \in H$, the derivatives of f up to certain order are absolutely continuous in H. Then, there exists a positive definite self-adjoint operator S^2 such that the following functional PCA decompositions are equivalent:*

- *The functional PCA of $S^2(X)$ with respect to $\langle \cdot, \cdot \rangle_\theta$, $S^2(X) = \sum_j \xi_j \gamma_{\theta,j}$;*
- *The functional PCA of $S(X)$ with respect to $\langle \cdot, \cdot \rangle$, $S(X) = \sum_j \xi_j S^{-1}(\gamma_{\theta,j})$;*
- *The functional PCA of X with respect to $\langle \cdot, \cdot \rangle_S$, $X = \sum_j \xi_j S^{-2}(\gamma_{\theta,j})$, with $\langle f, g \rangle_S = \langle S(f), S(g) \rangle = \langle S^2(f), S^2(g) \rangle_\theta$.*

Therefore, the eigenfunctions of the covariance operator $\mathscr{C}_{S(X)} = S\mathscr{C}_X S$ of $S(X)$ are given by $\beta_j = S^{-1}(\gamma_{\theta,j})$, where $\gamma_{\theta,j}$ are obtained by the penalized estimation procedure 2.44. Then, the basis β_j is orthonormal with respect to the usual inner product in H, so that the smooth random variable $S(X)$ can be approximated by the KL expansion

$$X^{[p]}(t) = \sum_{j=1}^{p} \xi_j \beta_j, \tag{2.45}$$

where $\xi_j = \langle \beta_j, S(X) \rangle = \langle \gamma_{\theta,j}, X \rangle$ and, by abuse of notation, $X^{[p]}(t)$ denotes a truncated representation of the smoothed random variable $S(X)$. The functional IC model presented in the following subsection, consists in performing the functional ICA of $X^{[p]}(t)$, which is equivalent to the multivariate ICA of the ξ_j's. In this model, it is assumed that the interesting non-Gaussian ICs lie in the space spanned by the first p eigenfunctions of the operator $\mathscr{C}_{S(X)}$.

Basis expansion implementation

To derive smoothed KL expansions using basis functions, here we adopt the B-spline functional PCA approach developed by Aguilera and Aguilera-Morillo [4], which incorporates a discrete penalty (P-spline penalty) on the orthonormality constraint described above. Consider the B-spline basis expansion of the covariance eigenfunctions $\gamma(t) = \phi(t)^\top b$, with $b = (b_1, \ldots, b_q)^\top$ being its vector of basis coefficients, and a roughness penalty function defined by $\operatorname{pen}_d(\gamma) = b^\top \mathfrak{P}_d b$, where $\mathfrak{P}_d \in \mathbb{R}^{q \times q}$ is the penalty matrix $\mathfrak{P}_d = \Delta_d^\top \Delta_d$, with Δ_d being a matrix representation of the d-order difference operator $R^{[d]}$. As in §2.5, we assume second order differences ($R \equiv R^{[2]}$) for defining the penalty function $b^\top \mathfrak{P}_2 b = (b_1 - 2b_2 + b_3)^2 + \cdots + (b_{q-2} - 2b_{q-1} + b_q)^2$. This way, the inner product in 2.26 is given in terms of B-splines expansions as

$$\langle f, g \rangle_\theta = \mathtt{f}^\top \mathcal{G} \mathtt{g} + \theta \mathtt{f}^\top \mathfrak{P}_2 \mathtt{g}, \tag{2.46}$$

with $f = \phi^\top \mathtt{f}$, $g = \phi^\top \mathtt{g}$ and $\mathcal{G}$ the Gram matrix of inner products between basis functions (see §2.5 for the notation). Then, the maximization problem in 2.44 is equivalent to solve the following matrix problem:

$$b_{\theta,j} = \operatorname{argmax} \frac{b^\top \mathcal{G} \Sigma_A \mathcal{G} b}{b^\top (\mathcal{G} + \theta \mathfrak{P}_2) b}, \tag{2.47}$$

subject to the constraint $b^\top (\mathcal{G} + \theta \mathfrak{P}_2) b_{\theta,k} = 0$ for all $k < j$, where $\Sigma_A = n^{-1} A^\top A$ and $\theta \geq 0$ is the penalty parameter used to control the trade-off between maximizing the sample variance and the strength of the penalty.

Because a B-spline basis is not orthonormal by construction, one can apply Cholesky factorization of the form $LL^\top = \mathcal{G} + \theta \mathfrak{P}_2$ in order to derive an isometry that allows us to operate in terms of usual geometrical structure. Then, the smoothed functional PCA corresponds to solve the eigenvalue problem

$$L^{-1} \mathcal{G} \Sigma_A \mathcal{G} (L^{-1})^\top v_j = \eta_j v_j, \tag{2.48}$$

where $v_j = L^\top b_{\theta,j}$, and the coefficients of $\gamma_{\theta,j}$ are $b_{\theta,j} = (L^{-1})^\top v_j$. We have obtained a set of orthonormal functions with respect to the inner product $\langle \cdot, \cdot \rangle_\theta$. Then, the jth smoothed principal component is given by

$$\xi_j = A\mathcal{G}b_{\theta,j} = A\mathcal{G}(L^{-1})^\top v_j. \tag{2.49}$$

Thus, the problem is reduced to the multivariate PCA of the matrix $A\mathcal{G}(L^{-1})^\top$, as discussed [4]. From the results in [222, 223], we now deduce the expression of the smoothing operator S that provides the equivalence between such multivariate PCA and the functional PCA of the smoothed data.

Proposition 12 (Vidal et al. [303], Proposition 2). *Given the basis expansion $X^{[q]}(t) = A\phi(t)$ of a sample of curves, the PCA of the matrix $A\mathcal{G}(L^{-1})^\top$ is equivalent to all functional PCA's in Proposition 11 with the operator S^2 defined as $S^2(f) = \phi(t)^\top(\mathcal{G} + \theta\mathfrak{B}_d)^{-1}\mathcal{G}\mathtt{f}$, with $f = \phi(t)^\top\mathtt{f}$.*

As a result, the principal components of $S(X_i)$ $(i = 1, \ldots, n)$ are given by $A\mathcal{G}(L^{-1})^\top V$ where V is the matrix whose columns are the eigenvectors v_j verifying Equation 2.48, and thus, the eigenfunctions β_j are $\beta_j = S^{-1}(\gamma_{\theta,j})$.

Having estimated the weight functions coefficients and principal components scores, assume 2.45 is truncated at some $p \leq q$. Then, the vector of smoothed sample curves is given by $X^{[p]}(t) = \Xi^{[p]}\beta(t)$, where $\Xi^{[p]} = (\xi_{ij}) \in \mathbb{R}^{n\times p}$ is the matrix whose columns are the first p principal components scores with respect to the basis of smoothed weight functions $\beta(t) = (\beta_1(t), \ldots, \beta_q(t))^\top$.

From now on, we only have to estimate the multivariate ICA of the matrix $\Xi^{[p]}$. Since $\Xi^{[p]}$ is uncorrelated, we can simply standardize this matrix as $\tilde{\Xi}^{[p]} = \Xi^{[p]}D_\eta$, where D_η is a diagonal matrix with entries $(\eta_1, \ldots, \eta_p)^{-1/2}$. Then, for all $h = \beta(t)^\top\mathtt{h}$, the kurtosis operator of the standardized curves $\mathbb{X}^{[p]}(t) = \tilde{\Xi}^{[p]}\beta(t)$ have the matrix expansion

$$\mathcal{K}_{\mathbb{X}^{[p]}}(h) = n^{-1}(\tilde{\Xi}^{[p]\top}D_{\tilde{\Xi}^{[p]}}\tilde{\Xi}^{[p]}\mathtt{h})^\top\beta(t), \quad \forall h = \beta(t)^\top\mathtt{h}, \tag{2.50}$$

where $D_{\tilde{\Xi}^{[p]}} = \operatorname{diag}(\tilde{\Xi}^{[p]}\tilde{\Xi}^{[p]\top})$. The eigenanalysis of $\mathcal{K}_{\mathbb{X}^{[p]}}$ leads to the diagonalization of the kurtosis matrix

$$\Sigma_{4,\tilde{\Xi}^{[p]}}u_l = \rho_l u_l \quad (l = 1, \ldots, p), \tag{2.51}$$

where $\Sigma_{4,\tilde{\Xi}^{[p]}} = n^{-1}(\tilde{\Xi}^{[p]\top}D_{\tilde{\Xi}^{[p]}}\tilde{\Xi}^{[p]}\mathtt{h})^\top$. The eigenfunctions of $\mathcal{K}_{\mathbb{X}^{[p]}}$ are $\beta(t)^\top u_l$, and the ICs are obtained as $\tilde{\Xi}^{[p]}u_l$. Alternatively one can consider $\Xi^{[p]}u_l$ (projecting on to the non-whitened principal components) and build an expansion via the eigengunctions of $\mathcal{K}_{\mathbb{X}^{[p]}}$, with any of the coefficient vectors previously discussed. See [303] for further details.

§ 2.7. Theoretical properties of the kurtosis operator

Discriminative properties of the kurtosis operator

Suppose that $X : \Omega \to H$ can be observed as a mixture of two subpopulations Π_k $(k = 0, 1)$, and the aim is to assign their sample paths into one of them. The two subpopulations are identified by the binary variable $Y = k$ when $X \in \Pi_k$. We denote by $\pi_k = P(Y = k)$ with $\pi_1 = 1 - \pi_0$. The prior probability of classification is usually unknown and in practice $\pi_0 = 1/2$ is often assumed. The estimation of the function that allows to optimally classify X corresponds to Fisher's discriminant problem. The purpose of this section is to show that the eigendecomposition of $\mathcal{K}_{\mathbb{X}}$ provides solutions to it.

Shin [275] proposed a seamless extension of Fisher's discriminant analysis in infinite dimensional settings we shall take into account in the following lines. We henceforth reset X to $\operatorname{ran}(\mathcal{C}_X)$, such the precision operator is possibly bounded under Picard's rule. Further, consider that class k has mean function $\mu_k = \mathbb{E}(X|Y = k)$ and $\mathbb{E}X = \mu = \pi_0\mu_0 + (1-\pi_0)\mu_1$. In principle, we do not impose distributional assumptions on X, but we consider $\mu_0 \neq \mu_1$ and equal class covariance operators. Within this functional framework, Fisher's discriminant problem consists in estimating a function f that maximizes the ratio

$$J(f) = \left\langle f, \mathcal{C}_X^W f \right\rangle^{-1} \left\langle f, \mathcal{C}_X^B f \right\rangle, \tag{2.52}$$

where $\mathcal{C}_X^B = \pi_0(1-\pi_0)\{(\mu_1 - \mu_0) \otimes (\mu_1 - \mu_0)\}$ and $\mathcal{C}_X^W = \pi_0\mathcal{C}_{X|Y=0} + (1-\pi_0)\mathcal{C}_{X|Y=1} = \mathcal{R}$ are respectively, the between and within-class covariance operator. We note that, $\mathcal{R} \equiv \sum_{j=1}^\infty \lambda_j \mathcal{P}_{\gamma_j}$ represents the common covariance operator in each population. Roughly speaking, the idea is to give large separation to the group means while, at the same time, keeping the variance between groups small. According to the law of the

total covariance, one can also write $\mathscr{C}_X = \mathscr{C}_X^W + \mathscr{C}_X^B$. Then, for a function f with expansion $f = \sum_j \mathtt{f}_j \gamma_j$, the covariance operator can be written in its spectral form as

$$\mathscr{C}_X(f) = \sum_{j=1}^{\infty} \{\lambda_j \mathtt{f}_j + \pi_0(1-\pi_0)\mathtt{v}_j \sum_{j'=1}^{\infty} \mathtt{v}_{j'} \mathtt{f}_{j'}\}\gamma_j, = \sum_{jj'=1}^{\infty} s_{jj'} \mathtt{f}_j \gamma_{j'}, \tag{2.53}$$

where $\mathtt{v}_j$ are the coefficients of the mean differences between classes in terms of γ_j's (i.e., $\mu_\Delta = \mu_0 - \mu_1 = \sum_{j=1}^{\infty} \mathtt{v}_j \gamma_j$) and

$$s_{jj'} = \begin{cases} \lambda_j + \pi_0(1-\pi_0)\mathtt{v}_j^2 & j = j' \\ \pi_0(1-\pi_0)\mathtt{v}_j \mathtt{v}_{j'} & j \neq j'. \end{cases} \tag{2.54}$$

Solutions to (2.52) are well-known. Here, we briefly debrief them for the sake of clarity.

Proposition 13 (Peña et al. [231], Lemma 1). *For some constant c, the function β with expansion $\mathfrak{F} = c \sum_{j=1}^{\infty} \lambda_j^{-1} \mathtt{v}_j \gamma_j$ is the maximizer of J in 2.52.*

According to this result, our goal is to show that the kurtosis operator has an eigenfunction that is equivalent to the function found in Proposition 13. This is proved in the next Proposition.

Proposition 14. *Let X be a random functional variable corresponding to the mixture of two groups Π_k $(k = 0, 1)$, with $\mu_k = \mathbb{E}(X|Y = k)$ and $\mu_0 \neq \mu_1$. Then, the kurtosis operator of the whitened functional variable $\mathbb{X} \mapsto \mathscr{K}_{\mathbb{X}}$ has an eigefunction that corresponds to $c \sum_{j=1}^{\infty} \lambda_j^{-1} \mathtt{v}_j \gamma_j$.*

Proof. Let $\tilde{\mu}_\Delta = \Psi(\mu_\Delta)$. From Proposition 1 in [296], it can be easily proven that $\tilde{\mu}_\Delta$ is an eigenfunction of $\mathscr{C}_{\mathbb{X}}$ since $\mathscr{C}_{\mathbb{X}}(\tilde{\mu}_\Delta) = P_{\overline{\text{ran}}(\Psi)}(\tilde{\mu}_\Delta) = 1 \cdot \tilde{\mu}_\Delta$; see also Corollary 1 in [231]. Thus, the operator maps a function in the domain of $\mathscr{C}_{\mathbb{X}}$ and it gives the original function again with some constant value. Now, consider the kurtosis operator expressed as follows:

$$\mathscr{K}_{\mathbb{X}} = \mathbb{E}\left[(\mathbb{X} \otimes \mathbb{X})^2\right] = \mathbb{E}\left(\|\mathbb{X}\|\mathbb{X} \otimes \|\mathbb{X}\|\mathbb{X}\right) = \mathscr{C}_{\|\mathbb{X}\|\mathbb{X}}, \tag{2.55}$$

and the function $\|\tilde{\mu}_\Delta\|\tilde{\mu}_\Delta$ where $\|\tilde{\mu}_\Delta\| = \langle \Psi(\mu_\Delta), \Psi(\mu_\Delta)\rangle^{1/2}$. Clearly, $\|\tilde{\mu}_\Delta\|\tilde{\mu}_\Delta$ is in the domain of $\mathscr{C}_{\|\mathbb{X}\|\mathbb{X}}$. Further,

$$\mathscr{C}_{\|\mathbb{X}\|\mathbb{X}}(\|\tilde{\mu}_\Delta\|\tilde{\mu}_\Delta) = \left(\sum_{j=1}^{\infty} \kappa_j \psi_j \otimes \psi_j\right)(\|\tilde{\mu}_\Delta\|\tilde{\mu}_\Delta) = \sum_{j=1}^{\infty} \kappa_j (\|\tilde{\mu}_\Delta\|\tilde{\mu}_\Delta) = \tau \cdot (\|\tilde{\mu}_\Delta\|\tilde{\mu}_\Delta), \tag{2.56}$$

where $\|\tilde{\mu}_\Delta\|\tilde{\mu}_\Delta$ is invariant to rotations of the eigenbasis of $\mathscr{C}_{\|\mathbb{X}\|\mathbb{X}}$ and τ is the trace of $\mathscr{C}_{\|\mathbb{X}\|\mathbb{X}}$, which by Proposition 6 we are assured of its finiteness. Then, note that $\|\tilde{\mu}_\Delta\|\tilde{\mu}_\Delta = \|\mu_\Delta\|\Psi(\tilde{\mu}_\Delta) = c \sum_{j=1}^{\infty} \lambda_j^{-1} \mathtt{v}_j \gamma_j$, proving that $\|\tilde{\mu}_\Delta\|\tilde{\mu}_\Delta$ is an eigenfunction of $\mathscr{K}_{\mathbb{X}}$ equivalent to the discriminant function found in Proposition 13. See also Theorem 1 in [231]. ∎

The duality of the kurtosis operator and near-perfect classification

In this subsection, we assume that $(X_{ik}, Y_{ik}) = \{(X_{1k}, Y_1), \ldots, (X_{nk}, Y_{n_k})\}$ $(k = 0, 1)$, is the empirical counterpart of (X, Y) defined as in Subsection 4.1, with sample means denoted by $\hat{\mu}_k$.

Next, we concisely review the asymptotic centroid-based classifier proposed by [77] which is constructed by projecting a newly observed function X^* on to a pre-chosen square-integrable function β on $\mathcal{I} = [0, T]$. Although this classifier reaches optimal performance in the case of a homoscedastic Gaussian mixture, here we focus on a more general scenario. The centroid classifier is there defined as $T_n(X^*) = D^2(X^*, \hat{\mu}_1) - D^2(X^*, \hat{\mu}_0)$, where $D(X^*, \hat{\mu}_k) = |\langle X^*, \mathfrak{F}\rangle - \langle \hat{\mu}_k, \mathfrak{F}\rangle|$. Assuming $\hat{\mu}_0 = 0$ and denoting $\hat{\mu}_1 \equiv \hat{\mu}_\Delta = \sum_{j=1}^{\infty} \mathtt{v}_j \gamma_j$, the asymptotic version of this classifier, T^0, assigns X^* to Π_1 if

$$(\langle X^*, \mathfrak{F}\rangle - \langle \hat{\mu}_\Delta, \mathfrak{F}\rangle)^2 - (\langle X^*, \mathfrak{F}\rangle)^2 < 0. \tag{2.57}$$

Now, to elucidate our ideas we formulate Theorem 2 in [77] and discuss their implications with regard to Picard's condition and the Hájek-Feldman dichotomy.

Theorem 5 (Delaigle and Hall [77] - Theorem 2). *Assume X_{ik} $(i = 1, \ldots, n; k = 0, 1)$, are non necessarily Gaussian, $\mu_0 \neq \mu_1$ and $\mathscr{C}_{X_k} = \mathscr{C}_X$. If Π_0 and Π_1 have prior probabilities π_0 and $1 - \pi_0$ respectively, and $\mu_0 = 0$, then*

1. *The missclassification probability for the classifier* T^0 *equals* $\mathrm{err} = \pi P(Q > v/2\sigma_Q) + (1-\pi)P(Q < -v/2\sigma_Q)$, *where* $Q = \langle X - \mu, \mathfrak{F} \rangle$ *and* $v = \langle \hat{\mu}_\Delta, \mathfrak{F} \rangle$.

2. *If* $\sum_{j=1}^{\infty} \lambda_j^{-1} \mathrm{v}_j^2 = \infty$, *by taking a sequence of classifiers build from* $\mathfrak{F}^{[q]} = \sum_{j=1}^{q} \lambda_j^{-1} \mathrm{v}_j \gamma_j$ *with* $q \to \infty$, *the minimal missclassification probability tends to* $\mathrm{err}_0 = 0$, *and perfect classification is then possible.*

Geometrically, from Theorem 5 one deduces that asymptotic perfect classification is related to the divergence in norm induced by the metric $\langle \Psi(\cdot), \Psi(\cdot) \rangle = \langle \cdot, \cdot \rangle_{\mathbb{M}}$, as $\|\hat{\mu}_\Delta\|_{\mathbb{M}}^2 = \sum_{j=1}^{\infty} \lambda_j^{-1} \mathrm{v}_j^2 = \infty$. Note this is the same as assuming that Picard's condition does not hold (the coefficients v_j decay slower than the corresponding λ_j's and therefore, $\hat{\mu}_\Delta \notin \mathrm{ran}(\Psi)$). On the other hand, if $\|\hat{\mu}_\Delta\|_{\mathbb{M}}^2 < \infty$, the minimum classification error is strictly positive, and perfect classification cannot be reached. Exploring nonconvergent paths seems a viable option, as it can still provide optimal solutions considering that $\mathrm{err}_0 \to 0$ along these paths.

The reason of the behavior described in Theorem 5 has a probabilistic interpretation by the Hájek-Feldman dichotomy. Two probability measures m_k $(k = 0, 1)$, are said to be equivalent $(m_0 \sim m_1)$ if they are absolutely continuous with respect to one another: i.e., if $m_0(B) = 0$ for every Borel set $B \in \mathbb{B}$, it holds $m_1(B) = 0$ (they have the same zero sets). Conversely, if $m_0(B) = 0$ and $m_1(B) = 1$, then we say that m_0 and m_1 are mutually singular $(m_0 \perp m_1)$ as B splits in two disjoint sets where m_0 and m_1 are respectively concentrated. The Hájek-Feldman dichotomy states that in infinite dimensions, two Gaussian measures have the critical property of being either equivalent or mutually singular.

Theorem 6 (Da Prato and Zabczyk [68] - Theorem 2.25)**.** *Let* $m_k = N(\mu_{m_k}, \mathscr{C}_{m_k})$, $(k = 0, 1)$, *be two Gaussian measures on* H. *Then,* $m_0 \sim m_1$ *iff, it holds:*

1. *Both measures have the same Cameron-Martin space, i.e.,* $\mathrm{ran}(\mathscr{C}_m^{1/2}) = \mathrm{ran}(\mathscr{C}_{m_0}^{1/2}) = \mathrm{ran}(\mathscr{C}_{m_1}^{1/2})$.

2. $\mu_{m_0} - \mu_{m_1} \in \mathrm{ran}(\mathscr{C}_m^{1/2})$.

3. $(\mathscr{C}_{m_0}^{-1/2}\mathscr{C}_{m_1}^{1/2})(\mathscr{C}_{m_0}^{-1/2}\mathscr{C}_{m_1}^{1/2})^* - I_H$ *is a Hilbert-Schmidt operator on* $\overline{\mathrm{ran}}(\mathscr{C}_m^{1/2})$.

If one of the above conditions is violated, then $m_0 \perp m_1$.

Now recall Theorem 5. Suppose that all functions of X_i are Gaussian via the measures m_k. In [34, Theorem 5], it has been proven that $m_0 \sim m_1 \Leftrightarrow \|\hat{\mu}_\Delta\|_{\mathbb{M}}^2 < \infty$ and $m_0 \perp m_1 \Leftrightarrow \|\hat{\mu}_\Delta\|_{\mathbb{M}}^2 = \infty$, thus explaining the mechanisms underlying the dichotomy found in [77]. This result follows from Theorem 6 and Parseval's formula as $\hat{\mu}_\Delta \in \mathrm{ran}(\mathscr{C}_X) = \mathrm{ran}(\mathscr{C}_X^{1/2})$ iff $\|\hat{\mu}_\Delta\|_{\mathbb{M}}^2 < \infty$. Let us note that $\mathbb{M}$ is here a Cameron-Martin space, which is defined by Picard's law and whose geometric structure is build from the factorization of $\mathscr{C}_X^{\dagger}$, i.e., $\langle \Psi(\cdot), \Psi(\cdot) \rangle = \langle \cdot, \cdot \rangle_{\mathbb{M}}$.

Assuming $\|\hat{\mu}_\Delta\|_{\mathbb{M}}^2 < \infty$ under finite dimension space dependency, yet, it can be computationally proven that $\mathrm{kurt}(\langle \mathbb{X}_i^{[q]}, \psi_j \rangle) \to 3$ $(j = 1, \ldots, q-1)$ and $\mathrm{kurt}(\langle \mathbb{X}_i^{[q]}, \psi_q \rangle) \to 1$ if the means differ enough and n diverges. This corresponds to the schematic of perfect classification in the Gaussian homocedastic case (maximum bimodality is then reached). By Theorem 5, it can be straightforwardly deduced that, in these cases, $\lambda_j \to 0$ rapidly, such that 0 becomes an accumulation point of the spectrum of the corresponding non-degenerate population covariance operator. One could further presume that $\langle X^{[q]}, \gamma_1 \rangle$ will also exhibit a kurtosis close to 1 in such cases.While solutions to choose the best $\mathfrak{F}$ are given by the last eigenfunction in the independent component expansion, they also exist arbitrarily in the tails of the principal component expansion [77], and eventually (when λ_j are strongly biased upwards due to multicollinearity en each group of functions) they concentrate towards the first principal component.

The above results can be seen parallel to those of Theorem 5: in fact, in the Gaussian case, half the absolute value of the kurtosis coefficient minus 3 will take values in $[0, 1]$ and can be interpreted as a probabilistic measure. It is insightful to give a formal argument in the current operating context.

Corollary 2. *Let* $\|\nu\|_{\mathbb{B}} = \sup\{|\nu(B)| \; B \subseteq \mathbb{B}\}$ *define a norm on the space of Borel measures. Then,* $m_0 \sim m_1 \Leftrightarrow \|m_0 - m_1\|_{\mathbb{B}} = 0$ *and* $m_0 \perp m_1 \Leftrightarrow \|m_0 - m_1\|_{\mathbb{B}} = 1$.

Proof. Is immediate from the properties described above. ∎

Not surprisingly, low kurtosis has been previously associated, with some reservations, to bimodality in symmetric distributions. Noting that $\mathrm{kurt}(\xi_j) = \langle \mathscr{K}_{\mathbb{X}} \psi_j, \psi_j \rangle = \kappa_j$, the spectrum of $\mathscr{K}_{\mathbb{X}}$ provides a unique avenue for assessing the trade-off between equivalence/singularity of two Gaussian measures on the sample paths of X, as well as a way to prospect the chances of correct classification, even in non-Gaussian scenarios.

§ 2.8. Simulated and real data examples

The following numerical studies correspond to the functional IC model in §2.5, which is used for functional data classification purposes.

Simulated data

To investigate the empirical performance of the proposed estimators discussed in §2.5, we conduct a study that extends the results of Simulation 2 in [231] using three possible taxonomies of mean differences. Let X_i $(i = 1, \ldots, n)$, be a mixture of two subpopulations Π_k, $(k = 0, 1)$, with $n_k = n/2$ curves sampled on a grid of 20 equispaced points on $t \in [1, T]$ with $T = 20$. Both groups have same quadratic covariance matrix $\text{cov}(t_j, t_{j'}) = \exp\{-(2\ell^2)^{-1}(t_j - t_{j'})^2\}$, $(j, j' = 1, \ldots, T)$, with $\ell = 15$. The data is then generated as

$$X_{ij} = \sum_{k=0}^{1} \left(\sum_{j=1}^{T} \lambda_j^{1/2} Z_{k,ij} \gamma_j + \mu_k + \epsilon_{k,ij} \right) \mathbb{I}\left(X_i \in \Pi_k\right), \tag{2.58}$$

where Z_{ij} are Gaussian random variables, $\epsilon_{k,ij}$ is an additive error term and $\mathbb{I}$ denotes the indicator function. Further extensions of the above model to non-Gaussian settings using $Z_{k,ij} \sim \exp(1) - 1$ can be also found in our results.

We consider the following versions of the above model: in Example 1 we define $\mu_0 = 0$ and $\mu_1 = 0.2\cos(3\pi t/T)$, the means differ in shape; in Example 2, $\mu_0 = 0.3\cos(3\pi t/T)$ and $\mu_1 = 0.2\cos(3\pi t/T)$, the means have equal shape and slightly differ in amplitude; in Example 3 we set $\mu_0 = 0.2\sin(3\pi t/T)$ and $\mu_1 = 0.2\cos(3\pi t/T)$, the means are equal in shape but dephased $\pi/2$. In all cases, $Z_{k,ij}$ are sampled from a standard normal distribution and $\epsilon_{k,ij} \sim \mathcal{N}(0, \sigma^2)$. We generated 200 datasets for each experiment with sample sizes $n_k = 30, 50$. The R package pfica [297] was used for the implementation of various functional pre-whitening methods via B-spline expansions with $q = 5$.

Results for $\sigma = 0$ are shown in Table A.1 in Appendix A. In all examples, the overall good behavior of kurtosis classifiers based on the last independent component (minimum kurtosis) is apparent, particularly for the smoothed kurtosis projections and large sample sizes. In Example 3, the PC with the lowest kurtosis coefficient performed notably well, similarly to their kurtosis peers and eventually outperforming the rates of the non-smoothed kurtosis. Regarding functional whitening, results indicate that classification optimization with the proposed operators is not that different, although Cholesky whitening reaches good performance in Examples 1 and 3, while zero-phase components analysis whitening does better in Example 2. In the non-Gaussian simulation, results are more balanced between both functional ICA's, although superior to the rest of competitors. Figure 2.1 further illustrates the effect of modulating the noise in the performance of the classifiers on a Gaussian scenario. Note that as σ grows, it exponentially worsens the classification rate. Notwithstanding, results for the smoothed kurtosis are very competitive for mild levels of noise. These analyses point to the importance of finding a good trade-off between groups when smoothing the data, as both noise and the type of smoothing (especially if it is homogeneous across curves) can undermine the effectiveness of these classifiers.

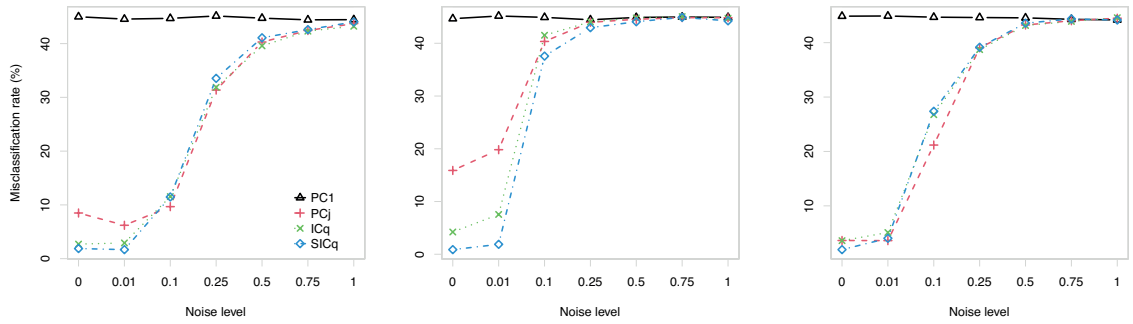

Figure 2.1: Comparison of the missclassification rate (mean values for training samples of size 200) for the three Examples 1-3 (from left to right) and different levels of noise (σ) using ZCA whitening.

Real datasets

Our methods are now applied to well-known datasets in the functional data literature. In the first example, we show that the smoothed kurtosis is able to find bimodality in the Canadian Weather data, which is usually treated as a discrimination problem of more than two groups. We consider a geographical division based on a west-east location distribution rather than the usual four climate regions. The whitening method

and penalty parameter was selected using cross-validation by minimizing the kurtosis coefficient of the projections on to ψ_q (Figure 2.2A) with $q \in \{5, \ldots, 34\}$, $\theta \in \{0, 100, \ldots, 10^8\}$. Results suggest the presence of bimodality in these data, and the few missclasified observations appear to be locations close to large bodies of water, commonly encountered in the west zone. Due the representativeness of these data, we further asses the asymptotic behavior of $\|\Delta\mathscr{C}_{\mathbb{X}[q]}\|_{HS} = \|\mathscr{C}_{\mathbb{X}[q]} - I_q\|_{HS}$ as q increases, to evaluate the consistency of the whitening procedures. As shown in Figure 2.2B most of the whitening procedures are (under mild conditions) consistent, with Cholesky whitening being the less consistent albeit the one that provides more interesting results.

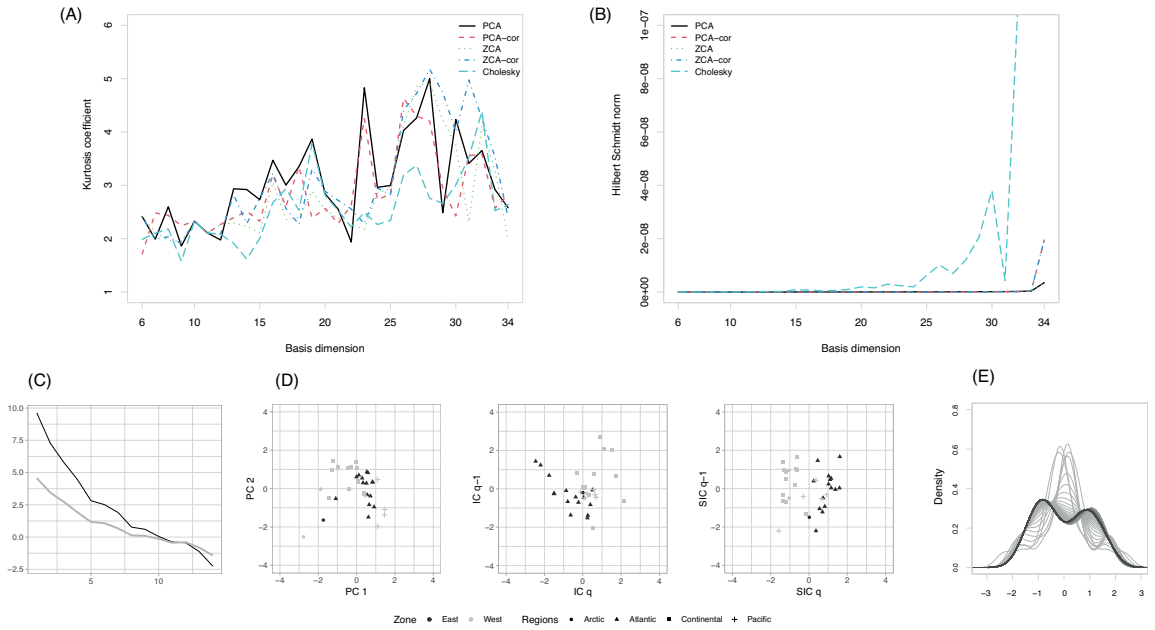

Figure 2.2: Canadian Weather data results. **(A)** Kurtosis coefficient of $\xi_{\theta,q}, \theta \in \{0, 100, \ldots, 3 \cdot 10^4\}$ for various B-spline basis dimensions (q) and whitening procedures. **(B)** The plot shows the asymptotic behavior of the whitening transformations when q grows up to n as evaluated by $\|\Delta\mathscr{C}_{\mathbb{X}[q]}\|_{HS}, q \in \{6, \ldots, 34\}$ **(C)** Picard's plot. The black line stands for the eigenvalues $\log(\lambda_j)$ and the grey one, for the means of absolute values of principal component scores given by $\sum_{i=1}^{n} |\langle X_i^{[q]}, \gamma_j \rangle|$ expressed in a logarithmic scale. **(D)** Scatter plots. From left to right: functional PCA, ICA and smoothed functional ICA using a basis expansion of $q = 14$ **(E)** Estimated densities of the vector $\xi_{\theta,q}$ for each lambda, showing the effect of smoothing the kurtosis operator.

In a second example, we consider the phoneme dataset as analyzed in [77]. The data were retrieved from the fds package [272] and consist of 400 log-periodograms constructed from audio recordings of males pronouncing the phonemes 'aa' as in dark and 'ao' as in water. The similarity between both groups of curves has been previously reported to pose a challenging problem of classification. In fact, we were neither able to find interesting projections with any of the proposed methods. As workaround, we propose to perform a functional PCA on each sample and use the basis function expansion

$$X_{k,i}^{[p]} = \sum_{j=1}^{p} \langle X_{k,i}^{[q]}, \gamma_{k,j} \rangle \gamma_{k,j} \mathbb{I}\left(X_i \in \Pi_k\right), \tag{2.59}$$

which takes the matrix form

$$X_k^{[p]} = (A_k \mathscr{G} b_k^\top) b_k \phi(t), \tag{2.60}$$

where $b_k = U_k \mathscr{G}^{-1/2}$, with $U_k \in \mathbb{R}^{p \times q}$ the matrix of eigenvectors of $n^{-1}\mathscr{G}^{1/2} A_k^\top A_k \mathscr{G}^{1/2}$ truncated at the p-row and $X_k^{[p]} = (X_{k,1}^{[p]}, \ldots, X_{k,n_k}^{[p]})^\top$. Taking $X^{[p]} = (X_1^{[p]}, X_2^{[p]})$ and using the coefficients in terms of basis functions pooled by rows, one can perform the functional ICA using these coefficients. However, the matrix $A = \{(A_k \mathscr{G} b_k^\top) b_k\}_{k=0}^{1}$ might have non invertible covariance matrix. Although this could be reversed with a suitable Tikhonov regularization, to avoid harming the whitening procedure, the best option is to truncate and perform the functional ICA on the p-principal components. Therefore, we use the first components up to the limit where the whitening transformation no longer meets the orthonormality

property. Results in Figure 2.2, taking $p = 8$ components, show the great improvement of performing functional ICA on these representations, which achieves near-perfect classification with an error rate of 0.125 %.

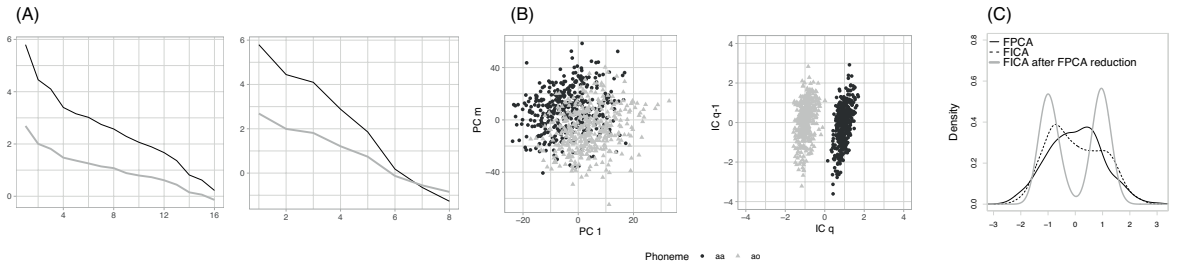

Figure 2.3: Phoneme data results. **(A)** Picard's plot before and after a functional PCA per group. **(B)** Scatter plots. From left to right: functional PCA (first component against component with lowest kurtosis), ICA (non-smoothed) after reduction. **(C)** Estimated densities of the vector of scores with lowest kurtosis using different reduction techniques.

§ 2.9. Discussion

In this chapter, we present a comprehensive overview of the theory behind Functional ICA. We establish the necessary conditions for the existence of a whitening transformation in infinite-dimensional spaces and define the class of whitening operators. Additionally, we have introduced two smoothed functional ICA models. The model discussed in §2.4 harnesses the infinite dimensionality of the estimates to evaluate the extent of non-Gaussianity within the sample, paving the way for optimal classification strategies. Current methods recast on "Gaussianizing" the data using a suitable whitening transformation, regularization, or a functional PCA reduction in order to balance and minimize the kurtosis of the eigenprojections. We have shown the proposed approach has competitive operating features in binary classification problems, both in Gaussian and non-Gaussian settings. The kurtosis operator and particularly its spectral attributes, offer a unique analytical pathway that can bring us nearer to achieving near-perfect accuracy, as elucidated by the Feldman-Hájek dichotomy. In contrast, the model discussed in §2.6 is primarily conceived to provide smooth estimates of non-Gaussian components, potentially enhancing their irreducibility by regulating the roughness of the estimates. Further simulations and applications not reported in this dissertation can be found in [303]. A more advanced/alternative version of this model is presented in the next chapter.

Acknowledgements: The authors wish to thank Prof. Marc Leman for providing valuable comments and Daniel Gost for helping with figures and formatting. (paper [303]); We wish to thank Prof. Jasson Vindas for providing valuable comments and suggestions. (paper [296]); We wish to thank Vadim Nikulin for providing valuable comments and suggestions on our analyses. (manuscript [300]). **Author contributions:** Conceptualization, methodology, software and formal analysis, M.V. and A.M.A.; writing—original draft preparation, M.V.; writing—review and editing, M.V., M.R. and A.M.A.; visualization, M.V.; supervision, A.M.A.; data collection and preprocessing M.V. and M.R. (paper [303]). **Data accessibility:** All R source code and simulation scripts with the results for the full simulation study in [296] are available through the GitHub repository: https://github.com/m-vidal/functional-whitening. (paper [296]); Interactive simulations are available through the Shiny app: https://mvidal.shinyapps.io/fica/. (manuscript [300]) **Conflict of interest:** The authors declare no conflict of interest.

3 | Functional ICA for EEG artifact removal

The work presented in this chapter includes the manuscript:

- [299] Vidal, M. and Aguilera, A. M. (2024). Wavelet thresholding on independent subspace factorizations of spatially indexed wide functional data for robust estimation of cortical activity. *Under review*.

In the field of neurophysiology, electroencephalography (EEG) represents one of the few techniques providing a direct measure of bioelectrical brain activity, as oscillations in excitability of populations of cortical pyramidal cells [309] contribute to variations in the electrical potentials over the scalp. Oscillations are characterized by intrinsic rhythms conventionally grouped into frequency bands, which are by now validated as markers of several neurocognitive phenomena [48]. However, despite the temporal resolution achievable with its high sampling rate, EEG is a technique that suffers from low signal-to-noise ratio. This is mainly due to the fact that the layers of tissue dividing the electrodes from the cortex act as a natural filter attenuating genuine brain activity, resulting in a combination of cortical and artifactual sources in the EEG signal. In addition, brain-related spectral features often overlap with artifactual activity in higher frequency bands, and particularly at lower frequencies most of the variance in the signal is explained by physiological sources outside the brain. For these reasons, analyzing EEG signals can ultimately be viewed as solving a source-separation problem with the goal of estimating brain potentials of interest.

This chapter has a predominant focus on methodology, and can be seen as a bridge between the theoretical and applied/experimental part of the dissertation. Here, we delve into the mathematical and probabilistic principles behind the reconstruction of artifactual activity from EEG signals. In [303], we attempted to provide a first approximation to this problem from a functional data perspective. While the mathematical model (see §2.6) itself is valuable, the application was limited to short time courses and artifacts assumed to be basically smooth. The limited applicability of this model led to a new manuscript three years after its initial publication, wherein more rigorous developments (both theoretical and practical) were addressed. Results are presented in the current chapter. The proposed methodology has been effectively applied in the investigation reported in Chapter 5, where we also delineate a pipeline for pre-processing EEG data during complex motor interactions.

Here, we argue the reconstruction of artifacts is related to the approximation of a function in a Hilbert basis that is a realization of a random variable taking values in a two-domain Hilbert space. A model for sparse optimization based on a fixed-point iteration over the spatial domain and posterior optimization in the temporal domain via wavelet thresholding is discussed under the paradigm of "wide functional data". Two criteria are introduced for selecting wavelet expansion coefficients in scenarios where noise lacks of a precise parametric specification: one based on multiplicative scaling and the other on the entropic NID (ENID), as introduced in Bruni et al. [45]. Through comprehensive numerical simulations and real data analyses of EEG data, we showcase the effectiveness of the proposed methods.

§ 3.1. Introduction

In the analysis of data assumed to be realizations of a random variable X taking values in a separable function space endowed with Hilbert structure, certain considerations are frequently made. These include assuming finite second moments and the Hilbert-Schmidt boundedness of the covariance operator, along with considering independent and identically distributed realizations of X. In practice, however, there are scenarios in which the sample covariance operator poses certain challenges: it is often not injective, and low sample sizes can lead to eigenvalues that are biased upwards, potentially lacking an upper bound, especially in high-dimensional settings. Furthermore, assuming that all realizations of X are independent is not always justifiable, especially when there exists repeated measures in time or space.

Let $L^2_{\mathcal{I}}$ denote the space of square integrable functions defined over a real compact interval $\mathcal{I} \subset \mathbb{R}$, equipped with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. Consider $\{X_i\}_{i=1}^m$ to be m realizations of a random variable $X = \{X(t), t \in \mathcal{I}\}$ taking values in $L^2_{\mathcal{I}}$ with $\mathbb{E}\|X\|^2 < \infty$, i.e., the theoretical covariance operator is here assumed to exist [125]. Since separability holds, for any orthonormal basis $(e_j)_{j\in\mathbb{N}}$ of $L^2_{\mathcal{I}}$, we can express these realizations of X as $X_i \approx \sum_{j=1}^q Z_{i,j} e_j$, where $Z_{i,j} = \langle X_i, e_j \rangle$ are the expansion coefficients. This property follows directly from the Projection Theorem which constitutes the fundamental tool for functional data

analysis [125, 180]. Spaces of smoother functions, such as reproducing kernel or Sobolev spaces, are often preferred to better capture the underlying topological features of X.

While under mild conditions (i.e, for a reasonable $m < q$) the sample covariance function of X_i is still computable, when $m \lll q$, it becomes ill conditioned and computationally intractable. In such extreme cases, we categorize data with these characteristics as *wide functional data*. One can envision these data as functions that exhibit fine detail throughout their domain, which may not necessarily be related to noise. Importantly, we assume our functions have finite energy in $L^2_{\mathcal{T}}$, i.e. they are Lesbegue square-integrable, although in practice wide data integration might get exhausted in common computers due to memory requirements. Therefore, alternative approaches are needed to effectively explore data over $\mathcal{T}$.

In the current context, our attention is drawn to a specific type of data that may not necessarily represent independent realizations of X. Dependencies could potentially arise in a parallel domain, such as space. These kind of data can be encountered in various scenarios, with a notable occurrence in neuroscience, e.g., during the pre-processing of artifacts or the analysis of long-term monitoring studies involving individual patients. Note that, by assuming dependencies in a secondary domain, the covariance matrix of $Z^\top = (Z_{j,i})_{q\times m}$ becomes a subject of interest.

Although not restricted to, here we consider a two-domain random variable

$$\mathcal{X} = \{X(t,s) : t \in \mathcal{T}, s \in \mathcal{S}\}, \tag{3.1}$$

taking values in $L^2_{\mathcal{T}\times\mathcal{S}}$. We address the problem of approximating a single realization of $\mathcal{X}$ denoted as $X_k \equiv X(t,s_k)$ $(k = 1,\dots,p)$, where s_k are indexed locations fixed on a compact plane (or other manifold structures), where functions in $\mathcal{T}$ emerge.

Now, suppose that X_k is observed at n equidistant points and has the form

$$X_{k,i} = \Phi(t_i,s_k) + \mathcal{B}(t_i,s_k) + \epsilon_{k,i} \quad (i = 1,\dots,n), \tag{3.2}$$

where Φ is an artifactual function with non-Gaussian components in $\mathcal{S}$ and sparse in $\mathcal{T}$, $\mathcal{B}$ is anisotropic and follows an unknown Gaussian distribution in $\mathcal{T}$ and $(\epsilon_{k,i})$ is a noise assumed to be independent in space, with each column following a (possibly correlated) normal distribution. Note that both Φ and $\mathcal{B}$ are random functions taking values in $L^2_{\mathcal{T}\times\mathcal{S}}$. In this paper we are interested to estimate $\mathcal{B}$ considering estimates of Φ when $p \lll n$. This problem is genuinely motivated by the analysis of contaminated EEG data, where muscular artifacts (here represented by Φ) are exacerbated by volume conduction, leading to challenges in isolating the brain activity of interest (i.e. $\mathcal{B}$).

It has been argued that in the analysis of large-scale data, sparsity is meant to be reinforced to detect interpretable low-dimensional structures with statistical significance [37]. Sparsity is an appealing property since it compressively encodes relevant information in a few entries of the expansion coefficients of a function. Statistically, this is a particular way of being non-Gaussian, reason why sparsity shares a close connection with independent component techniques [57]. However, while sparsity implies a form of non-Gaussian behavior, it does not guarantee independence [72] (independent variables can indeed be present in datasets that do not exhibit sparsity). Here, we propose a model (see Figure 3.1A) that consists of the analysis of subspaces emerging from the spatial structure of $X = (X_{k,i})_{p\times n}$ and spanned by independent components (§3.2). Wavelet techniques are then applied to find a sparse approximation of Φ among a variety of spatial factorizations of X. We therefore review some concepts of wavelet theory and derive a suitable method for multiresolution analysis to apply wavelet denoising techniques (§3.3). Two non-parametric methods for the selection of expansion coefficients are then introduced (§3.4). Both of them do not directly depend on the variance of the detail coefficients, rather on the global regularity properties of the original signal, allowing the discrimination of coefficients that significantly reflect the signal of interest.

The outlined model can be viewed as an alternative method to the functional independent component analysis (ICA) model in [303], offering the advantage of circumventing the challenging task of estimating large covariance matrices. Additionally, it eliminates the need for an excessive number of B-spline knots, providing more versatility for denoising and the potential to effectively handle a large number of covariates, all within a reasonably fast computation time. It also extends prior wavelet ICA schemes that take into account the spatial structure of the data; see [10] and references therein. We also introduce a method for simulating EEG data (§3.6) and provide numerical evidence illustrating why current pre-processing practices may not always be reliable. Finally, some real data analyses are discussed.

Notation

For a particular matrix, take $A \in \mathbb{R}^{p\times p}$, a selection of its vector-rows is denoted as $A_{1:p,\bullet}$ and vector-columns as $A_{\bullet,1:p}$. A single one of its vector rows or columns is written as $A_{1,\bullet}$ or $A_{\bullet,1}$, respectively. The operator $\mathfrak{C} : \mathbb{R}^p \to \mathbb{R}^p$ performs cumulative sum. $\mathfrak{O}$ stands for the numerical order. By some abuse of notation, $\|\cdot\|$ will denote the L^2 and Euclidean norm.

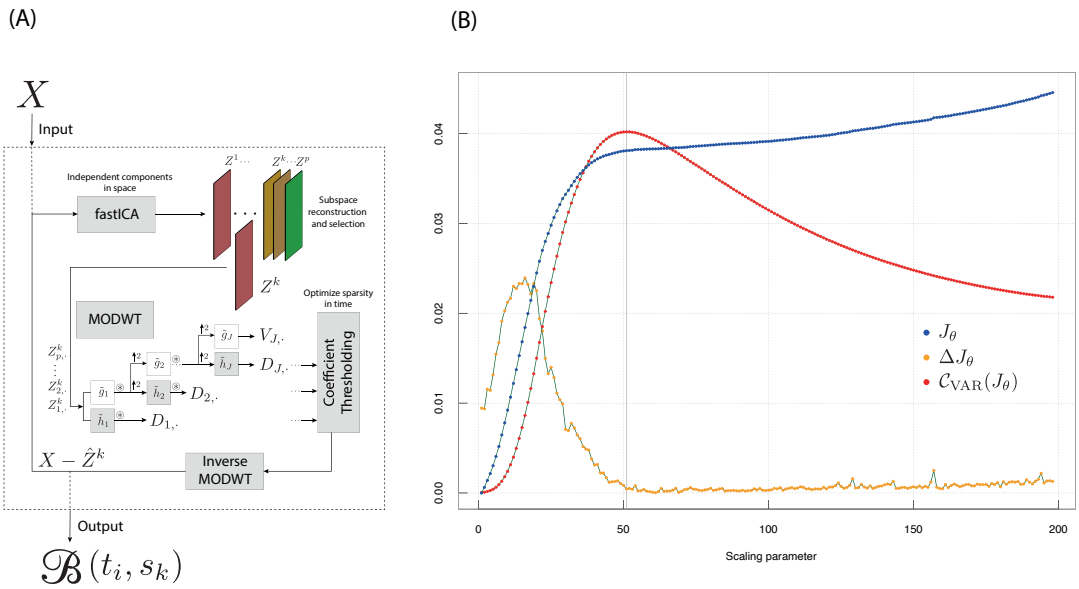

Figure 3.1: Wavelet functional ICA model for artifact reduction. **(A)** Main model representation **(B)** Exemplification of multiplicative scaling (J_θ was scaled for the sake of its representation).

§ 3.2. Factorization into subspaces spanned by spatial ICs

The spatial independent component model applied to wide data

Assume $X = (X_{k,i})_{p\times n}$ with $p \lll n$ can be expressed as linear transformation of an unobserved matrix $S \in \mathbb{R}^{p\times n}$ of mutually spatially independent components, i.e. the probability distribution of $S_{1:k,\bullet}$ factorizes as $P_{S_{1:k,\bullet}} = \prod_{k=1}^{n} P_{S_{k,\bullet}}$. The spatial independent component (spatial IC) model can be expressed as

$$X = \overline{X} + AS \tag{3.3}$$

where $A \in \mathbb{R}^{p\times p}$ is an unknown invertible square matrix containing vectors of spatial weights for each $S_{k,\bullet}$ and $\overline{X} = X1_n/n$ is the spatial mean. The task of spatial ICA is therefore to infer the matrix S from X. Due to the variance indeterminacy in the model 3.3, a common pre-processing step is to "whiten" or map to orthogonality X. This transformation has some desirable properties to fasten the convergence of some ICA algorithms. A whitening transformation $\Psi(\cdot) : \mathbb{R}^{p\times n} \to \mathbb{R}^{p\times n}$ is derived from the factorization $\Sigma^{-1} = \Psi\Psi^\top$ where $\Sigma^{-1} = \{(X - \overline{X})(X - \overline{X})^\top/p\}^{-1}$ is the covariance of the transpose of the original data. As a result, $\tilde{X} = \Psi(X - \overline{X})$ becomes spatially decorrelated (or isotropic, i.e., the spatial covariance matrix of $\tilde{X}$ is I_p). Let us note that Ψ is not uniquely determined and many whitening procedures can be devised due to the rotational freedom the transformation possesses; see [296] for further details. This property is the basic principle of ICA.

Under the above assumptions and assuming $\Psi = \Sigma^{-1/2}$, the model 3.3 can be straightforwardly reformulated as

$$\tilde{X} = \Psi(X - \overline{X}) = \Psi(AS) = (AA^\top)^{-1/2}AS = \tilde{A}S. \tag{3.4}$$

Given that $\tilde{X}$ is isotropic and S has independent components, the matrix $\tilde{A}$ is orthogonal. At this point, we might further suppose that A is orthogonal in 3.3 when assuming $\overline{X} = 0$. Therefore, one can only recover S up to a permutation and the sign [64], which means there exists a demixing matrix $W \in \mathbb{R}^{p\times p}$ satisfiying $WA = DP$, where D is a diagonal matrix with diagonal elements being 1 or -1 and P is a permutation matrix.

Fixed-point iteration on isotropic data using non-linearities

After pre-whitening, the aim is to estimate a demixing matrix for $\tilde{X}$, we denote by W^*. A fixed-point iteration based on a non-Gaussian optimization is then used. This iteration is at the core of the popular FastICA algorithm [127], which is presented in its basic form in the next lines.

Let $g : \mathbb{R} \to \mathbb{R}$ be a function measuring non-Gaussianity that operates elementwise to the entries of a scalar vector. We call the function g nonlinearity, which is usually defined as $g(\cdot) = G'(\cdot)$, where G is twicely differentiable nonlinear and non-quadratic function with $G(0) = 0$. Typical examples of $G(\cdot)$ include $x^4/4$ (*kurtosis*), $-\exp(-x^2/2)$ (*gauss*) or $\log\cosh(x)$ (*tanh*). All of these nonlinearities are smooth and

even, and can be bounded by a polynomial function. The optimization problem in FastICA summarizes in the contrast function

$$\underset{\|\mathtt{w}\|=1}{\mathcal{J}}(\mathtt{w}) = n^{-1} G\left(\mathtt{w}^\top \tilde{X}\right) 1_n, \tag{3.5}$$

where $\mathtt{w}$ is a vector of p entries on the unit hypersphere $\mathbb{S} \equiv \{\mathtt{w} \in \mathbb{R}^p \mid \|\mathtt{w}\| = 1\}$. Under Langrangian conditions, the extreme value of 3.5 can be obtained via solutions to the equation $\tilde{X} g(\tilde{X}^\top \mathtt{w}) + \beta \mathtt{w} = 0$. This problem is solved using the following iterative method:

1. Chose an initial guess for $\mathtt{w} \in \mathbb{S}$.
2. Iterate

$$\begin{aligned} \mathtt{w}^+ &\leftarrow \tilde{X} g(\tilde{X}^\top \mathtt{w}) - n^{-1} \mathtt{w} g'(\mathtt{w}^\top \tilde{X}) 1_n \\ \mathtt{w}^+ &\leftarrow \frac{\mathtt{w}^+}{\|\mathtt{w}^+\|}, \end{aligned} \tag{3.6}$$

   until convergence.
3. An estimate $\hat{\mathtt{w}} = \mathtt{w}^+$ of $W^*_{1,\bullet}$ is obtained.

Note the above algorithm only extracts one demixing vector. Therefore, a sequential or deflationary orthogonalization procedure is usually conducted to estimate the rest of vectors. One can add the following orthogonal constrain in between the two steps in 3.6:

$$\mathtt{w}^+ = \mathtt{w}^+ - \sum_k (\mathtt{v}_k \mathtt{v}_k^\top) \mathtt{w}^+, \tag{3.7}$$

where $\mathtt{v}_k$ are previously obtained demixing vectors ("deflated" observations). Each extracted $\hat{\mathtt{w}}$ in the fashion described above is a column of W^*, and $S = W^* \Psi (X - \overline{X})$ can be finally obtained. This algorithm is called one-unit FastICA as it extracts one demixing vector at a time in a similar vein to projection pursuit techniques.

Subspace reconstruction and expansion

Once the matrix W^* is obtained, finding the inverse of $W = W^* \Psi$ gives the mixing matrix A to recover X. Given that $X = AS$ assuming $\overline{X} = 0$, one can perform the following factorization of X as

$$X = \{\underbrace{A_{\bullet,1} S_{1,\bullet}}_{Z^1 \in \mathbb{R}^{p \times n}} + \cdots + \underbrace{A_{\bullet,k} S_{k,\bullet}}_{Z^k} + \cdots + \underbrace{A_{\bullet,p} S_{p,\bullet}}_{Z^p}\}. \tag{3.8}$$

where $\{Z^k\}_{k=1}^p$ are a set of subspaces spanned by each $S_{k,\bullet}$. In the following lines, we discuss some of their statistical properties. All Z^k's are not isotropic while they are a.s. spatially *irreducible* [111]: i.e., for any Z^k, the matrix A in the model $Z^k = AS$ vanishes, indicating that Z^k is not separable into lower independent components. Having singular covariance matrix is a necessary but not sufficient condition for irreducibility; however, non-Gaussian random vectors often exhibit irreducibility under such conditions. Additionally, the Z^k's are pairwise orthogonal (i.e., $Z^k Z^{k'\top} = 0_p$ for all $k \neq k'$) and any linear combination of them, say $Z^{k_1} + \cdots + Z^{k_m}$ $(m < p-1)$, is orthogonal to the rest of Z^ks and of its complement. It is worth noting that applying wavelet denoising directly to $S_{k,\bullet}$ would compromise these properties, thus failing to ensure the level of independence that spatial ICA provides.

§ 3.3. Wavelet approach

A critical step before applying wavelet techniques is the selection of Z^k's, which is discussed in Sections 3.6 and 3.7. Assume some of these Z^k's, or linear combinations of them, correspond to the function in 3.2 containing artifactual components. The next step involves approximating each of the selected Z^k's using a suitable Hilbert basis in the temporal domain with the aim of optimizing the irreducibility (consequently the independence) of the selected components. We now recall some concepts of wavelet theory to later present our wavelet thresholding approach.

Preliminaries: wavelet transform and multiresolution analysis of L^2

A wavelet on $\mathbb{R}$ is an oscillatory function $\psi \in L^2_{\mathbb{R}}$ satisfying $\int_{-\infty}^{+\infty} \psi(t)\mathrm{d}t = 0$ (has zero average), $\|\psi\| = 1$ (is normalized at 1), centered in the neighbourhood of $t = 0$ and decaying to 0 as $t \to \pm\infty$. Other desirable properties of ψ include regularity, having d vanishing moments, i.e., $\int_{-\infty}^{\infty} t^p \psi(t)\mathrm{d}t = 0, d \in \mathbb{N}_0$ (the ability to represent polynomial functions) or compact support (for further details, see [71]). The continuous wavelet transform of a function $f \in L^2_{\mathbb{R}}$ with respect to ψ is the linear functional

$$[\mathcal{W}_\psi f(t)](b,a) = \langle f(b+ax), \psi^*(x)\rangle = \left\langle f(t), \frac{1}{a}\psi^*\left(\frac{t-b}{a}\right)\right\rangle \quad (b,a) \in \mathbb{R}\times\mathbb{R}_+, \tag{3.9}$$

where a is a scale parameter and b a translation parameter and ψ^* is the complex conjugate of ψ. The above transform it is usually performed in a dyadic fashion (a and b are based on powers of 2), which gives rise to the notion of multiresolution analysis (MRA).

Definition 6. *A function $\phi \in L^2_{\mathbb{R}}$ is a scaling function of a MRA of $L^2_{\mathbb{R}}$ if the following conditions hold:*

1. *The family of translated functions $\{\phi(\cdot - \delta)\}_{\delta\in\mathbb{Z}}$ is a complete orthonormal system of $L^2_{\mathbb{R}}$.*
2. *The linear spaces $\mathscr{H}_0 = \mathrm{span}\{\phi(\cdot-\delta)\}, \ldots, \mathscr{H}_j = \mathrm{span}\{\phi(2^j\cdot - \delta)\}, \ldots$ are nested; i.e, $\mathscr{H}_{j-1} \subset \mathscr{H}_j, \forall j \in \mathbb{N}$.*
3. *$\overline{\cup_{j\geq 0}\mathscr{H}_j} = L^2_{\mathbb{R}}$ (the closure of $\cup_j \mathscr{H}_j$ is dense in $L^2_{\mathbb{R}}$, i.e. $\cap_j \mathscr{H}_j = \{0\}$).*

Note that here, the index j indicates the scale (or level of resolution) and, as it increases to ∞, the precision of the approximation increases too. Further, observe that each subspace $\mathscr{H}_j$ consists of functions that are piecewise constant over intervals of exactly twice the length of those for $\mathscr{H}_{j-1}$. From points (i) and (ii) it immediately follows that the the functions $\phi_{j,\delta}(x) = 2^{j/2}\phi(2^j x - \delta), \delta \in \mathbb{Z}$, form an orthonormal basis of the space $\mathscr{H}_j, j \in \mathbb{N}$, as these spaces are just scalings of $\mathscr{H}_0$.

The multiresolution approximations of a function $f \in L^2_{\mathbb{R}}$ are given by the projections P_j on to $\mathscr{H}_j$, $P_j f = \sum_\delta \langle \phi_{j,\delta}, f\rangle \phi_{j,\delta}$, with $\langle \phi, f\rangle = \int_{-\infty}^{+\infty} f(t)\phi^* \mathrm{d}x$. The range of the difference between two successive approximations $P_{j+1}f - P_j f$ corresponds to the orthonormal complement of the space $\mathscr{H}_j$ in $\mathscr{H}_{j+1}$. The information contained in these subspaces are relevant for reconstructing f since

$$\mathscr{H}_j = \mathscr{H}_0 \oplus \left(\bigoplus_{\ell=0}^{j-1} \mathscr{H}_\ell^\perp\right), \tag{3.10}$$

where $\mathscr{H}_\ell^\perp \equiv \mathscr{H}_{\ell+1} \ominus \mathscr{H}_\ell$ with $\mathscr{H}_\ell \perp \mathscr{H}_{\ell'}, \ell \neq \ell'$. Thus, one would like to find basis functions that span the spaces $\mathscr{H}_\ell^\perp$. Using ϕ, it is possible to construct via standard ways a corresponding wavelet function ψ, so that $\{\psi(\cdot-\delta)\}_{\delta\in\mathbb{Z}}$ is an orthonormal basis of $\mathscr{H}_0^\perp$. Similarly, $\psi_{\ell,\delta}(x) = 2^{\ell/2}\psi(2^\ell x - \delta), \delta \in \mathbb{Z}$ will form an orthonormal basis for $\mathscr{H}_{j>0}^\perp$, as these spaces are mutually orthogonal and obtained (from each other) by scalings. We note that the functions ϕ and ψ are roughly referred to as father and mother wavelet, respectively.

Under the above considerations, one can expand a function $f \in L^2_{\mathbb{R}}$ as the convergent series expansion

$$f = \sum_{\ell\in\mathbb{Z}}\sum_{\delta\in\mathbb{Z}} d_{\ell,\delta}\psi_{\ell,\delta}, \quad \|\hat{f}\|^2_{L^2_{\mathbb{R}}} = \sum_{\ell,\delta\in\mathbb{Z}} d^2_{\ell,\delta} < \infty, \tag{3.11}$$

where $d_{\ell,\delta} = \langle f, \psi_{\ell,\delta}\rangle = \int_{-\infty}^{\infty} f(x)\psi^*_{\ell,\delta}(x)\mathrm{d}x$ are the wavelet coefficients of f in terms of the wavelet basis. The relation of the wavelet transform in 3.9 with the wavelet coefficients is then $d_{\ell,\delta} = 2^{-\frac{\ell}{2}}\mathcal{W}_\psi f\left(\delta 2^{-\ell}, 2^{-\ell}\right)$. MRA has previously been used in functional data studies to estimate a function-on-function linear regression model, elucidating the relationship between lupus severity and stress levels in patients with this autoimmune disease [5].

Discrete non-decimated multiresolution analysis

As noted above, the dilation and translation parameters a, b are assumed to vary continuously over $\mathbb{R}$ (with the constraint $a \neq 0$). If these parameters take discrete values, then the transform is called discrete wavelet transform (DWT). Here, the aim is to perform MRA on $\mathbf{z}_i \equiv Z^k_{i,\bullet}$, for each i and a fixed k, using Mallat's pyramidal algorithm [186] but without using dyadic subsampling (decimation). This corresponds to the maximal overlap discrete wavelet transform (MODWT) [232, 307], in which the number of coefficients at each resolution level end being n. The aforementioned technique offers several advantages, including

its insensitivity to vector length or circular translations as well as improved frequency localization due to the redundancy introduced by the overlap, resulting in enhanced denoising capabilities and better ability to mitigate the boundary effects commonly encountered when decimation is conducted.

We use an even length scaling filter $\{g_l\}_{l=0}^{L-1}$ and wavelet filter $\{h_l\}_{l=0}^{L-1}$ with $L \leq n$, that are discretely compactly supported filters of the Dauchebies class [71, Chapter 6]. By definition, these filters sum zero, have unit square norm and are orthogonal to its even translations. As in [307], filters are additionally selected to be quadrature mirror, ensuring that they meet the orthogonality condition $h_l = (-1)^l g_{L-l-1}$ or $g_l = (-1)^{l+1} h_{L-l-1}$ for $l = 0, \ldots, L-1$. Furthermore, filters are normalized as follows: $\tilde{g}_l = g_l/\sqrt{2}$ and $\tilde{h}_l = h_l/\sqrt{2}$. This procedure allows to preserve the energy when the filters are upsampled by 2^{j-1}, which is achieved by padding out zeros between successive elements of the filter, that is,

$$\tilde{h}_j \equiv [\tilde{h}_0, \underbrace{0, \ldots, 0}_{2^{j-1}-1 \text{ zeros}}, \tilde{h}_1, \underbrace{0, \ldots, 0}_{2^{j-1}-1 \text{ zeros}}, \ldots, \tilde{h}_{L-2}, \underbrace{0, \ldots, 0}_{2^{j-1}-1 \text{ zeros}}, \tilde{h}_{L-1}].$$

Observe that this operation is akin to dilating the wavelets, thereby enabling the formation of a multiresolution analysis without decimation. If $V_{0,i} \equiv \mathbf{z}_i$, the jth-level of the pyramid algorithm is then given by the following circular filterings:

$$D_{j,i} = \sum_{l=0}^{L-1} \tilde{h}_l V_{j-1,(i-2^{j-1}l) \bmod n}, \quad V_{j,i} = \sum_{l=0}^{L-1} \tilde{g}_l V_{j-1,(i-2^{j-1}l) \bmod n},$$

where $D_{j,i}, V_{j,i}$ represent the jth-level wavelet and scaling coefficients, $i = 0, 1, \ldots, n-1$ the number of samples and mod the modulo operator. The reconstructing coefficients are then

$$V_{j-1,i} = \sum_{l=0}^{L-1} \tilde{h}_l D_{j,(i+2^{j-1}l) \bmod n} + \sum_{l=0}^{L-1} \tilde{g}_l V_{j,(i+2^{j-1}l) \bmod n}.$$

The coefficients of both MODWT and DWT share the same nominal frequency band at each level of resolution. Nevertheless, the MODWT is not an orthonormal transform of $\mathbf{z}_i$ and if n is an integer multiple of 2^J, the MODWT has a computational complexity of $\mathcal{O}(n \log_2 n)$ multiplications whereas the DWT solely of $\mathcal{O}(n)$.

§ 3.4. Wavelet thresholding for non-necessarily white noise

Let $D = (D_{j,i})_{J \times n}$ be the matrix that contains the coefficients of the MODWT performed on $\mathbf{z}_i$. Considering that the wavelet transform of a well-behaved function typically exhibits sparsity, only a small portion of the wavelet coefficients will have significant values, while the rest will be relatively small and can be considered negligible. Therefore, when a coefficient $D_{j,i}$ is relatively small, it is justifiable to treat it as predominantly noise and set it to zero; conversely, if it is significantly large, retaining it is a reasonable choice. This corresponds to the shrinking policy/rule called hard thresholding [82, 83] and defined here by the estimator $\eta_H(D_{j,i}, \alpha_j) = D_{j,i}\mathbb{1}\{|D_{j,i}| > \alpha_j\}$ for some threshold values α_j. In this paper, our focus centers on the aforementioned estimator in the context of the model 3.2. An intriguing aspect lies in the estimation of Φ, since the conventional reliance on the assumption of a white noise for the selection of a suitable threshold no longer holds. Two non-parametric solutions to chose level-dependent thresholds are presented in the following lines.

Multiplicative scaling of Walden's MODWT level-dependent threshold

Walden [307] proposed a robust alternative to the so-called universal threshold $\alpha \equiv \{2\sigma_\epsilon^2 \log(n)\}^{1/2}$, where σ_ϵ is the standard deviation of the detail (or finest level) coefficients, $D_{1,\bullet}$. It is therefore assumed that $D_{1,\bullet}$ are noise dominated. While this approach has been proven asymptotically optimal [82], it can erroneously set to zero certain coefficients. The common estimator for σ_ϵ is then replaced in [307] by

$$\hat{\sigma}_{\text{MAD}} \equiv \frac{\text{median}\left\{\left|D_{1,0}\right|, \left|D_{1,1}\right|, \ldots, \left|D_{1,\frac{n}{2}-1}\right|\right\}}{0.6745}, \tag{3.12}$$

where 0.6745 is a rescaling factor, so that the correct variance can be returned in case of Gaussian white noise. The resultant threshold is typically levelwise adjusted as $\alpha_j \equiv \{2\sigma_j^2 \log(n)\}^{1/2}$ with $\sigma_j^2 = \sigma_\epsilon^2/2^j$, and

adapted to the MODWT considering the estimator $2^{1/2}\hat{\sigma}_{\text{MAD}}$, where the scaling factor $2^{1/2}$ accounts for the relation $D_{1,i} = 2^{1/2}D_{1,2i+1}$.

The current approach may indeed be effective in optimizing the reduction of the error term in 3.2. However, it may not be as efficient in separating latent components of Φ from $\mathcal{B}$. Some packages offer the possibility to heuristically tailor the denoising level by either scaling α_j or by incrementing the median value in 3.12. However, finding a good trade-off between both parameters is not an easy task. Here, we concentrate on estimating a multiplicative scaling factor for Walden's MODWT level-dependent threshold via cross-validation schemas without altering 3.12.

For $\theta \in \mathbb{R}_+$, consider $\alpha_j^\theta \equiv \theta\{2\sigma_j^2 \log(n)\}^{1/2}$ where the estimator for σ_j^2 is given by $2^{1/2-j}\hat{\sigma}_{\text{MAD}}$. We then aim at finding a θ that maximizes the cumulative variance of the loss function

$$J_\theta \equiv \frac{1}{p}\sum_{k=1}^{p}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(Z_{k,i}^k - \hat{Z}_{k,i}^{k,\theta}\right)^2\right\}^{1/2}, \tag{3.13}$$

where $\hat{Z}^{k,\theta}$ is a "denoised" Z^k using α_j^θ for different values of θ. The optimization problem is then defined as

$$\text{argmax}_{\theta\in\mathbb{R}_+}\mathfrak{C}_{\text{VAR}}(J_\theta), \tag{3.14}$$

where $\mathfrak{C}_{\text{VAR}}(\mathtt{x}) = (n-1)^{-1}\{\mathfrak{C}(\mathtt{x}^2) - n^{-1}\mathfrak{C}(\mathtt{x})^2\}$ is the cumulative variance calculated for any $\mathtt{x} \in \mathbb{R}^n$.

We now discuss the rationale underlying this approach. The function J_θ is assumed monotonically increasing and grows fast as we progressively move away from Z^k by increasing θ. Nonetheless, J_θ starts growing slowly as $\hat{Z}^{k,\theta}$ becomes sparser. As a result, $\mathfrak{C}_{\text{VAR}}(J_\theta)$ reaches a peak and decreases along with the first derivative of J_θ converging to 0 (indicating that no significant variance is added to the model, see Figure 3.1B). This peak (depicted by a grey vertical line) can be interpreted as a stability/fixed point where the behavior of J_θ starts changing. Investigating the vicinity of this maximum can lead to optimal outcomes, as shown by our simulations.

Entropic normalized information distance

We now present an alternative coefficient selection approach based on a paper by Bruni et al. [45]. There, it is assumed that the expansion coefficients originate from two separate and independent sources, namely, the most "representative" and the "less representative" coefficients. This method boasts a distinctive advantage: there is no prerequisite knowledge needed about the statistical nature of the noise, a requirement often taken into account in other approaches [18]. The method is established within an information-theoretic framework, introducing a novel measure called entropic normalized information distance (ENID), which is fundamented the so-called Vitanyi's normalized information distance and shaped by the notion of differential entropy. Although ENID is formulated in terms of the detail coefficients, it can be applied, without loss of generality, to coarser levels or the whole coefficient set, as proposed here.

Consider the vectorization of the matrix D, $\mathtt{d}_j = \text{vec}(D)$ $(j = 0, \ldots, p \times N)$. Suppose that the absolute value of $\mathtt{d}_j$ are rearranged in decreasing order and normalized on $[0, 1]$, i.e, $\tilde{\mathtt{d}}_j = \{1 \geqslant \tilde{\mathtt{d}}_{p\times N-1} \geqslant \cdots \geqslant 0\}$. Then, for all $\beta \in \{1, \ldots, p \times N - 1\}$, the ENID is defined as

$$\text{ENID}_\beta = \frac{\left|\left|E_{[0,1]}\right| - \min\left\{\beta\left|E_{[0,\beta]}\right|, (1-\beta)\left|E_{[\beta,1]}\right|\right\}\right|}{\max\left\{\beta\left|E_{[0,\beta]}\right|, (1-\beta)\left|E_{[\beta,1]}\right|\right\}}, \tag{3.15}$$

where $E_{[\cdot,\cdot]}$ denotes the differential entropy at a predefined interval. Numerical schemes for the approximation of E are provided in [45] §2.3. The index entry that realizes the minimum of ENID_β allows to identify in $\tilde{\mathtt{d}}_j$ (when not normalized) the wavelet coefficient used for thresholding. Then, in $\alpha_j \equiv \{2\sigma_j^2 \log(n)\}^{1/2}$ we define the estimate for σ_j^2 as $|\mathtt{d}_j|/2^j$ where $\mathtt{d}_j$ represents the wavelet coefficient related to the minimum ENID_β.

Remark 2. *In order to fasten computation when handling wide data, one can model $\tilde{\mathtt{d}}_j$ as a $L^2_{\mathcal{J}}$ function using some basis functions representation. Given that the approximation coefficients maintain the probabilistic properties of the functional approximation (due to the existence of an isometric isomorphism between $L^2_{\mathcal{J}}$ and the coefficient space), E can be computed on them. Note that ENID will be drastically reduced to a few data points, therefore some interpolation technique has to be used to identify the wavelet coefficient of interest.*

In essence, ENID gauges the dissimilarity between two coefficient sets as the difference between the overall complexity and the less complex of the two subsets concerning the regularity of the signal, as reflected by the decreasing rearrangement of the coefficients. When ENID reaches a minima, this indicates that the entropy of one coefficient set is close to the overall entropy and therefore the remaining one has little impact on the randomness of the other.

§ 3.5. Model summary

An overview of the proposed algorithm is depicted in Figure 3.1.

Input: a wide data matrix $X \in \mathbb{R}^{p\times n}$ ($p \lll n$). *Output*: $\mathcal{B}(t_i, s_k) \in \mathbb{R}^{p\times n}$.

1. Perform spatial ICA (e.g. using FastICA) on X and derive $Z^1, \ldots, Z^p$ to make a selection of them.
2. For each selected Z^k perform the MODWT to each of its rows. Extract the corresponding matrices of coefficients.
3. Use one of the techniques described in §3.4 to conduct hard thresholding on the wavelet coefficients.
4. Invert all MODWT transforms and subtract the denoised Z^k's from X.
5. Repeat until no artifactual components are left.

The above algorithm is not restricted to a fixed number of iterations (in our data analyses we consider a single one); rather, it can be performed until the residual signals meet a predefined criterion.

Proposition 15. *Let $Z^k = A_{\bullet,k}S_{k,\bullet}$ be the expansion of an arbitrary spatial IC and $\mathcal{V}(Z^k)$ a non-linear mapping that uniformly denoises the rows of Z^k. Then, $\mathcal{V}(Z^k) = A_{\bullet,k}\mathcal{V}(S_{k,\bullet})$.*

Proof. The proof is trivial since in the inner product $A_{\bullet,k}S_{k,\bullet}$, each entry of the vector $A_{\bullet,k}$ is multiplied by each of the corresponding rows of $S_{k,\bullet}$, therefore weighting p times the vector $S_{k,\bullet}$. If, however, $\mathcal{V}(\cdot)$ applies non-uniformly to each row of Z^k, the above relation does not hold. ∎

We observe that Proposition 15 allows to notably reduce the computational cost by applying the denoising procedures to $S_{k,\bullet}$ instead of Z^k. While there seem to be no apparent disadvantages in directly denoising $S_{k,\bullet}$, note that the spatial weights in $A_{\bullet,k}$ do not correspond to those of $\mathcal{V}(S_{k,\bullet})$ and $\|\mathcal{V}(S_{k,\bullet})\| < 1$. Ultimately, the problem reduces to whether selecting or not an orthonormal basis for projection, since the proposed procedures for denoising are scale invariant. For the subsequent simulations and analyses, we adhere to our original approach.

§ 3.6. Numerical simulations

In this section, various numerical simulations are conducted to show the performance of the proposed methods. The artificial data is generated from the following model:

$$X_{ij} = \sin(\tau_{1;i,j}) + \tilde{A}_{i,j}\cos(\tau_{2;i,j}) + \tilde{K}_{i,j} + \epsilon_{i,j} \quad (i = 1, \ldots, n; j = 1, \ldots, p), \tag{3.16}$$

where, for all i, $\tau_{1;j} = \pi\mathfrak{C}(|\mathtt{x}_j|), \mathtt{x} \sim \mathcal{N}(0, 0.02)$; $\tilde{A}_{i,j} = A_{i,j}\mathtt{c}_{1;j}$, where $\mathtt{c}_{1;j}$ is a vector of weights of p equidistant observations in $[-90, 90]$ and $A_{\bullet,j}$ contains log-normal densities ($\mathcal{N}_{\log}(3, \sigma), \sigma \sim \mathcal{U}[0.2, 0.8]$) generated in varying intervals ranging from 1 in the lower bound at the extremals; $\tau_{2,j} = 2\pi\mathfrak{C}(|\lfloor \mathtt{x}_j \rfloor|)/12500, \mathtt{x} \sim \mathcal{N}(80, 180)$, where $\mathtt{x}_j$ changes of value (phase) at each 250 observations; $\tilde{K}_{i,j} = K_{i,j}\mathtt{c}_{2;j}$ where $\mathtt{c}_{2;j}$ is a vector of weights with p equidistant entries and $K_{\bullet,j}$ are linear combinations of two artifacts generated from modified functions detailed in [82] (artifact 1 "Bumps"; artifact 2 "Doppler"); $\epsilon_j \sim \mathcal{N}(0, 0.05)$. Note that the vectors $\mathtt{c}_{1;j}, \mathtt{c}_{2;j}$ are used as spatial weights of current spiking oscillations and artifacts. For artifact 1, $\mathtt{c}_{2;j}$ is a set of p equidistant observations in $[15, 5]$ and for artifact 2, $\mathtt{c}_{2;j} \in \{0, 0, \ldots, 3\}$. In this paper, we consider $p \lll n$ to be defined by $2^J = n, J \in \mathbb{N}_0$ with $p \leqslant 2\log_2(n)$, thus we take $n = 2^{15}$ in our simulations.

The data generated by model 3.16 bears certain similarities to multichannel EEG signals. In EEG, brain signals are typically aperiodic and exhibit intricate variability in amplitude characterized as local bursts with exponential decay [49]. These characteristics are effectively captured by the phase configurations of the trigonometric terms in 3.16 and their respective scalings using the log-normal density generator. Contamination of EEG brain signals (by blinks, muscular activity, cable movements...) is roughly represented by the artifacts contained in $\tilde{K} = \{\tilde{K}_{i,j}\}$. Artifact 1 appears at all channels with modulated amplitudes that become less pronounced for a higher p, eventually merging with the large scale processes (Figure 3.2A - first row). This artifact is sparse in time, strongly correlated in space/time while assumed independent of the rest of processes. Instead, artifact 2 appears at one single channel, thus, it is uncorrelated in space and not assumed spatially independent, at least within the first regions of its temporal domain (Figure 3.2A - second row).

Figure 3.2: Simulation Results **(A)** Estimation of Artifact 1 and 2 in Simulation 1 (l4), displaying three spatial components. **(B)** Estimation of Artifact 3 in Simulation 3 (l6).

Our simulations are assessed using four key performance metrics described below. The first, and the simplest one, is a multichannel RMSE defined as

$$\text{MRMSE} = \frac{1}{p}\sum_{k=1}^{p}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{G}_{k,i} - \hat{X}_{k,i}\right)^2\right\}^{1/2}, \tag{3.17}$$

where $\mathcal{G}$ is the ground truth signal and $\hat{X}$ is an estimate of $\mathcal{G}$ after the model is applied. To analyze the level of residual noise in $\hat{X}$ and the degree of similarity with $\mathcal{G}$, we employ a variance- and correlation-based signal-to-noise ratio (SNR) measures adapted from [286]. Here, the multichannel SNR variance measure is defined through RMS as

$$\text{SNR}_{\text{var}} = \left[\frac{1}{p}\sum_{k=1}^{p}\left\{10\log_{10}\left(\frac{\sigma^2_{\mathcal{G}_{k,\cdot}}}{\sigma^2_{X_{k,\cdot}}}\right) - 10\log_{10}\left(\frac{\sigma^2_{\mathcal{G}_{k,\cdot}}}{\sigma^2_{\hat{X}_{k,\cdot}}}\right)\right\}^2\right]^{1/2}, \tag{3.18}$$

where $\sigma^2_{\mathcal{G}_{k,\cdot}}$ is variance of the ground truth signal, $\sigma^2_{X_{k,\cdot}}$ is the variance of the original signal and $\sigma^2_{\hat{X}_{k,\cdot}}$ the estimate of $\sigma^2_{\mathcal{G}_{k,\cdot}}$ after the model is applied. Conversely, the SNR correlation measure is defined as

$$\text{SNR}_{\text{cor}} = \frac{1}{p}\sum_{k=1}^{p}\left(1 - \frac{\rho_{\mathcal{G}_k} - \rho_{(\mathcal{G}_k, X_k)}}{\rho_{\mathcal{G}_k} - \rho_{(\mathcal{G}_k, \hat{X}_k)}}\right). \tag{3.19}$$

where $\rho_{\mathcal{G}_k}$ denotes the autocorrelation of $\mathcal{G}_k$ at lag 1, $\rho_{(\mathcal{G}_k, X_k)}$ and $\rho_{(\mathcal{G}_k, \hat{X}_k)}$ are, respectively, the cross-correlation between $\mathcal{G}_k$ with the original signal and the estimate of $\mathcal{G}_k$. We note that the mean of all correlations obtained in 3.19 is estimated via Fisher transformation. As $\mathrm{SNR}_{\mathrm{cor}} : \mathbb{R}^p \to [0, 1]$ decreases to 0, this means that the shape of $\hat{X}$ diverges from the shape of the ground truth signal. Further, one can intuitively observe that an increase in the level of $\mathrm{SNR}_{\mathrm{var}}$ suggests the presence of residual noise in $\hat{X}$ when $\mathrm{SNR}_{\mathrm{cor}}$ is also high. Finally, we used the normalized compression distance (NCD) as defined in [45].

Model configurations: deflation FastICA with *tanh* (see [306]). We used PCA whitening as it tends to enhance data compression [296], which can be advantageous for lower spatial resolutions.

- **Simulation 1**. *Objective*: We investigate the quality of extracting artifacts 1 and 2 and reconstructing the underlying process after artifact subtraction. Two scenarios are considered: one involving subspace factorization as depicted in Figure 3.1 (single iteration), and the other consisting of the direct application of wavelet thresholding to the original data. *Remarks*: the metrics for the residual process after subtraction of artifacts are calculated using the corresponding thresholding methods on the detail coefficients. Artifacts 1 and 2 where detected using the correlation coefficient between the ground truth artifact and denoised versions of S. *Results*: as reported in Table A.2, the direct application of wavelet thresholding significantly degrades all performance metrics when compared to the selective removal of artifacts through spatial factorization. The SNR levels also suggest that the residual process is spuriously modulated when direct wavelet thresholding is employed, with this issue appearing to be less pronounced when using multiplicative scaling (MS). Regarding the artifact estimation, it is evident that both MS and ENID outperform universal thresholding by a significant margin, while MS reaches an overall higher performance. We found an aliasing effect after subspace reconstruction of artifact 2 which was overcome in our analyses (metrics are derived only for the channel containing the artifact). As expected, our methods for denoising artifact 2 partially failed due to mixings with the residual process. See also Figure 3.2A.

- **Simulation 2**. *Objective*: We investigate the quality of extracting artifact 1 in the following setting: the number of bumps increase and are randomized in location, amplitude and frequency; artifact 2 is randomly changing of channel. Within this scenario, we forsee that a prior PCA reduction could help improve the the results. *Remarks*: We compare performing FastICA using in the fixed-point iteration $\Psi_{1:r,\bullet}(X - \overline{X})$ (with $r = 2$) instead of using all rows of Ψ. *Results*: As shown in Table A.3, an increase in the number of bumps leads to an exponential increase in MRMSE. However, restricting FastICA to the first components notably improves these rates, as demonstrated in Table A.4, for both MS and ENID.

- **Simulation 3**. *Objective*: In this third simulation we investigate the extraction of a new artifact that is characterized by white noise amplification during a short temporal period (Figure 3.2B) affecting the the whole spatial domain (we use the weights determined for artifact 1) in a homogeneous way (the noise is the same across all channels). We study the ability of our method to estimate artifact 3 and the residual process after artifact subtraction. *Remarks*: Again, the metrics for the residual process are calculated using the corresponding thresholding methods on the detail coefficients. $\mathrm{ISNR}_{\mathrm{cor}}$ is calculated for the region of the domain where noise amplification occurs. *Results*: See Table A.5 and Figure 3.2B. The large values for $\mathrm{ISNR}_{\mathrm{var}}$ indicate that our methods are relatively inefficient in denoising this artifact. Meanwhile, the low $\mathrm{ISNR}_{\mathrm{cor}}$ values suggest that the shape of the artifact is almost unrecoverable. This observation holds for the residual process as well, although MS appears to outperform ENID to some extent.

§ 3.7. Applications to EEG data

Arguably, the pre-processing of EEG data stands out as one of the most intricate challenges, requiring a rather crafted practices or intelligently driven approaches concerning the selection and removal of artifacts. A common practice involves the manual selection of artifacts, either through spatial representations of each vector in A or by examining S. In contrast to our simulations, the ground truth of the artifact is now unknown. Further, an artifact may overlap in multiple spatial components, and while we assume they are temporally sparse, its occurrence is often unknown.

The data used in this section was recorded at ASIL lab as part of the project with protocol no. 2022-33, which received approval from the local ethics committee at the Faculty of Arts and Philosophy (Ghent University). The participant provided written consent to take part in the study. A 64 channel EEG device recording at 10 KHz was used in the experiment. During the recording, the participant performed several

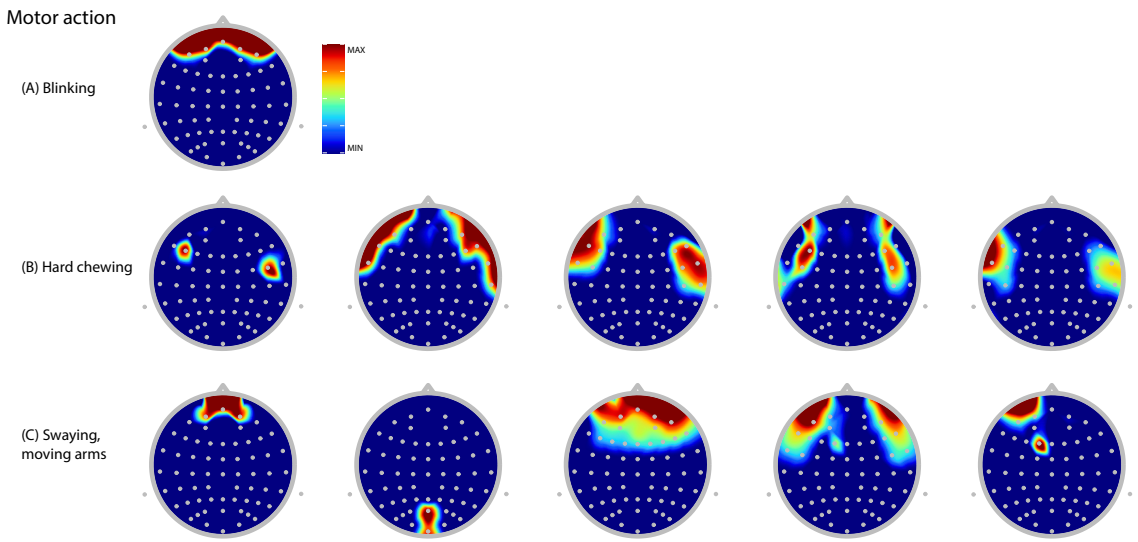

Figure 3.3: A selection of brain topographies illustrating target areas related to potential artifacts in the different recordings of motor actions. These topographies are built by interpolating the entries of $A_{\bullet,k}$ across a circular field, i.e., the channel coordinates on the hemisphere projected on to a compact circular field. Blinking and chewing artifacts (A,B) are usually characterized by strong prefrontal and localized dorsolateral prefrontal activations, respectively. More complex motor tasks, such as swaying (C) –which might involve eventual blinking as well (C-3)– also show clear spatial activations in the medial frontal/dorsolateral prefrontal and occipital cortices (C-1,2,4). Observe that the artifactual activity is mostly originated and localized at the extremals of the field.

movements and was asked to synchronize them with four auditory beats, each lasting 2 seconds. The task was divided into two separate time frames, with pauses in between. The beat appeared four times with a different sound during the pauses to give a sign to the participant of the beginning/end of the task. The recorded motor actions included blinking (Art1), intensive chewing (Art2), and swaying while moving the arms (Art3). We will be working with three multivariate discrete signals represented by matrices of size $64 \times n$. Our goal is to derive a suitable representation of EEG cortical activity by mitigating the artifactual components as estimated through the application of the proposed model. Artifacts were visually selected, as the use of semi- or automated detection methods was beyond the scope of this paper.

Data were solely demeaned and band-pass filtered between 0.5 Hz and 45 Hz to remove slow drifts and attenuate muscular activity. To assess the results, we use an SNR measure defined as the quotient between the mean of the maximum absolute voltage across channels in the original signal and the cleaned signal, expressed in decibels. Calculations were confined to the time frames with artifactual activity. We selected the best-localized wavelet family based on the artifact characteristics, utilizing wavelets with either 2 or 6 vanishing moments. For Art1 (l2), SNR values were $\text{SNR}_{\text{MS}} = 1.457681$ and $\text{SNR}_{\text{ENID}} = 1.145932$; for Art2 (l2), $\text{SNR}_{\text{MS}} = 4.742761$ and $\text{SNR}_{\text{ENID}} = 1.056919$, and (l6) $\text{SNR}_{\text{MS}} = 0.2877005$ and $\text{SNR}_{\text{ENID}} = 0.8506382$; for Art3 (l2), $\text{SNR}_{\text{MS}} = 4.327495$ and $\text{SNR}_{\text{ENID}} = 0.9719538$. Visual inspection (see Online Supplementary Material) revealed satisfactory results for Art1; however, for Art2 (l2), residual artifacts persisted with both methods, albeit reduced when using higher vanishing moments. Notably, ENID outperformed MS in reducing lower frequency artifacts in Art3. Here, SNR levels can be interpreted as indicators of artifactual activity, particularly regarding the low frequency artifacts.

§ 3.8. Discussion

In this paper, we concentrate on the mathematical and statistical principles underlying the reconstruction of artifactual activity from EEG signals. We approach this problem by considering artifacts belong to a high-dimensional space dominated by sparsity. Typical pre-processing techniques primarily involve spatial covariance estimation while overlooking the temporal dimension of the data. This has inherent drawbacks because spatial and temporal structures are related but fundamentally distinct, with critical topological information residing in the time domain. Although in [303] we aimed to address this concern, the proposed model does not take into account the spatial structure while lacks of asymptotic robustness, leaving significant gaps. Instead, here we present a model that captures linear information in the spatial

domain while also exploiting the temporal structure in a non-linear fashion using wavelet techniques on the spatial IC expansions. Our spatial approach relies on FastICA due to its simplicity and common usage in artifact detection problems [79]. Additionally, FastICA is known for its good asymptotic behavior [252], while is likely to outperform other estimators such as FOBI [218]. Other ICA methods as well as different FastICA configurations can certainly be considered in future comparative studies.

We introduced two approaches for thresholding wavelet coefficients, both of which have demonstrated their effectiveness especially in scenarios where the definition of noise (or what is meant to be denoised) lacks of a clear parametric specification. After the spatial factorization, these methods aim to enhance the irreducibility of the spatial ICs, thereby providing genuine independent spatial representations. The MS method has been shown to be notably effective in estimating stereotyped artifacts. Nevertheless, its success relies on selecting an appropriate scaling series, as well as a suitable number of vanishing moments, to ensure proper convergence. The advantage of ENID is that it always converges regardless of any parameter configuration, although in certain cases it can be less effective in recovering stereotyped artifacts. The performance of both techniques when applied to EEG data depends on their adaptability to the complex temporal topology of the artifact taxonomy and the type of embedded noise (if one assumes the presence of negligible white/coulored noise, reverting to the methods' original formulation may be convenient). Our simulations and real data analyses yield three crucial insights for practitioners: (*i*) the direct application of denoising techniques to estimate artifactual activity can lead to spurious modulation of brain activity when artifact attenuation/removal is conducted (*ii*) if artifacts are not suitably denoised, their subtraction will lead to brain activity loss (*iii*) identifying certain artifacts may rely on knowledge of the covariance structure in the temporal domain, which is not easily estimable in wide data settings.

Acknowledgements: We express our gratitude to Prof. Vittoria Bruni (Sapienza University of Rome) for her insightful suggestions during our presentation at the IMACS World Conference 2023, and to Prof. Marc Leman for allowing us to carry out tests and all simulations on a 90-core computer for parallel processing. **Data accessibility:** Data and scripts will be made available in due time on https://github.com/m-vidal. **Conflict of interest:** The authors declare no conflict of interest.

4 Study I: Modeling pupil data during musical tasks of variable emotionality

This chapter includes the paper:

- [302] Vidal, M., Onderdijk, E., K., Aguilera, A. M., Six, J., Maes, P.-J., Fritz, T. H., and Leman, M. (2023). Cholinergic-related pupil activity reflects level of emotionality during motor performance. *European Journal of Neuroscience*, **n/a**:1–14.

Pupillometry has gained increased attention in recent years, driven partly by research linking pupil activity to axonal diffusion of several neurotransmitter systems [138]. This technique offers a cost-effective and portable means of measurement, making it particularly versatile for use during various motor activities. Unlike EEG recordings, it does not suffer from typical artifact effects caused by muscular activity. However, the majority of pupil dilation changes can be attributed to autonomic regulation in response to luminance variations; for instance, a basic eye blink reflex can induce turbulent effects in the mid- and high-frequency band of the pupil signal, as shown in §4.4. Therefore, estimating cognitive/neurotransmitter-related pupil activity is not straightforward, and prior pre-processing of pupil data[1] is mandatory to derive reliable scientific results.

In this chapter, we present a paper that integrates both methodological and experimental results related to estimating pupil activity during musical tasks of variable emotionality. Emotional motor control, which involves the neural mechanisms regulating physical responses to emotions, is in its neurological organization substantially distinct from voluntary motor control [120, 295]. Here, we aim to identify its neurological signature with pupillometry, examining slow and phasic fluctuations previously related to the activity of cholinergic and noradrenergic axons. We conducted pupillometry in a population of trained singers who sequentially performed different motor tasks determined by the structure of a musical piece and designed to vary according to the degree of emotional engagement. We hypothesized that actively engaging in musical behavior by singing and moving along with the music (movement+singing) would elicit stronger emotional responses from participants, resulting in a qualitatively distinct pattern of baseline and phasic pupil activity compared to control conditions that were comparable but designed to be less emotionally engaging. The proposed paradigm is investigated within the framework of the generalized arousal (GA) hypothesis [235–237]. GA is here defined through the estimation of dominant dynamics using a multivariate functional PCA (§4.4) of pupil data.

Abbreviations: LC, locus coeruleus; NE, norepinephrine; BF, basal forebrain; Ch, cholinergic; ACh, acetylcholine; GA, generalized arousal; CWT, Continuous Wavelet Transform; KL, Karhunen–Loève; ROE, response to ocular event; NM, no movement only listening; M, body sway to music; NMS, singing but no body sway allowed; MS, body sway plus singing.

§ 4.1. Introduction

Optimal levels of arousal are critical for perceptual and cognitive functions, given that arousal modulates entire classes of responses to various events, for example making an organism more responsive to sensory stimuli, more ready to execute voluntary motor activity and more emotionally responsive [235]. Physiologically, regulation of arousal and autonomic function are related to the activity of the locus coeruleus norepinephrine (LC-NE) and basal forebrain cholinergic (BF-Ch) systems [16, 27, 28, 35, 90, 177, 264]. The LC-NE activity plays an important role in enhancing the processing of information salience [294, 308] and has been shown to have an influence on decision making [74, 84]. BF-Ch activity is a key component to promote sensory perception [239] and in emotion regulation [21, 29, 30, 193, 238]. Furthermore, its engagement is particularly characteristic as an integral aspect of motor activity, for example during locomotion [251] or other types of body movements independent of locomotion [213].

Prior studies in humans and non-human animals have demonstrated that, under isoluminance conditions, there exists a causal relationship between pupil fluctuations and the activity in the LC and the BF-Ch

[1] A comprehensive and up-to-date guide to pupillometry can be found in [95].

neurons [44, 74, 138, 139, 205, 213, 251]. However, neuromodulatory mechanisms underlying movement control and their corresponding effects on pupil behavior are not yet well understood. Neurophysiological work in rodents has shown that the pupil tends to dilate concurrently with activity of the NE and Ch axons before locomotion onset [250]. This dilation is prolonged along with a sustained Ch axonal activity until motor offset, showing a hallmark latency to reach baseline levels [194, 250, 304]. BF-Ch inputs have also been related to microdilations induced by small body movements [213]. NE phasic activity tracks transient and differential dilations during motor and passive states [139, 251]. A recent study has shown, however, that these projections are more likely to be related to infrequent and large dilation events, suggesting that inferencing on repeated measures could increase the accuracy of the NE axonal estimates [196]. Note that, while arousal and motor activity have been shown to be to some degree independent in how they modulate firing in cortical circuits [304], their interplay seems to notably contribute to functional plasticity in the cortex, for example enhancing learning [11].

The behavior of BF-Ch and LC-NE systems in humans with respect to motor functionality and how it relates to pupillary changes have been less investigated. Nevertheless, several studies on visuo-motor tasks have reported that modulation of pupil size is dominated by the motor response rather than other cognitive factors; see [190] and references therein. The intensity of physical exercise in the absence of visual cues has been previously associated with an increased baseline pupil diameter [117, 323], which when performed with moderate intensity was comparable to when participants performed mental arithmetic tasks [117]. Subsequent studies have corroborated these findings using measures of peak oxygen consumption ($\dot{V}_{O_2}$) and minute ventilation ($\dot{V}_E$), further demonstrating that exercise-intensity-dependent pupil dilation was exponentially correlated to these physiological measures [164]. Other studies have tested the effects of single bouts of exercise on cognitive inhibitory control as measured by pupillometry following the physical activity, suggesting that task enhancement was independent to some degree of LC-related pupil activity [195, 274]. Instead, findings have shown that choline supplements for boosting cholinergic activation enhanced performance accuracy over velocity during visuo-motor aiming tasks, which translated into a relative decrease in pupil size compared to when movements were faster and less accurate [208, 209].

Here, we investigate the inverse problem of inferring the activity of brainstem arousal systems from a blinded-inference paradigm perspective, using pupil recordings in humans. We formulate a frequency-specific schema of analysis based on prior investigations on motor tasks in rodents [251] that is determined by the structure of a musical piece. Our analytical methodology is motivated by the "generalized arousal" (GA) hypothesis [235–237]. Several components of the nervous system such as the medullary reticular formation, thalamus and cortex contribute to GA, which is crucial for initiation of any behavior during arousal states. The relevance of nucleus gigantocellularis neurons, whose activity is related to serotonin and ACh, together with adrenergic projections from LC have been reported to play a role in modulating GA [52, 181, 188, 287]. Here, we assume that the confluence of these neural mechanisms makes pupil activity a potential candidate whose latent dynamics might serve to model GA function. The proposed methodology extends previous efforts of GA analysis to functional data [248] through a reduction method that allows capturing variability in a population who sequentially perform different motor tasks in a musical context that are designed to vary with the degree to how emotionally engaging they are.

Participants with musical training were recruited to perform under different movement conditions, to emotionally engage the performers in various degrees during their motor tasks. Four conditions were compared: (i) no movement (NM), only music listening, (ii) body sway to music (M), (iii) singing but no body sway allowed (NMS), (iv) body sway plus singing (MS). We hypothesize that actively engaging in musical behavior while singing and moving along to the music will more strongly engage participants emotionally and lead to a qualitatively different pattern of tonic and phasic pupil activity. Given the relevance of GA for emotional motor control in the basal forebrain [188], we speculate that GA modeling of BF-ACh pupil related activity might be a way of objectively quantifying emotional response whereas transient fluctuations related to LC-NE activity occurring in parallel in higher subbands might reflect other cognitive parameters possibly related to attention.

§ 4.2. Results

GA levels of pupil related low frequency cholinergic function change with emotionality of motor performance. The MFPCA of the smoothed pupil data reported a 36.29% of explained variance for the GA component, which has previously been reported to account for less than half of the variance [52]. Levels of GA were quantified by calculating the L^2 log-norm (from the baseline) of the KL curves. We found higher dilation rates of tonic activity in all motoric conditions compared to NM (Figure 4.1E), with MS leading to a significant increase compared to the other conditions. By contrast, no differences were found between the pupil dilation rates of M and NMS, despite levels of motor activity being highly differentiated

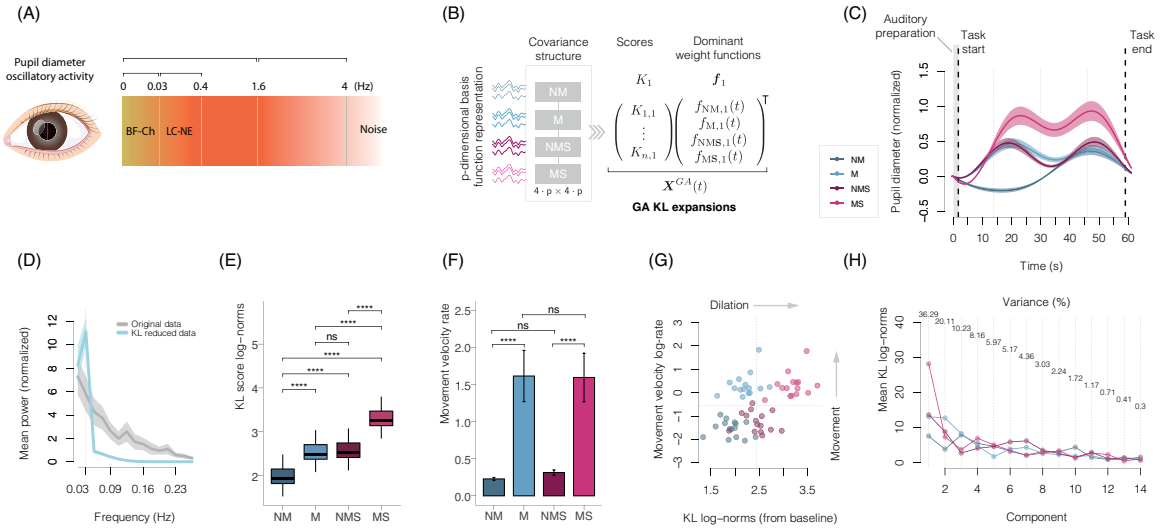

Figure 4.1: Neuromodulatory modeling of pupil data **(A)** Schematic showing the division of the pupil signal in two frequency bands (0–1.6/1.6–4 Hz) as proposed by Peysakhovich et al. [233] and levels of coherency found by Reimer for the cholinergic axons (> 0.03 Hz) and noradrenergic axons (0.03-0.4 Hz). Note that LC-NE pupil related activity (in red) is generally not restricted to values in the band 0.03-0.4 Hz. **(B)** Functional data model of Pfaff's generalized arousal applied to the data. The model is based on one factor (dominant eigenfunctions) Karhunen Loève (KL) representation, estimated from a multivariate functional principal component analysis. **(C)** KL mean curves representing BF-Ch pupil related activity during the performance task. The grey dashed lines delimit section changes (see Figure 4.2A). **(D)** Spectral analysis comparing the original dataset to the KL reduced curves. **(E)** Comparison of baseline pupil rates of the KL curves for each condition (L^2 log-norms). **(F)** Comparison of force sensor platform velocity rates derived from the computation of the instantaneous amplitude via Hilbert transform. **(G)** Baseline pupil rates are plotted as a function of the platform velocity rates; the dashed lines represent the means for each group *(H)* Means of the KL pupil rates for GA component and specific arousal forms $\{K_2 \boldsymbol{f}_2(t) \cdots K_n \boldsymbol{f}_n(t)\}$. A higher velocity decay in mean rates is observed in NMS and MS compared to the other conditions - Error bands were calculated at a 95% bootstrap confidence interval. Statistical comparisons were made using the Wilcoxon signed-rank test (with Bonferroni-Holm correction for multiple comparisons): $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$; $^{****}p < 0.0001$; n.s., not significant. The error bars indicate standard errors.

(Figure 4.1F,G). Motor activity rates (at least in terms of body sway) did not differ between M and MS.

Amplitude of sensorimotor engagement determines change in LC related pupil behavior during chorus recapitulation. It can be argued that the human fascination with music to a great degree relates to music's capacity to dynamically vary expectation [157]. A systematic variation of such aspects of stimulus property over time is an inherent quality of the stimulus material used in the present study. Accordingly, we looked at LC-related pupil behavior in relation to musical structures that are known to systematically influence expectations, verse and chorus (Figure 4.2A). This also corresponds to predictive coding theories [17, 99] that state that the brain is a solver of likelihood functions, leading to the neuronal codes that predict sensory perception, so that cognitive processing is influenced by previous exposure. We accordingly investigated if pupil indices are modulated with chorus recapitulation. We used Pfaff's GA formula for MFPCA reduction, although now applied to each part of the musical structure. Dominant fast and transient dilations can be seen as GA approximations in a lower time scale. Pupil behavior differences between chorus and chorus recapitulation were observed. As shown in Figure 4.2B, an overall decrease in pupil dilation was found during the movement conditions along the different parts of the formal structure. A higher level of dilation was observed in the singing conditions during chorus and first verse compared to NM, which was significantly higher for θ_1, θ_2 during the chorus. In the recapitulation, M, NMS and MS were more prominently lowered in θ_2, θ_3. This is apparent in the scatterplots (Figure 4.2C), where the singing conditions shifted left in the axis of the KL log-norms as a function of the velocity rate measured with the movement sensors. We also found evidence of this pupil constriction in a post-hoc analysis testing differences across subbands by pooling all scores of each respective chorus part as a single variable. Results suggest a general decrease of dilation indices in NMS and MS (Figure 4.2D). No significant differences

were reached in M and NM.

behavioral data. The discrepancy between the two experts' evaluations of the singer's performance was small ($T^2 = 1.922, p = 0.0941$), and differences between the singing conditions were not significant ($T^2 = 1.0622, p = 0.396$). The averages across conditions of each item to be evaluated were generally high (Intonation: NMV = 8.333±1.582, MV = 8.566±1.006; Rhythm: NMV = 9.466±1.008, MV = 9.6±0.723; Fluency: NMV = 9.233 ± 0.971, MV = 9.466 ± 0.571. The analysis contrasting the voice recordings between the NMV and MV conditions show that the quality of the performance in both conditions was similar (cosine similarity: 0.97 ± 0.01; Pearson's correlation: 0.87 ± 0.05 - means and SD across participants using a suitable score transformation).

Participant's subjective ratings of the level of excitement and absorption (Figure 4.2E) show that a significant percentage of participants had a higher emotional experience when singing and moving along to the music compared to when they were only listening ($p > 0.05$, using Wilcoxon signed-rank test with Bonferroni-Holm correction).

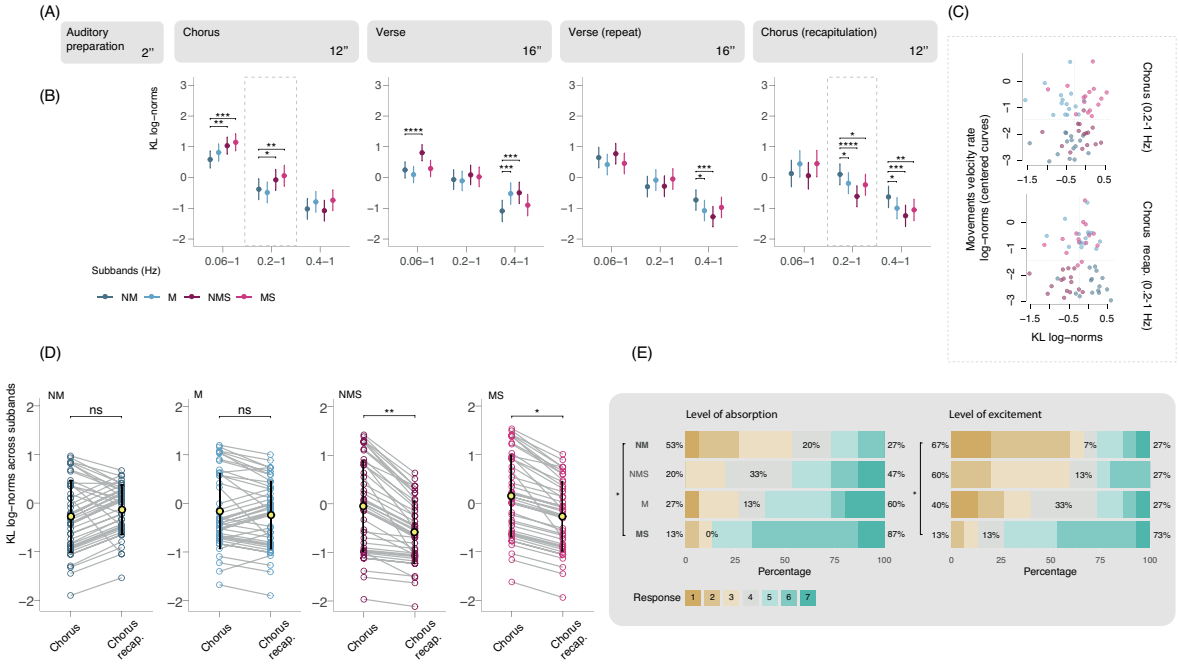

Figure 4.2: Modeling of LC-NE activity (rapid pupil fluctuations) and behavioral data **(A)** Schematic summarizing the formal structure of the musical piece the participants were asked to perform. **(B)** Comparison of GA pupil rates across different frequency subbands related to LC-NE activity. - Statistical significance is measured as detailed in Figure 4.1 with respect to the control condition NM, bars represent standard deviations. **(C)** Phasic pupil rates in the subband 0.2-1 Hz are plotted as a function of the locomotor activity (force sensor platform velocity rates); the dashed lines represent the means for each group. **(D)** Pooled dilation rates across all subbands ranging from 0.06 to 1 Hz. - Statistical tests were conducted using the two sided Wilcoxon signed-rank test for single comparisons; $P < 0.05$ effects at MS were lost for multiple testing when using Bonferroni-Holm correction. **(E)** Perceived emotional attributes. - Comparisons were made using Wilcoxon signed-rank test with Bonferroni-Holm correction.

§ 4.3. Discussion

Musical activities are known to readily evoke emotional experiences, and singing as other expressive vocalizations can be regarded to often be a rather intense emotional activity [100, 106, 122, 268, 271]. Body movement and sway has been described to happen automatically during intense musical engagement [58, 170], possibly through so-called empathic gestural attuning [169], where musical features such as beat, rhythm, melody, timbre may trigger body gesture associations. Introducing additional constraints on physical movement is often compulsory in Western classical professional musicians, and it also demonstrates to the audience an enhanced control that is perceived as an attribute of professionalism. Very relevant for the current study design, using musical engagement allowed for experimentally structuring the investigation of emotionality of task. At the same time, introducing the requirement to inhibit body sway during the task, more specifically, singing while additionally having the constraint not to move along and sway,

allowed for less emotional experience of the performer than singing while being allowed to move freely with the body.

Our analyses suggest that inhibition of body sway during singing results in an enhanced pupil baseline tone that is comparable to when participants did not sing but only swayed along to the music. It might be that synchronizing with either voice or body for a musician engages similar cognitive processes to a comparable degree, and that engaging in synchronization of both voice and body sway does combine to a different quality of immersive and emotional experience. This corresponds to the behavioral findings showing higher level of emotionality during MS. Notably participants performed singing with a similar quality with and without body sway as assessed by evaluators, supporting the notion that the observed differences were indeed due to differences in emotionality and not to other parameters such as task difficulty. Current analyses rather show that cognitive load corresponds to values of LC-NE subbands (Figure 4.2B, chorus), where singing led to an increased pupil rate compared to no singing conditions. No differences were found between pupil rates in conditions NMS and MS, which reflects participants' musical proficiency and further supports this interpretation of the results.

Following the discussion in [251], a modulation of slow frequency cholinergic activity and related pupil activity could be associated with pre-motor planning, arousal or both. While certainly controlling only voice and controlling body sway correspond to an involvement of different somatotopically organized premotor representations [175] and selective neural populations in the auditory cortex [220, 282], their effect in terms of pupil activity is comparable (Figure 4.1E). This rather seems to indicate that the cholinergic associated pupil activity reflects arousal level. As shown in Figure 4.1C, level of pupil dilation increases from no movement during music listening, to singing while not moving along to the music, to singing while moving along to the music. Given that arousal is a consistent dimension of emotion in arousal models, and musical activities readily evoke emotion, we argue that in this musical context a variation of action as part of musical tasks systematically vary level of emotionality. Supporting this hypothesis, cholinergic contributions to emotion regulation are well-known in the literature [21, 29, 30, 193] also during diverse musical and motor experience [39, 135, 154]. We are aware that performing different motor acts, for example singing or swaying, involves a different set of physiological resources that correspond to a different set of neural activation patterns that cannot only be regarded emotional. However, from a perspective of our knowledge of a dichotomy of voluntary and emotional motor control [120, 122], we know that movement and emotionality are entwined. The current findings on the variability of the slow pupil dynamics are therefore congruent with the GA hypothesis and related theories on the emotional motor system which associate mitigation of emotional cues with action [188].

Musical structure and musicians' sway during performance has been previously described to covary systematically [81]. Because body sway was one major aspect of the task, we took into account the formal structure of the musical piece when investigating the time-course of pupil dilation. Furthermore, pupil behavior in relation to musical structure is applicable to understand rapid pupil fluctuations (0.06-1 Hz) that have been observed to relate to LC activity, which is often modulated during changes in stimulation, especially if such changes relate to the task. We observed that the pupil first dilated and then attenuated over time and repetition of the music. Larger and rapid pupil dilations have been associated to cortical desynchronization (suppression of low-frequency fluctuations), a form of attentional state required to accurately process sensory information [115, 194, 304], also in the auditory domain [230, 262]. Although this pattern of dilation and posterior attenuation when musical sections repeated was apparent in all motor conditions (Figure 4.2B), it was particularly robust during the singing conditions as shown in our post-hoc analyses (Figure 4.2D).

Such attenuation might correspond to the pupil's ability to respond to information from dynamic environments, possibly reflecting a process of active inference [157] in musical performance, even when visual information is controlled for. Pupil activity in correspondence to multimodal cognitive rather than visual processing has previously been shown. For example, higher learning rates in a prediction task were associated with a smaller baseline pupil diameter [211], or biases on subject's internal beliefs related to higher pupil dilation in dynamic environments [94, 162]. In addition, motor engagement during performance seems to enhance this aspect of pupil activity given that attenuation of LC related pupil indices occur mainly during motor conditions as shown in our analyses (Figure 4.2B,D). Similar results have also been reported during motor tasks, reflecting motor control and learning [209, 314].

Furthermore, we have introduced an unsupervised algorithm that allows to detect ROE of different duration and amplitude regardless of the artefact benchmark. The method works in a nonlinear fashion, which makes it suitable for non-stationary environments and only recasts on a dispersion parameter that can be relaxed according to the levels of contamination. Turbulence due to ROE is then minimised, providing a more precise estimate of the neurological processes under investigation. With this, dissociable traits in the pupil behavior related to the LC and BF-ACh activity found in previous research can be

discovered. However, simultaneous measurements of neuroimaging and pupillometry techniques are necessary to corroborate how pupil size and the activity of these neuromodulators, and possibly others, are related and interact during motor performance of variable emotionality.

§ 4.4. Materials and methods

Participants

Fifteen participants (all female, aged 21-50), with formal and/or informal musical education (mean = 11.46 years; SD = 6.51), gave written consent for taking part in the study, approved by the local ethics committee (Faculty of Arts and Philosophy, Ghent University). They were right-handed and had normal vision, or vision correction, normal hearing and were neurologically healthy. All participants were screened on their experience with singing in a choir, which ensured they had experience with singing a steady part in the presence of other voices. They received voucher credit for their participation.

Acquisition of pupil data and other modalities

Pupil diameter was recorded at 30 Hz frequency rate using Pupil Lab's wearable headset with integrated cameras (Logitech C930e) directed towards the right eye. A force sensor platform with 59 cm radius was used to control for body movement. This platform consists of a plate with four sensitive weight sensors underneath (one at each corner) to register variations in pressure related to body movement. Each sensor is captured with the 13bit Analog Digital Converter of a micro-controller (Teensy 3.1, PJRC) at 120Hz. The sensor data is wrapped into MIDI packets. This allows recording audio and sensor values in sync using standard digital audio workstation software: in this case, Ableton Live 9 was used. The participant's singing voice was recorded with a Shure Beta 87A microphone. Next to this, a decibel meter (UNI-T UT352) was used to monitor and review the volume level before the start of the experiment (to limit the effect of loudness).

Procedure and rating of perceived level of emotionality

The experiment was conducted in a dark room with steady LED luminance of less than about 30 cd/m^2 (Uni-T Luminometers UT381). The participants stood on the platform, facing three speakers at a distance of 2 meters. During conditions NM and NMS they had to stand still looking at a white cross placed on the middle speaker whereas in conditions M and MS, gaze movement was allowed facing forward horizontally where the speakers were located. Gazing upwards or downwards was advised against. Five-point calibration and validation were performed before the start of the experimental session. Hereafter, we asked participants to stand on their assigned spot in the room and to sing the rehearsed melody by heart. This allowed us to check whether participants learned the melody to complete the various tasks with fluency. Participants performed with an instrumental music piece including a three-part vocal harmony. In conditions NMS, MS, they were singing the middle voice from the three part harmony, which was in the singing conditions not audible in the music presentation. The order in which every condition was presented was randomized for all participants. For further details concerning the experimental setting and stimuli, see Supplementary Material.

After performing each task, participants were asked to rate their perceived level of excitement and absorption (degree of correspondence with the activity) on a Likert scale from 1 "low" to 7 "high level". We used the terms absorption and excitement as attributes of positive emotionality.

Analysis of the singing performance

Recordings of the singers were evaluated by two experts with more than 20 years of musical training and teaching experience in musical institutions. They were asked to rate on a scale from 1 (very inaccurate) to 10 (very accurate) the following items: intonation, rhythm and fluency (smoothness of the performance, as in [221]). The recording files were blinded and randomized before their presentation to the evaluators. Furthermore, acoustic similarity between NMS and MS conditions was measured on the recordings with the singing part isolated from the instrumental by calculating cosine-similarity and Pearson's correlation on spectrogram-like representations (frequencies were in Mel).

General data preprocessing and statistics

Outlying data produced by blinks taking values of zero (or close to it) were removed from 100 ms before until 200 ms after the observation (6.962% ±6.158 of missing data). As an alternative to interpolate the removed values, we imputed them using an algorithm based on a vector autoregressive model with heavy-tailed Student's t distributed innovations that is robust against outliers [320]. This procedure reconstructs the missing data using stochastic parameters amenable for heavy-tailed and sparse high-dimensional data.

Subsequently, the pupil time series were mean-centered and normalized to unit variance. Standardization across subjects allowed to control for differential sensitivity to the ambient luminance, as to compute higher-order moments. Unless otherwise stated, the median pupil size of the interval from 200 ms to the start of the auditory stimuli was used as baseline. Similarly, the data recorded with the force platform was downsampled at 30 Hz, low-pass filtered (1.5 Hz) and normalized. To express this data as dynamic firing rates, we calculated the first derivative and its Hilbert envelope or instantaneous amplitude.

All data and statistical analyses were performed using custom-made R scripts [244] (ver. 4.2.0). Statistical significance was measured against an alpha level of 0.05.

Estimation of cognitive-related pupil activity

Turbulent-like dynamics can be encountered in many physiological phenomena, such as in cardiac [269] or brain imaging signals [76, 88]. In the current research scenario it can be understood as chaotic dynamics that appear as a distortion of a more parsimonious state by various physiological parameters. More precisely, in pupillometry turbulence is observed after changes in retinal illuminance (for example, when blinking), as the pupil tends to rapidly constrict and re-dilate transiently to tonic levels, drawing the shape of a negative gamma function in the signal (Figure 4.3A,E). Luminance-related changes directly modulate the feedforward response of V1 (striate cortex) that is independent of psychological factors like attentional load [41], suggesting that changes in pupil size due to autonomic regulation also affect visual perception to some degree. Note that while blinks are often considered informative of attentional or other arousal states [80, 146, 270], their effect on pupil size is probably rather related to basic visual function.

Pupil responses produced by subtle changes in luminosity can act as confounders, notoriously biasing the results of a subsequent data reduction [152, 317]. The intensity, duration and frequency of pupil occlusion together with abrupt visual changes resulting from movement, can generate turbulent flows in the signal hampering estimating pupil behavior related to cognitive processes. We observed that blinks are more likely to occur during movement and vocalization tasks compared to the control task NM ($p < 0.05$ using permutation test for location based on means applying the Box–Cox transformation; see [61]), which is consistent with previous findings reporting an increasing blink rate due to complex motor activity during speaking [46]. Furthermore, responses to ocular events (ROE) might overlap with other ROE in such conditions, which can lead to a non-linear distortion of the signal. Taking into account all these facts, a systematic correction of ROE is necessary to provide reliable results about levels of pupil-linked arousal.

To detect these turbulent flows, we applied a novel unsupervised algorithm that accounts for slow and high frequency ocular responses. The model uses nonlinear internal vector spaces of the pupil signal reconstructed by a third-order Butterworth band-pass filter with a fixed higher cutoff frequency and variable low pass frequency cutoff to gradually remove localized abnormal oscillations.

ROE reduction algorithm

Let $x_1, x_2, \ldots, x_t, \ldots, x_n$ be an univariate time series of n pupil diameter measurements and denote by $\mathrm{BP}_{\underline{\omega},\overline{\omega}_j}(\cdot)$ a battery of third-order band-pass Butterworth filters with fixed cutoff frequency $\underline{\omega}$ and variable low pass frequency cutoff $\overline{\omega}_j$ ($j = 1, \ldots, m$). Then, the ROE reduction consists in the following iterative process.

For each filtering step $\overline{\omega}_j$, we repeat:

1. Calculate $\hat{x}_t = \mathrm{BP}_{\underline{\omega},\overline{\omega}_j}(x_t)$ and determine the set of turbulence onsets t_{τ_l} ($l = 1, \ldots, L_j$) such that $\Delta\hat{x}_{t_{\tau_l}} < \mathrm{Mdn}_{\Delta\hat{x}} - k \cdot \sigma_{\Delta\hat{x}}$, where Δ denotes the differencing operator and k is a dispersion hyperparameter.

2. For each t_{τ_l} at the jth filtering step, let $\{t_{a_l}, \ldots, t_{\tau_l}, \ldots, t_{p_l}\}$ be the set of observed time points where t_{a_l}, t_{p_l} represent the index positions related to the smallest nearest neighbouring peak of $\hat{x}_{\tau_l}$ and the relative position of this peak before or after t_{τ_l}. Then, the subtraction can be performed using baseline correction as

$$(x_{t_{a_l}:t_{p_l}} - x_{t_{a_l}}) - (\hat{x}_{t_{a_l}:t_{p_l}} - \hat{x}_{t_{a_l}}) + x_{t_{a_l}}, \tag{4.1}$$

   where $x_{t_{a_l}:t_{p_l}}$ is the snippet containing the ROE.

3. Update x_t for each τ_l and then go to the next filtering step.

Parameter selection and validation

To determine the parameters $\underline{\omega}, \overline{\omega}_j, k$, we recorded a participant who was asked to blink four times synchronized with an auditory beat (of 2 s duration) in two time frames separated by pauses. The beat appeared

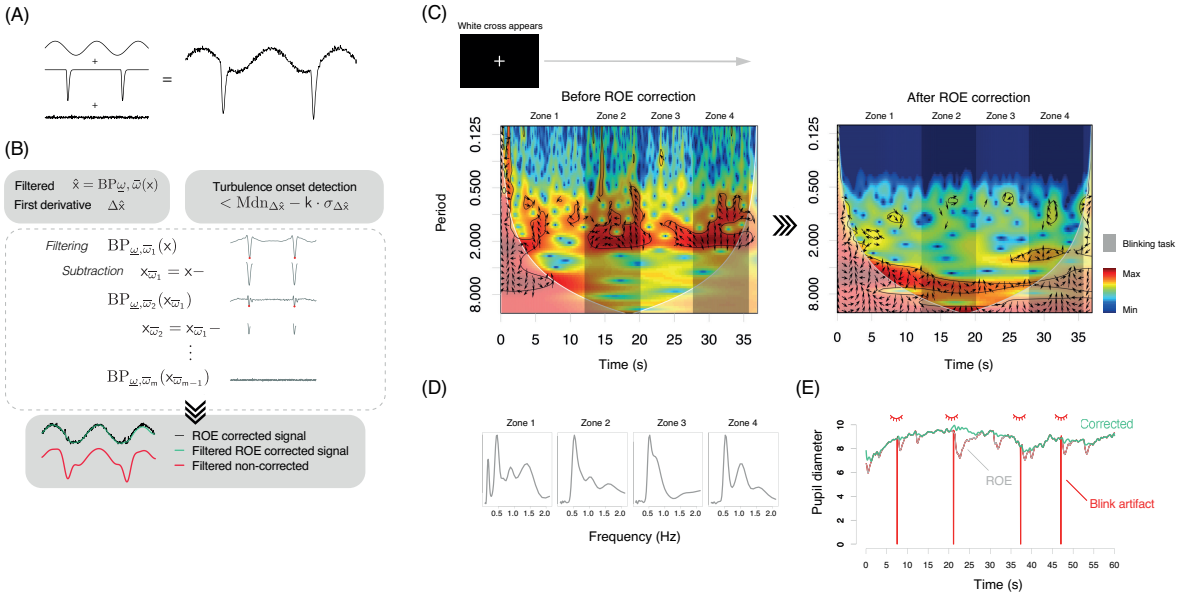

Figure 4.3: Removal of ROE: toy simulation and real data examples showing how the ROE algorithm works. (A) Simulated contaminated pupil data. (B) Downwards oscillations corresponding to ROE are detected using a battery of band pass filters and progressively removed at each smoothing step subtracting the area of the curve generated below the nearest low-peak of the oscillation period to the turbulence onset.(C) CWT before and after the application of the ROE correction (with final low-pass filter at 1.6 Hz). Zones shaded in grey depict the ROE phase changes during the blinking task (4 blinks) with contours enclosing turbulent flow fields (regions with a significance level > 0.95% according to a χ^2 test). *(D)* Series showing the row sums of the matrix containing significance levels of the CWT for the different zones marked in the CWT before ROE correction. (E) Application of the algorithm (with $k = 3$) to real data retrieved from the Python package FIRDeconvolution [153].

4 times with a different sound during the pauses to alert the participant of the beginning/end of the blinking task. Blinks were intentionally performed longer to see their effect on the signal. During the pauses, eventual and faster blinks also occurred. We recorded pupil activity in a dark environment where a white fixation cross was projected during the experiment. This produced a slow ROE related to changes in the luminance level (Figure 4.3C). For further details, see Supplementary Material Online.

The continuous wavelet transform (Morlet wavelet) was performed on the data to examine the pseudo-frequency scales characterizing the dynamics of turbulence produced by ROE. Time-power spectra bias-corrected [182] and levels of significance (χ^2 test, see [290]) are shown in Figure 4.3. The ROE algorithm was applied in two steps from the baseline frequencies $\underline{\omega}_1 = 0, \underline{\omega}_2 = 0.25$ and $\overline{\omega}_{1,j} = (0.03, 0.045, \dots)$, $\overline{\omega}_{2,j} = (0.5, 0.515, \dots)$ that allowed to detect slow and rapid ROE, respectively. We chose $\underline{\omega}_2$ guided by the CWT (continuous wavelet transform) results analyzing the series obtained by summing the values of each row of the matrix containing significance levels (zone 1, Figure 4.3D). The minima found (0.278 Hz) after the first local maximum (0.175 Hz), which roughly delimits the upper frequency range for the low-frequency ROE observed in zone 1, served as a point of reference. The value of $\underline{\omega}_{1,1}$ was determined by Reimer's top frequency threshold for the slow ACh activity, and $\underline{\omega}_{2,1}$ again summing the values of each row of the matrix containing significance levels, but now selecting the second peak of frequency (~ 0.5 Hz) which was noticeable across all zones analyzed with blinking activity, whether voluntary or not. The mth values of the $\overline{\omega}_j$'s vectors can be chosen based on a frequency limit above which pupil activity is rarely attributable to physiological sources (e.g. 4 Hz as suggested in [152, 233]). Lastly, the corrected pupil data can be low-pass filtered at 4 Hz or at lower frequencies for subsequent analyses (we used 1.6 Hz, see Figure 4.3C).

The dispersion hyperparameter (k) from the filtered signal's first derivative median value is fixed to identify the abnormal oscillatory changes. Because blink responses contribute more variance to the signal in movement-controlled conditions [152], a lower value than $k = 3$ (which is often conventionally used) provided more realistic estimations. Thus, for $\underline{\omega}_2$, we modeled the dispersion hyperparameter as the decreasing exponential function $k = 3 \cdot e^{-b}$, where b denotes the participants' blink rate. The turbulence onset is then defined as a low peak of velocity that surpasses the established dispersion threshold. Application of the ROE algorithm to the data and other external data (Figure 4.3C) shows that the algorithm is capable to identify all ROE turbulence right after the blink artifact but also others that occur directly after ROE

indistinctly of the artifact benchmark, possibly due to autonomic regulation [192]. More research is needed to refine parameter selection, for example, in different or even variable luminosity conditions assessed by multiple recording devices.

Hilbert space modeling of generalized arousal function

Fourier basis approximations of slow BF-Ch pupil related activity

Axonal projections from the cholinergic neuromodulatory neurons during locomotion have been found to be coherent with pupil oscillations in low frequencies (< 0.03 Hz) [251]. These frequencies operate on a timescale beyond what is often described as moment-to-moment changes/fluctuations, therefore we examined pupil data across the duration of the entire musical excerpt. To characterize cholinergic activity through pupil measurements, in a first step pupil curves are regressed out as smooth functions using a Fourier basis. The choice of a Fourier basis instead of other systems (B-splines, Wavelets) is supported by the assumption that the shape of the pupil is a perfect circle as well as the apparent periodic behavior of the data. The dimension of the basis was selected in order to minimize the root mean square error (RMSE) between the observed data and the evaluated Fourier basis approximation (discrete predicted values). To shape the baseline modulations of interest, we selected the basis dimension on a low range ($p < 11$). This allowed to find a RMSE trade off between middle and low-frequency rhythmicity, notably reducing oscillatory activity (> 0.05 Hz, see Figure 4.1D) to levels of coherency previously found [251].

Fourier basis approximations of phasic LC-NE pupil related activity

To quantify variations of noradrenergic activity through pupil diameter, we bandpass-filtered the data from 0.06 to 1 Hz (θ_1). This frequency band includes part of Reimer's subband for NE axonal activity (0.03-0.4 Hz) and ranges upwards with the high-frequency threshold determined in a recent study in humans [200] that is based on previous research in rodents [194, 250]. Above 1 Hz, results might be more accurately quantifiable in a more fine grained time scale. Note, however, that human pupillary oscillations in higher subbands have been linked to luminance effects rather than other cognitive factors [207, 233].

Crucially, we are interested in representations of the pupil data less susceptible to baseline effects, as activity in noradrenergic projections in cortex, which is characterized by localized bursts, tracks phasic changes in pupil diameter with an observable causal effect on the pupil size gain [139, 196, 251, 294]. Therefore we selected a high-pass cutoff of 0.06 Hz to ensure certain degree of stationarity, while also preventing overlap with slow BF-Ch pupil related activity. To this end, here we further propose a form of deconvolution by narrowbanding the signal into two additional subbands, from 0.2 to 1 Hz (θ_2) and from 0.4 to 1 Hz (θ_3). This technique is used as an alternative to taking first differences, which although it has previously allowed to establish a number of correlations with NE axonal activity [139, 251], after a first normalisation of the pupil data, statistical effects are more unlikely to survive when differentiating (which is also a kind of normalization); see Discussion in [296].

To enhance the functional representations of the data filtered above 0.06 Hz given their high density rhythmicity, we cut the data into four parts according to the formal structure described in Figure 4.2A. In order to accurately represent the shape of the rapid oscillations that were observed, we fitted a standard Fourier basis with a larger dimension ($p = 19$). Since the filtered curves tend to be rather stationary when slow oscillations are reduced ($p > 0.01$ on 83.88% of the data, augmented Dickey-Fuller test) the pupil series were mean-corrected instead of baseline aligned.

Multivariate functional GA model of pupil data

Within this setting, we resort to Pfaff's GA form, an elementary form of arousal we denote by A_1 expressed in combination to other specific forms $A_2, \ldots, A_n$ as

$$\text{Arousal} = F(K_1A_1 + K_2A_2 + \cdots + K_nA_n), \tag{4.2}$$

where F is a mapping (non necessarily linear) and $K_1, K_2, \ldots, K_n$ are scores reflecting traits of the individual [235]. Eq. 4.2. can also be expressed dynamically as a differential equation [52]. Statistically, we interpret Eq. 4.2 by means of a multivariate functional principal component analysis (MFPCA), which takes into account response dependencies through the pairwise cross-covariance functions [2, 134]. This modeling strategy naturally extends previous linear formulations of the problem [104, 243] to the functional case. In our approach, however, K_1 is a vector of scores obtained from the projection of the original functional data on to a vector of weight functions (one for each condition) with a major variance contribution. The GA we define in this context is then the set of random functions $\boldsymbol{X}^{GA}(t) = K_1\boldsymbol{f}_1(t)$, where $\boldsymbol{f}_1$ is a vector of weight functions and $\boldsymbol{X}^{GA}$ is a one factor representation of the original data (see Figure 4.1B), also known as the truncated Karhunen-Loève (KL) expansion.

Next we formalize the above interpretation of Pfaff's GA model. We assume that the pupil data have been approximated by a finite linear combination of non-linear Fourier functions (see, for example, [248]). Consider a sample of multivariate functional observations over a closed interval,

$$\boldsymbol{X}_i(t) = (X_{i1}(t), \ldots, X_{iH}(t))^\top; \quad (i = 1, \ldots, N; t \in \mathcal{J}), \tag{4.3}$$

containing the pupillometric curves of N individuals in H experimental conditions. These functional approximations have the basis expansion $\boldsymbol{X}_i(t) = \boldsymbol{\Phi}(t)\mathrm{c}_i^\top$, where $\boldsymbol{\Phi}(t)$ is a diagonal matrix of p dimensional vector valued Fourier functions with dimension $H \times (p_1 + p_2 + \cdots + p_H)$, that is,

$$\boldsymbol{\Phi}(t) = \begin{pmatrix} \phi_{11}(t) & \cdots & \phi_{1p_1}(t) & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \phi_{21}(t) & \cdots & \phi_{2p_2}(t) & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \phi_{H1}(t) & \cdots & \phi_{Hp_H}(t) \end{pmatrix}, \tag{4.4}$$

and

$$\mathrm{c}_i = (c_{i11}, \ldots, c_{i1p_1}, c_{i21}, \ldots, c_{i2p_2}, \ldots, c_{iH1}, \ldots, c_{iHp_H}), \tag{4.5}$$

its vector of coefficients. If all coefficients c_i are pooled together by rows, we can rewrite the above expression as $\boldsymbol{X}(t) = \boldsymbol{C}\boldsymbol{\Phi}^\top(t)$. The MFPCA is then obtained by solving the eigenequation

$$\boldsymbol{\Phi}(s)\Sigma_C \boldsymbol{G}\boldsymbol{b}_m^\top = \lambda_m \boldsymbol{\Phi}(s)\boldsymbol{b}_m^\top, \tag{4.6}$$

where $\Sigma_C = \frac{1}{n}\boldsymbol{C}^\top\boldsymbol{C}$, $\boldsymbol{G} = \int_{\mathcal{J}} \boldsymbol{\Phi}(t)^\top\boldsymbol{\Phi}(t)\mathrm{d}t$ and $\boldsymbol{b}_m^\top$ is the vector of coefficients of $\boldsymbol{f}_m(t) = \boldsymbol{\Phi}(t)\boldsymbol{b}_m^\top$, a vector of weight functions obtained from the spectral decomposition of the variance-covariance function $\Gamma(s,t) = \boldsymbol{\Phi}(s)\Sigma_C\boldsymbol{\Phi}^\top(t)$, and $\lambda_m \in \mathbb{R}$ its associated eigenvalues. This problem has the algebraic solution in the matrix eigenproblem $\Sigma_C\boldsymbol{G}\boldsymbol{b}_m^\top = \lambda_m\boldsymbol{b}_m^\top$, where $\boldsymbol{b}_m$ are in turn the eigenvectors of the matrix Σ_C provided that $\boldsymbol{G}$ is the identity matrix. This way, the principal component scores are $K_{im} = \mathrm{c}_{im}^\top\boldsymbol{b}_m$ and the corresponding KL expansion for the GA component is

$$\boldsymbol{X}_i^{GA}(t) = \boldsymbol{\mu}(t) + K_{i1}\boldsymbol{f}_1(t), \tag{4.7}$$

where $\boldsymbol{\mu}(t)$ is a vector of mean functions of $\boldsymbol{X}_i(t)$ and $\boldsymbol{f}_1(t)$ is a vector of eigenfunctions (one per condition). Specific forms of arousal are then similarly obtained as

$$\{\boldsymbol{X}_i^{(2)}(t) = \boldsymbol{\mu}(t) + K_{i2}\boldsymbol{f}_2(t), \ldots\}. \tag{4.8}$$

Note that the MFPCA is reduced to the multivariate PCA of the Fourier coefficients concatenated across trial conditions within each subject.

As the MFPCA (or second order models) are based on a Gaussian assumption, some unusual observations might significantly influence the estimation of the covariance and the quality of the estimators of its spectrum. Therefore, prior application of the method we checked whether this assumption was satisfied. After ROE removal, no multivariate functional outliers were detected [15], meaning that these dominant dynamics we retrieved from our data are robust and consistent for inferencing. Computational implementations of the MFPCA can be found in the R packages fda, funHDDC and MFPCA.

Acknowledgements: We wish to thank Prof. Mark Cohen (UCLA Brain Research Institute) for providing valuable comments and suggestions on our study. We also would like to thank Prof. Gert Holstege, Prof. Chris Lowry and Prof. Arno Villringer for their contributions in discussions about the emotional motor system. We thank Kevin Smink for his help in making the accompanying/explanatory animation video. **Author contributions:** Conceptualization, M.V., K.E.O. and M.L.; data curation, M.V. and J.S.; formal analysis, M.V. and A.M.A.; funding acquisition, A.M.A. and M.L.; investigation, K.E.O. and J.S.; methodology, M.V. and K.E.O.; project administration, M.V., K.E.O.; resources, K.E.O. and J.S.; software M.V.; supervision, A.M.A, P-J.M., T.H.F. and M.L.; validation, M.V. and A.M.A.; visualization, M.V.; writing – original draft, M.V.; writing – review & editing, M.V., T.H.F. and M.L. All authors read and approved the final manuscript. **Data accessibility:** All data underlying the present findings are fully available and can be shared from upon request. Routines to compute the ROE algorithm and further examples are available at https://github.com/m-vidal/pupil-turbulence-removal. Supplementary material online can be found in https://onlinelibrary.wiley.com/doi/full/10.1111/ejn.15998. **Conflict of interest:** The authors declare no conflict of interest.

5 Study II: Turbulence modeling of EEG signal during embodied virtual interactions

The current chapter includes the manuscript:

- [301] Vidal, M., Moura, N., Aguilera, A. M., Fritz, T. H., and Leman, M. (2024b). Geometric-based turbulence analysis of EEG signals for modeling emotional arousal during active immersive virtual interactions. *Under review*.

The current investigation's contribution is critical for understanding the neurology of emotion and emotional motor control, and can be regarded as an extension of the research conducted in [302] (Chapter 4). The study emphasizes the relevance of embodied musical interactions in comprehending the brain's operations under emotional states and reveals that brain function, operating at a less turbulent level, can be facilitated through virtual external agents, thereby engaging in a more parsimonious/homeostatic state that characterizes the sense of emotionality. We assessed turbulence of dominant brain dynamics in EEG recordings of 30 healthy participants during musical tasks in a virtual reality setup, allowing for variation in both task demand and degree of emotionality. Conditions included singing, swaying, responding to a virtual conductor of variable expressivity, having your own body movements mirrored by a virtual agent, and combinations thereof. In §5.4, we introduce multivariate functional ICA based on [300, 303] that is sequentially performed throughout the signal, allowing to quantify the degree of turbulence intensity in large time courses. This model aligns with the one outlined in [302], yet it emphasizes the analysis of generalized arousal function from a spatio-temporal perspective, while benefits from the finer temporal resolution EEG recordings offer.

Abbreviations: ENID, entropic normalized information distance; EE, emotional engagement; GA, generalized arousal; GOF, goodness of fit; LL, log-likelihood; MEM, mixed-effects model; MVR, movement velocity rate; TD, task demand; NM.S.NA, no movement singing, no avatar; M.S.NA, movement singing no avatar; M.NS.Mir, movement no singing avatar mirroring; NM.S.Rob, no movement singing robotic avatar; NM.S.Exp, no movement singing expressive avatar; M.S.Mir, movement singing avatar mirroring; VA, virtual agent.

§ 5.1. Introduction

Singing is one of the most archaic and refined forms of human emotional expression. From an evolutionary perspective, the intimate relationship between singing and emotion can be explained by the adaptive functions of music [96, 109], ranging from mating selection [118, 133] to social bonding [43, 265, 313] or care giving [62, 140, 173]. If, on the one hand, the act of singing embodies intense affective responses which facilitate performers experiencing emotions themselves, it also induces an emotionally contagious effect that amplifies interpersonal interactions.

Intricately connected to the emotional motor system [120], vocal production crucially engages premotor interneurons in the nucleus retroambiguus, a region of the brainstem connected to nuclei and higher brain regions controlling motor functions [121]. This process is influenced by two neural pathways: one through the anterior cingulate cortex and the midbrain periaqueductal gray for voluntary initiation and emotional control, and the other involving the primary motor cortex and subcortical loops that modulate vocal motor commands [142, 229]. Singing (or listening to human singing) elicits unique neurological patterns, involving specific brain areas such as the insula and parietal regions, anterior superior temporal gyrus, among others potentially influencing interactions with networks of selective neural populations in sensory-motor areas beyond the auditory cortex [150, 151, 175, 220, 282, 318]. Research in humans and non-human animals suggest that brainstem neurotransmitters play a significant role in shaping singing behavior [135, 265, 302].

Given that vocal production engages both voluntary and emotional motor systems, with singing being perceived as a rather emotional experience, we believe that the current paradigm is adequate to investigating neurological signatures of emotional arousal states. We hypothesize that active participation in musical tasks, including singing and moving along to the music, will more strongly involve participants

emotionally, hence constituting an optimal scenario to study the neurology of emotion and emotional motor control. In this study, we use the context of an immersive virtual reality experience including several singing and expressive movement task combinations. We experimentally introduce three different task-specific conditions: (1) singing (singing vs. non-singing), (2) moving (moving vs. non-moving), and (3) virtual agent (VA) interaction (expressive vs. robotic vs. mirroring vs. no VA). Combined, the previous conditions produce six different tasks: without VA, non-movement singing (NM.S.NA) and movement singing (M.S.NA), and with VA, movement non-singing mirroring VA (M.NS.Mir), movement singing mirroring VA (M.S.Mir), non-movement singing robotic VA (NM.S.Rob), and non-movement singing expressive VA (NM.S.Exp). The dual movement paradigm has been widely adopted in music research [23, 155, 257, 302], to cite a few. We further incorporate virtual reality as novel means to explore the social dimension underlying music experience, ensuring an environment that closely mimics real-world conditions while allowing for the use of complex physiological apparatus.

Modeling brain arousal through a geometric-based turbulence approach

The relevance of brainstem function to singing brings forth the notion of generalized CNS arousal (GA) [52, 148, 189, 235, 236]. GA is considered the composite result of multiple neuromodulatory systems operating at a high hierarchical level in the brain, commonly manifested through behavioral activation. This phenomenon has been widely investigated using electrophysiological measures [52, 103, 126, 253]. While existing studies have proposed several quantitative descriptors for GA [52, 236, 241, 243, 302], our work introduces a Hilbertian framework for analyzing turbulent EEG dynamics, capturing dominant spatio-temporal brain activity patterns aimed at reflecting GA function.

In recent years, differential equation modeling using ensembles of Stuart-Landau (Hopf) oscillators have been applied to characterize turbulence in fMRI and MEG datasets [75, 76, 88], with limited exploration in the context of EEG studies. These approaches, however, provide a compact interpretable model in terms of the relationship between derivatives, while other reduction techniques achieve more effective dimension reduction from a geometric point of view [9, 31, 255]. Nuanced variability can be further enhanced through multivariate considerations [256], encompassing models that factor in multiple dimensions, including time, space, and experimental conditions altogether, facilitating precise characterization of dynamic interdependencies inherent in neuroscientific data.

Here, we work under the paradigm of second-generation functional data [159]. In the analysis of these data, typically assumed to belong to an infinite-dimensional Hilbert space, complex dependencies between functional observations are considered. Given the non-Gaussian nature of turbulence, we propose a functional independent component analysis (ICA) [87, 303] extended to high-dimensional multivariate functional data that vary in spatial and temporal location to analyze turbulent flows on the cortical field. The model bears resemblance to methods that use the Karhunen-Loève decomposition for analyzing turbulent dynamics [31], although here adapted to the functional ICA paradigm and the notion of GA. We aim to uncover the presence of turbulent-like brain dynamics through the spectral analysis of a kurtosis kernel function, examining its temporal effects on dominant spatial representations of EEG brain activity. We also anticipate that the latent temporal dynamics under study, represented by the kernel eigenfunctions, operate within a regime of near-criticality. The model is depicted in Figure 5.1D.

§ 5.2. Results

This experimental study adopted a naturalistic neuroscience approach to preserve participants' organic singing and motor behavior. Six conditions (see Figure 5.1A) were compared varying in the degree of emotional engagement during embodied virtual interactions. We selected the alpha (8-13 Hz) and high gamma (50-80 Hz) frequency bands due to their significance in emotional processing (see Discussion).

Self-reports. Graphical representation of the self-report results is shown in Figure 5.1B,C. Emotional engagement was rated higher in the mirroring (M.S.Mir, $p = 0.01$, and M.NS.Mir, $p = 0.0005$) and expressive (NM.S.Exp, $p = 0.01$) conditions, when compared to the robotic condition (NM.S.Rob). Similar results were found for mirroring (M.S.Mir, $p = 0.005$, and M.NS.Mir, $p = 0.018$) and expressive conditions (NM.S.Exp, $p = 0.027$) compared to the non-movement singing no avatar condition (NM.S.NA). Absorption was rated higher in the singing mirror (M.S.Mir, $p = 0.006$) and expressive conditions (NM.S.Exp, $p = 0.012$) compared to the robotic condition (NM.S.Rob). Absorption was also rated higher in the mirroring (M.S.Mir, $p = 0.002$, and M.NS.Mir, $p = 0.022$) and expressive conditions (NM.S.Exp, $p = 0.003$) compared to the non-movement singing no avatar condition (NM.S.NA). In both variables, no effects were found with the movement singing no agent condition (M.S.NA). For the variable control, no significant effects were found. Interaction levels with the virtual avatar were rated higher for the mirroring

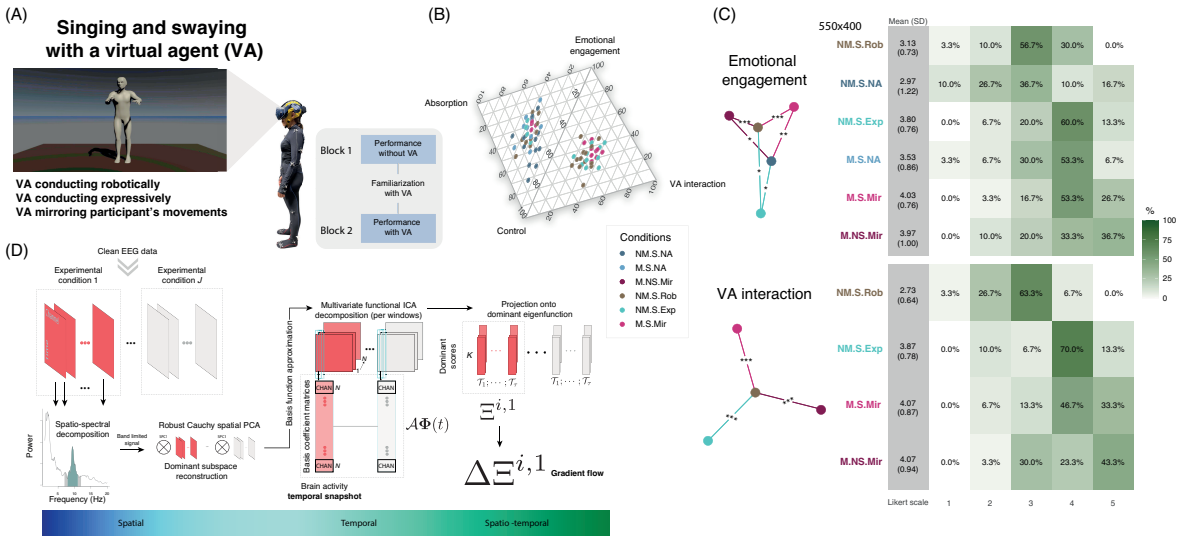

Figure 5.1: Experimental paradigm, behavioral data results and schematic of the brain activity modeling **(A)** Participants engaged in a set of conditions combining singing, swaying and interacting with three virtual agents conducting music (robotically, expressivelly and mirroring participants' movements), while wearing a mocap suit, VR headset and EEG cap. The experiment was organized in two blocks of randomized conditions, without and with VA. A short familiarization task with the VA took place before the VA block. Abbreviations for conditions: NM (no movement), M (movement), S (singing), NS (non-singing), NA (no VA), Mir (mirroring VA), Rob (robotic VA), Exp (expressive VA). **(B)** Behavioral data: pyramid relating the different items rated by the subjects (on the vertices); higher deviations from the center represent heterogeneity in behavior patterns. **(C)** On the left, graphs of the rated items directed with colors towards the condition with higher rates with the indication of the test significance. On the right, percentage distributions of the participants' emotional engagement and VA interaction ratings (1 to 5) across task conditions. **(D)** Mathematical model of Pfaff's generalized arousal to study turbulence behavior in the EEG signal. - Statistical comparisons were made using the two sided Wilcoxon signed-rank test: $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$; $^{****}p < 0.0001$; n.s., not significant.

(M.S.Mir, $p < 0.0000$, and M.NS.Mir, $p < 0.0000$) and expressive conditions (NM.S.Exp, $p < 0.0000$) in comparison to the robotic condition (NM.S.Rob). Among the participants, 21 reported that the expressive VA had a positive impact on their performance, while 9 preferred the VA mirroring their movements.

Dominant brain dynamics during motor-related emotional arousal states exhibit scale invariance patterns. Music is often perceived as timeless, having the power to evoke intense emotions and transiently mediating a subjective experience to be connected with others. This distinctive quality of music is explored here by investigating the scale invariance of EEG brain dynamics during the temporal course of the musical piece. We reconstructed the curves $\psi_1^j(\cdot)$ across all the domain and followed the method in [69] who divides the time series into non-overlapping segments at different time scales: a total of 54 levels, factors of 56 s (the duration of the musical piece) ranging from 28 to 0.002 s. We then calculated the mean fluctuation for each segment (see formula no. 4 in [69]) and averaged them at each level. To study the presence of scale invariance, we opted for fitting a continuous log-normal distribution instead of a power law. This choice was motivated by the fundamental nature of lognormal behavior in reflecting the complex structural and functional organization of the brain [50]. Additionally, log-normal distributions offer improved modeling of tail behavior, particularly in the presence of extreme values. Results suggest that dominant brain dynamics during motor-related emotional arousal states are (in general) scale invariant (Table 5.1). Levels of significance were derived following goodness of fit (GOF) tests based on the Kolmogorov-Smirnov statistic [63], via bootstrapping (1000 iterations) on the data. We observe that the results for the gamma band are statistically more consistent, exhibiting an overall better GOF and showing less sensitivity to the level of significance.

Dominant brain dynamics during motor-related emotional arousal states are distinctly turbulent. As a kurtosis value of 3 corresponds to that of a Gaussian distribution, this specific threshold can be taken to determine a cutoff point to discern the transition from the stability inherent in a Gaussian scenario to a turbulent state of non-Gaussian behavior. To investigate the presence of turbulence, our approach involves an information measure based on the differential entropy called Entropic NID (ENID) [45], that

Table 5.1: The table shows the results of fitting a continuous log-normal distribution to the temporal structure of the data. The parameters μ and σ are the estimated mean and standard deviation. The p-value and goodness of fit (GOF) are obtained on a basis of a bootstrapping procedure from the Kolmogorov-Smirnov test ($p > 0.05$ indicates that one can not rule out the log-normal model).

Condition (alpha band)	p-value	GOF	μ	σ
NM.S.NA	0.0060	0.1254	2.3412	0.0181
M.S.NA	0.0960	0.1150	2.3479	0.0061
M.NS.Mir	0.1320	0.0922	2.2919	0.0294
NM.S.Rob	0.1290	0.0960	2.3145	0.0198
NM.S.Exp	0.0660	0.0967	2.2988	0.0278
M.S.Mir	0.8140	0.0868	2.3333	0.0017
Condition (gamma band)	p-value	GOF	μ	σ
NM.S.NA	0.7350	0.1090	2.5173	0.0010
M.S.NA	0.1910	0.1165	2.5505	0.0026
M.NS.Mir	0.6000	0.1072	2.5442	0.0015
NM.S.Rob	0.2850	0.1098	2.5669	0.0024
NM.S.Exp	0.3370	0.1110	2.5660	0.0021
M.S.Mir	0.7280	0.0916	2.5532	0.0017

Table 5.2: Mean turbulence velocity levels on the alpha and gamma band. Comparisons and consistency.

Condition (alpha band)	F	p-value adj.	Power [0-1]	N
NM.S.NA - M	794	0.0000	0.8660	10
NM.S.NA - M.NS.Mir	740	0.0001	0.9550	20
NM.S.NA - NM.S.Exp	669	0.0090	0.9090	30
NM.S.NA - M.S.Mir	812	0.0000	0.8830	10
M - NM.S.Rob	113	0.0000	0.8360	10
M.NS.Mir - NM.S.Rob	178	0.0003	0.9160	20
NM.S.Rob - NM.S.Exp	656	0.0160	0.8240	30
NM.S.Rob - M.S.Mir	810	0.0000	0.8670	10
NM.S.Exp - M.S.Mir	645	0.0250	0.5690	30
Condition (gamma band)	F	p-value adj.	Power [0-1]	N
NM.S.NA - M.NS.Mir	698	0.0020	0.8200	20
NM.S.NA - NM.S.Exp	793	0.0000	0.8080	10
M.S.NA - M.NS.Mir	639	0.0430	0.7400	30
M.S.NA - NM.S.Exp	751	0.0000	0.9470	20
M.NS.Mir - M.S.Mir	178	0.0004	0.8450	20
NM.S.Rob - NM.S.Exp	693	0.0020	0.8600	20
NM.S.Exp - M.S.Mir	101	0.0000	0.8780	10

aims to separate realizations of a random variable in two (as much as possible) statistically independent subsets: here, those kurtosis coefficients of ξ_ℓ^{ij}'s > 3 (for all i, j) attracted to the vicinity of 3, and those who depart from it. By applying ENID to the inverse of the coefficients (the result should then reflect the optimal separation point after 3), divergence from a Gaussian equilibrium occurs at a kurtosis coefficient threshold of 8.2453 and 6.6926 for the alpha and gamma bands respectively. This indicates that turbulence is more prominent in the gamma band, as evidenced by ENID separating faster from 3. As our model yields the maximized kurtosis in time-space for each condition rather than each participant, the counts of values exceeding this threshold are reported on a per-condition basis (proportion of counts alpha; gamma): NM.S.NA 82.2025; 54.9400%; M.S.NA 28.2024; 52.3809%; M.NS.Mir 33.2257; 52.3809%; NM.S.Rob 68.8168; 59.5272%; NM.S.Exp 58.1586; 35.9873%; M.S.Mir 30.9515; 39.0046%.

Lower turbulence intensity reflects higher level of motor-related emotionality. We initially compared potential differences in turbulence intensity within the alpha and gamma bands (results are reported on Table 5.2 and Figure 5.2). Our analysis across both bands revealed varying significance levels, par-

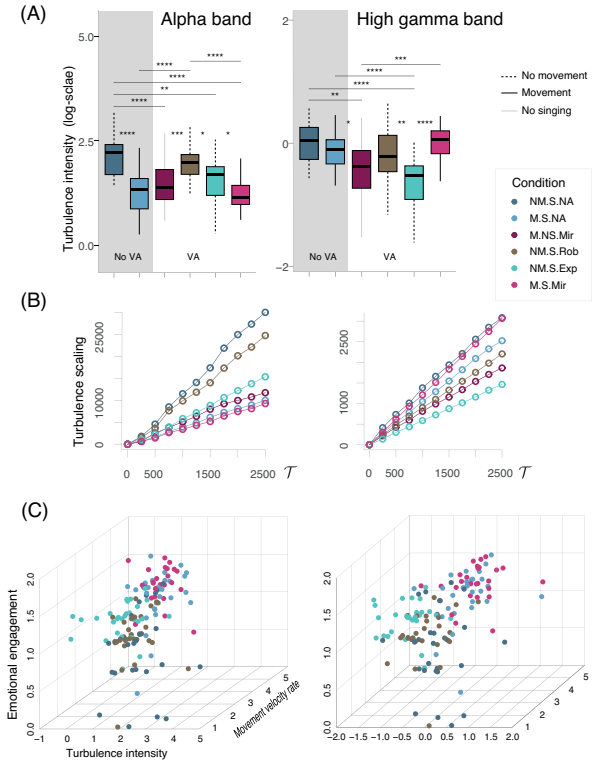

Figure 5.2: Turbulence analysis results. **(A)** Boxplots of turbulence intensity across conditions in alpha (8-13 Hz) and high gamma (50-80 Hz). The centre line of each boxplot represents the data median and the bounds of the box show the interquartile range. The whiskers represent the bottom 25% and top 25% of the data range. **(B)** Scaling of turbulence (i.e. the cumulative sum of the turbulence intensity) across conditions and bands sampled at each 250 observations for a better visualization **(C)** 3D scatterplots showing interactions between the variable emotional engagement, turbulence intensity and movement velocity rates in a log-scale - Statistical significance is measured as detailed in Figure 5.1

ticularly noteworthy in the cases of NM.S.NA and NM.S.Rob, which exhibited elevated contributions to turbulence intensity with respect to the rest of conditions. Further, turbulence intensity was higher for the mirroring conditions (M.NS.Mir, M.S.Mir) and M.S.NA in the gamma band, while NM.S.Exp exhibited the lowest level. For the current results, empirical power of Wilcoxon tests was assessed using simulations on non-Gaussian data $\exp(1) - 1$ to validate the sensitivity of the effects (in Table 5.2, if $N < 30$ it means 100% of reliability) which proved to be highly reasonable. Subsequently, we used a mixed-effects model (MEM) via restricted maximum likelihood leveraging subject/condition variation, focusing exclusively on the singing conditions (only M.NS.Mir was excluded). The response variable was turbulence intensity, while the predictors included the rated level of emotional engagement (EE) and the movement velocity rate (MVR) (see Materials and Methods), all in a log-scale. For the alpha band, the model output (AIC = 239.5063, Log-Likelihood/LL = -114.7531) indicates a significant effect of EE ($\beta = -0.0487, p = 0.048$) and strong effect of MVR ($\beta = -0.2194, p < 0.000$) on turbulence intensity with intercept correlation (IC) of -0.728 for EE and -0.219 for MVR. Note that IC measures the degree of association between the intercept and each fixed effect; no random slopes were included in our models to avoid complexity and overfitting. For the gamma band, the model output (AIC = 183.3267, LL = -86.6633) indicates a significant effect of EE ($\beta = -0.3622, p = 0.0063$) and MVR ($\beta = 0.0987, p = 0.0001$) with a IC of $-0.736, -0.224$ respectivelly. Figure 5.2C shows current interactions in 3D scatterplots.

Lower turbulence intensity in higher gamma band corresponds to level of emotionality and emotional motor control. In the current MEM, we introduced a combined effect of (MVR + Task Demand), where Task Demand (TD) is a predictor that assigns values of 1, 2, or 3 to the conditions according to level of demand (moving, singing, interacting with the VA or any combination of them). For the alpha band, the model output (AIC = 230.794, LL = -110.3972) indicates a strong effect of MVR + TD ($\beta = -0.2066, p < 0.000$) on turbulence intensity with an IC of -0.191. No significant effects were found for EM ($\beta = 0.0274, p = 0.1737$). For the gamma band, the model output (AIC = 188.8453, LL = -89.4226) indicates significant effects of EE ($\beta = -0.3442, p = 0.0085$) and MVR + TD ($\beta = 0.0716, p = 0.001$) on turbulence intensity with an IC of -0.726 for EE and -0.202 for MVR + TD . By adding the inter-

action term EM:Control to the model (AIC = 191.5129, LL = -90.7564), we found effects of EE:Control ($\beta = -0.2143, p = 0.0204$) and MVR + TD ($\beta = 0.0617, p = 0.0014$) on turbulence intensity with an IC of −0.569 and −0.328, respectively.

§ 5.3. Discussion

We have shown that dominant dynamics of EEG oscillatory activity during motor-related emotional arousal are turbulent, scale invariant and therefore working at a dynamical regime of near-criticality. Evidence presented here suggests that turbulence modulation in alpha band activity (8-13 Hz) is fundamentally mediated by motor function and associated level of task demand. We found a dichotomy between alpha and high gamma band (50-80 Hz) dynamics that suggests that, even in the absence of overt bodily movement, emotionality is revealed by the integrated interpretation of both bands.

While alpha-band activity has been a longstanding subject of psychophysiological investigations, its systematic examination in relation to motor-specificity of alpha emotional modulation has only emerged in few studies [105, 119, 278, 310], see also [311]. Extensive work suggests that alpha oscillations play a crucial role in optimizing cognitive resources by selectively dampening neural responses to non-pertinent information [97]. Decline of alpha power in the extended motor system has been shown to engage in neuronal spiking whereas increased alpha power exhibits phase synchronization due to rhythmic inhibition of neuronal firing. This supports the idea that alpha oscillations serve as an informative reflection of the motor system's state, acting as predictive markers of the overall performance [112, 113]. Our findings suggest that alpha turbulence down-modulation, at least in terms of the dominant dynamics under study, is produced by the influences of motor functionality in relation to the level of task demand. This is clearly illustrated by the gradual decrease in turbulence intensity, progressing from no movement during singing, to singing while observing the VA conducting, and ultimately singing while mirroring the movement on the VA (Figure 5.2A). We speculate that this phenomenon could possibly be linked to volume conduction effects arising from low-frequency cholinergic axonal diffusion during states of motor planning and movement [183, 251, 302], or to recent evidence on the control of inhibitory neurons by the cholinergic system [210]. The observed enhancement of alpha turbulence intensity in the NM conditions (Figure 5.2A) is in agreement with previous findings [105, 257] suggesting that synchronization of mu activity (8–12 Hz) in left and midline somatomotor area indicates active inhibition of motor urges, which here aligns with participants' inclination to move during the various singing tasks (25 out of 30 participants reported levels ≥ 3 of involuntary urge to move in restricted movement conditions). The influence of motor afferents in alpha modulation is reinforced by the fact that motor brain regions, particularly basal ganglia, cerebellum, and premotor cortices including area 55b (a restricted region in the right hemisphere), are consistently activated during music listening, even in the absence of overt bodily movement [277].

Although our study did not generate ample statistical evidence to validate turbulence alpha down-modulation as a robust signature of emotional arousal, we found that high gamma activity, as supported by current investigations [55, 316], was a more suitable candidate for this purpose. Aside from the effects of movement and task demand, results in gamma turbulence modulation suggest balanced contributions of emotional engagement and emotional motor control. Particularly relevant to these findings is that high-frequency brain activity supports the existence of GA function [52, 103], often evidenced in humans as a surge of gamma power linked to conscious processing during comatose states [315]. Based on our results (Figure 5.2A), we hypothesize that GA function plays a complex role in shaping motor-related emotional arousal, reflecting a mind-body interplay as proposed in [108] (i.e., coordinating and integrating motor functions to convey emotional expression). Remarkably, participants' sense of emotion (Figure 5.1C) was significantly correlated to down-modulation of turbulence in the gamma range, apparently implying a process of embodiment [169], considering that this was observed in conditions with movement and/or watching natural movement (M.NS.Mir, NM.S.Exp). In complement, NM.S.Exp was the condition in which participants perceived the greatest increase in their performance level, reinforcing the connection between gamma turbulence down-modulation and emotional motor control. The observed dichotomy in condition M.S.Mir, where alpha turbulence decreases while gamma activity intensifies, suggests an increase in attentional demands resulting from more complex social interactions [149, 285, 311], and possibly linked to heightened noradrenergic axonal activity [70, 216]. Interestingly, this dichotomy appears to be compatible with participants' reported level of emotional engagement. Recent investigations have established connections between attention and emotion employing models that link noradrenergic and cholinergic activity to distinct pupillary signatures [22, 302].

Some limitations were identified in the current research. We found challenging to find precise topographical descriptors of the turbulent states under study due to the intermittent and chaotic nature of turbulence. Future efforts may aim to establish tailored measures of dissipation, metastability, func-

tional connectivity, among others, as well as possibly extending the proposed modeling on manifold structures [56, 92, 228] contributing to a better understanding of the spatio-temporal aspects of GA function. Secondly, considering that lower turbulence levels occurred in conditions involving just singing or just watching movement hence implying lower cognitive workload, our study emphasizes the need for further clarification of the role of attention in emotional processing during motor tasks. Using immersive VR is necessary to enhance the tradeoff between internal and external validity in psychophysiological research. Although current VR still constitutes a reductionist version of in-person musical interactions, our work unveils that it can evoke emotionality in simulation contexts when genuine ones are intangible, offering promising clinical and educational applications [54, 240].

§ 5.4. Materials and methods

Participants

We recruited 31 healthy volunteers (mean age, 27.6451 ± 6.5347 years; 21 female) based on the inclusion criteria of being right-handed, having normal or corrected-to-normal vision, normal hearing, and no history of neurological or psychiatric disorders. Participants reported a median of 10 years of musical education (range 1-10 years) despite not being music professionals or trained singers, fulfilling the required musical skills to perform the task. Musical scores were sent to the volunteers 48h prior to the experiment. All of them were requested to restrict caffeine intake or other stimulants in the data collection day. The experiment took ~2h and a compensation of 20€ voucher was given upon completion. After screening, data from one participant was discarded due to bad electrode conductivity leading to a sample of $N = 30$ (mean age, 27.9333 ± 6.443 years; 20 female). Although not included, 4 extra participants took part in pilot tests to asses the feasibility of the experimental procedure. This study was approved by the Ethics Committee of the Faculty of Arts and Philosophy of Ghent University (protocol no. 2022-33). Written informed consent was obtained from all subjects involved in the study.

Screening

Before the experiment, all 31 initial recruits took part in a training and screening session with a laboratory technician and a musical expert. This was done to make sure subjects were capable of singing the music stimuli correctly by heart and moving under a reasonable degree of freedom while using the equipment (mocap suit and VR headsets). After short testing their ability to move, participants were asked to sing the song with the musical accompaniment under the guidance of the musical expert to check whether they were able to perform with fluency. Participants were naive to the purposes of the study, although they were informed that performance quality would not be under analysis, but rather their engagement and perceptions about the activity. Afterwards, instructions were read to the participants before starting the experiment.

Post-hoc performance quality assessment

Following the same procedure as in [302], two musical experts performed an a posteriori quality assessment of the singing recordings. Audio recordings were presented in randomized order and evaluators rated them on a scale from 1 (very inaccurate) to 10 (very accurate) in the following items: intonation, rhythm, fluency, and memory. The discrepancy between the two evaluators on the singers' performance was not significant ($T^2 = 2.875, p = 0.095$). Following Koo and Li's reliability levels [160], we found good reliability of absolute agreement (Intraclass Correlation Coefficient = 0.864) and consistency (ICC = 0.865). No differences were found between the singing performance comparing the two experimental blocks ($T^2 = 1.0896, p = 0.3544$), therefore the evaluators reached the consensus that all participants were able to keep good performance levels.

Task

At the beginning of each trial, key instructions appeared in the virtual environment, such as "Move. Sing along." or "Do not move. Do not sing.", at 3.7 m distance from the participant virtual view and 14 s before starting singing. A cue of 4 beats was included to signal the start of the music performance. Participants had to sing pronouncing "la" (instead of "ta") in order to limit artifactual effects in the EEG signal.

The experimental task was divided in two blocks, respectively, without and with VA. In the first block, there were two conditions: non-movement singing (NM.S.NA) and movement singing (M.S.NA). Prior to the VA block, participants had a training familiarization with the VA, both conducting and mirroring their movements. In the second block, there were four conditions: movement non-singing mirroring VA (M.NS.Mir) and movement singing mirroring VA (M.S.Mir), non-movement singing robotic VA

(NM.S.Rob), and non-movement singing expressive VA (NM.S.Exp). After performing each trial, participants were asked to rate their perceived level of emotional engagement, absorption (degree of correspondence with the activity), control and interaction (only second block) on a Likert scale from 1 "low level" to 5 "high level". They also rated their levels of stress at the beginning and the end of each block, with the intention of discarding trials in which levels above 3 were reported. No participant reported levels of stress > 3. The distribution of the ratings was: level 1 - 60.4838%, level 2 - 25.8064%, level 3 - 13.7096%. As in [119], we further asked participants to indicate us if they felt general discomfort, nausea, dizziness, headache, blurred vision, and difficulty concentrating. Some participants felt discomfort in the nasal area due to the headset's weight, which was cautioned by accommodating a cushion between the nose and the headset's support zone. At the end of the experiment, participants were asked a forced-choice question, "Which of the agent behaviors (robotic, expressive, mirror, or none of them) do you believe had the most positive impact on your singing performance?" to reiterate their VA preferences. Lastly, we ask them to rate the involuntary urge to move during the conditions they were not allowed to.

Data acquisition

Participants' movement was recorded with a 16 -infrared camera optical motion capture system (Qualisys, Sweden) using a sampling rate of 120 Hz. The acquisition software was the Qualisys Track Manager (QTM) 2023. Participants wore a suit where 42 reflective markers were placed following the Qualisys full body biomechanical model. This model was adopted due to its capability of producing realistic projections of the subjects' movements.

EEG data was recorded at 1 kHz with ANT-Neuro *eeg*^{TM} *mylab* systems using a 64 channel headset (10-10 system, with Ag/AgCl electrodes). One electrooculogram (EOG) electrode was placed below and next to the right eye. Recordings were conducted using a referential montage, with electrode CPz as reference. To reduce tension on electrode cables and allow unrestricted movements, the amplifier was positioned on an elevated table behind 1 m distance of the participant. Mobility was measured to freely sway (with sufficient cable length) one step ahead and laterally. Impedance levels were monitored using the eego^{TM} software to ensure they remained below 20 kΩ.

The VR headset was carefully placed over the EEG cap. Participants were equipped with HTC Vive Pro 2 headsets and followed the standard calibration procedure recommended by the manufacturer. The mocap data were streamed to a standard digital audio workstation software (Ableton Live 9) for synchronization with Unity (Unity Technologies, consumer version 2023.2.13) allowing the VA to mirror participant's movements. Vocal performance was recorded using a Shure Beta 87A microphone placed in the ceiling above the participant. Additionally, a decibel meter (UNI-T UT352) was employed to oversee and assess the volume levels before the commencement of the experiment, aiming to mitigate the impact of loudness.

Stimuli

The virtual environment was designed in Unity. We used a gender-neutral VA in a room with plane size limited at 5×10 m with low visual impact colors [259]. Participants were standing in the middle of the room 3.7 m distance from the VA. Initially, a light grey cross was projected on to the middle of the scene to help participants fix their gaze on a point. The movements of the VA conducting robotically or expressivelly were recorded previously from a professional conductor instructed to perform the gesturing accordingly. Auditory stimuli were the same as in our previous experiment (see Supplementary Material in [302]), where participants found it easy to memorize and adapt to their tonal range. See supplementary material.

Pre-processing of neuroimaging data

All pre-processing was performed in R [244] using custom made scripts. Routines were conducted separately for data recorded per participant and condition.

Detecting and removing artifacts in EEG signals during vocal tasks and body sway poses a complex challenge. Particularly, singing implies generation of artifacts from hypoglossal movement, involuntary clenching and by contractions of the neck and facial muscles. In addition, blink activity tends to be more prevalent in such conditions [302]. On the other hand, body movements can induce cable sway, muscle tension, and heightened heart rate variability. The critical mixture of artifactual potentials over the scalp field due to volume conduction, rather requires that the method used for identification and removal suitably adjusts to their topological features in time and space. The approach considered here bears resemblance to a multi-band component analysis [137], which allows to target artifacts according to the band-width in which they arise more predominantly. In a first stage, line noise interference was removed using a fourth-order Pei-Tseng notch filter centered at 50 Hz on the raw signal and FastICA (PCA whitening, parallel extraction with logcosh) was performed on the data high-pass filtered at 20 Hz (forward-backward 4th

order Butterworth filter - 4Bw), a common low threshold for the spectral bandwidth of muscle activity [206]. In order to select artifactual components, we inspected spatial topographies and applied wavelet shrinkage (MODWT-l4 [232]) to the vectors of the source matrix and visualized them using line references of the timmings when the participants pronounced "la" to change pitch. We enhanced their selection considering a median standard deviation threshold (as weighted by the number of channels containing absolute voltages > 100μV) of the norm of the transformed source vectors with the Teager-Kaiser operator (TKO) [143]. The selected components were individually backprojected for removal and denoised through wavelet shrinkage, as per the methodology outlined in [298], aiming to minimize modulation and preserve brain activity to the greatest extent possible. On average, 19.0776 denoised and temporally-sparse high-frequency components per subject were removed, which is considered a reasonable number according to current investigations in the area [178, 206].

Subsequently, adaptive notch (bandwidth: 0.1) filtering [25] was performed on the data around spectral peaks exceeding the default threshold of 8 standard deviations using non-overlapping widows of a 50 Hz step from 48 Hz via FFT. We followed this protocol since muscle activity can mask sources of line noise (possibly produced by the VR headset system [312]) while induce their spectral distortion. Channels containing absolute voltages > 100μV above 20 Hz were denoised (MODWT-Cl4) on the four coarsest decomposition levels (> 30 Hz) by shrinking to 0 the coefficients surpassing the universal threshold of their related TKO transformation. For the sake of smoothness, a Gaussian kernel was applied to these decomposition levels using Scott's bandwidth after performing the shrinkage. Pathological cases of noise corrupted channels were visually inspected and reconstructed via spline spherical interpolation.

In a second stage the data was referenced to robust average and FastICA was conducted on the broad-band pre-processed signal after a PCA reduction. A PCA usually enhances the estimation of high amplitude components corresponding to blinks, body movements or cable sway, by restricting their mixture with other PCs when ICA is performed. Outlying and sparsest spatial components were semi-automatically detected using the norms and the index of sparsity defined in [322] on the vector columns of the estimated mixing matrix. To minimize the impact on brain activity, wavelet denoising (l4) was once again applied. The removal process was evaluated using a depth statistics [65] on the median absolute voltage and the norm of the Fourier spectrum (1-15 Hz) pooling all channels across subjects and conditions. Outlying channels and associated trials were inspected to further detect artifactual ICA components which were removed until sufficient depth consistency was achieved (only mastoids were left as residual outliers, which were not used in our analyses). Results were visually validated and, on average, 1.7583 artifactual components were removed from the signal. In a last round, the signal was again examined with ICA to identify residual artifactual activity and cardiogenetic components, whose detection is known to be improved under more stationary conditions.

Robust subband estimation and EEG data representation

The pre-processed EEG data was filtered (4Bw) in the alpha (8-13 Hz) and high gamma (50-80 Hz) bands using spatio-spectral decomposition (SSD) [214]. This technique aims to find linear filters maximizing power in the frequency band of interest while minimizing power in the neighboring "flanking" frequency bins. Prior studies have shown the ability of the technique to enhance the robustness and interpretability of results [51, 119, 267], as well as to increase classification accuracy in BCI applications [116]. The SDD parameters were set as follows: bandpass signal ±2 Hz, bandstop noise ±1 Hz, bandpass noise ±4 Hz. The number of SDD components (filters) were selected via perpendicular line method, resulting in an average of 35.9222 and 37.8388 filters being retained for each band. Subsequently, Cauchy robust PCA [91] was applied to the filtered signal to represent the data in terms of the dominant spatial eigenvector. This technique, well-suited for high-dimensional data, utilizes a Cauchy likelihood instead of a Gaussian likelihood to ensure maximum robustness in component estimation, thereby notably reducing the leakage of possible artifact residual.

Hilbertian spatio-temporal model of dominant dynamics for turbulence analysis

Suppose we have the EEG data $\{X_{ij}(t_1, s_k), \ldots, X_{ij}(t_m, s_k), \ldots X_{ij}(t_M, s_K)\}_{i,j=1}^{N,J}$, where N are the number of participants, J the number of experimental conditions and K the number of channels. The matrices $(X_{ij})_{K\times M}$ had been filtered using the spatio-spectral decomposition at the frequency band of interest and then expressed in terms of the first spatial Cauchy principal component. Although the data is observed at $t_1, \ldots, t_M$ time points, we assume these are realizations of J spatio-temporal random variables taking values on the space of square integrable functions $L^2_{T\times S}$. Given that our data comes in a wide format ($N \lll M$), the interval $T \subset \mathbb{R}$ representing the temporal domain is partitioned in $\mathcal{T}_1, \ldots, \mathcal{T}_\ell, \ldots, \mathcal{T}_\tau$ subintervals, possibly intersecting in a region of its domain.

Consider the basis expansions approximations at any $\mathcal{J}$ of $N \cdot K$ curves $X^j(t) = A^j\phi(t), t \in \mathcal{J}$ where A^j is $N \cdot K \times p$ coefficient matrix of the participant's curves pooled per channel and $\phi(t) = (\phi_1(t), \ldots, \phi_p(t))^\top$ is a basis of p functions. Note A^j represent sequences of matrices in the direction of the experimental conditions and X^j their respective vector of functions. We will work with the expansion $\boldsymbol{X}(t) = \mathcal{A}\Phi(t)$ where $\boldsymbol{X}(t) = [X^1(t); \ldots; X^J(t)]_{N\cdot K\times J}$, $\mathcal{A} = [A^1; \ldots; A^J]_{N\cdot K\times p\cdot J}$ and $\Phi(t)$ is matrix of size $p \cdot J \times J$ with J p-dimensional basis functions in its diagonal.

Following [296], we can map $\boldsymbol{X}(t)$ to orthogonality, i.e, $\boldsymbol{X}(t) \to \mathbb{X}(t) : \operatorname{cov}(\mathbb{X}(t)) = \mathrm{Id}.$, using the factorization of the matrix $\mathcal{G} = \int_{\mathcal{J}} \Phi(t)\Phi(t)^\top \mathrm{d}t$ and its inverse. Then, we consider the projections

$$\xi^j = \int_{\mathcal{J}} X^j(t)\psi_1^j(t)\mathrm{d}t, \tag{5.1}$$

where $\psi_1^j(t)$ is a function embedded in the dominant eigenfunction $\psi_1(t)$ obtained via spectral decomposition of the kurtosis kernel function admitting the basis expansion

$$\operatorname{kurt}(\mathbb{X})(t,\cdot) = \Phi(\cdot)^\top\mathcal{G}^{-1/2}\underbrace{\left(\frac{1}{N\cdot K}\mathcal{G}^{1/2}\tilde{\mathcal{A}}^\top D\tilde{\mathcal{A}}\mathcal{G}^{1/2}\right)}_{\Sigma_{\tilde{\mathcal{A}}\mathcal{G}^{1/2}}}\mathcal{G}^{-1/2}\Phi(t), \tag{5.2}$$

where $D = \operatorname{diag}\left(\tilde{\mathcal{A}}\mathcal{G}\tilde{\mathcal{A}}^\top\right)$, $\tilde{\mathcal{A}}$ is the a coefficient matrix $\mathcal{A}$ after whitening and $\Sigma_{\tilde{\mathcal{A}}\mathcal{G}^{1/2}}$ is its kurtosis matrix.

By solving the eigenvalue problem $\Phi(t)^\top\Sigma_{\tilde{\mathcal{A}}\mathcal{G}^{1/2}}\mathcal{G}\boldsymbol{b}_s^\top = \kappa_s\Phi(t)^\top\boldsymbol{b}_s^\top$, we get a set of eigenvalues $\kappa_1 \geqslant \cdots \geqslant \kappa_{p\cdot J}$ and associated eigenvectors $\boldsymbol{b}_s$ which allow to compose $\psi_s(t) = \Phi(t)^\top\mathcal{G}^{-1/2}\boldsymbol{b}_s^\top$, the eigenfunctions of $\operatorname{kurt}(\mathbb{X})$ which have unit norm and are pairwise orthogonal. If we take the following division of the coefficients $[b_1, \ldots, b_p; b_{p+1}, \ldots, b_{2\cdot p}; \ldots; b_{(N\cdot K-1)\cdot p+1} \ldots, b_{N\cdot K\cdot p}]$, one can easily obtain the dominant functions $\psi_1^j(t)$ by expanding each coefficient trunk by $\phi(t)$ and obtain Equation 5.1.

In our model, the functions $\psi_1^j(t)$ are projected on to each univariate functional dataset X^j, thus preserving the participant's dimension across conditions, albeit at the cost of having non-uncorrelated projection scores. By subsequently performing at each $\mathcal{J}_\ell$ the above operations, we will get τ realizations of a discrete spatio-temporal random field, i.e, $\Xi^{ij} \equiv \{\xi_1^{ij}; \ldots; \xi_\ell^{ij}; \ldots; \xi_\tau^{ij}\}$ where ξ_ℓ^{ij} is a univariate vector of K entries that has been reorganized participant wise. The differentiation of Ξ^{ij} in time gives the gradient flow or fluctuation matrix and for all $\ell > 1$, the turbulence intensity is defined as

$$\left\langle \|\xi_\ell^{ij} - \xi_{\ell+1}^{ij}\|^2 \right\rangle_\ell, \tag{5.3}$$

where $\langle\cdot,\cdot\rangle_\ell$ indicates average in the temporal direction.

Turbulence model setup

For a system of overlapping windows, we determined a hop size of 20 ms according to the latencies that characterize interneuronal information transmission [132]. Window sizes of 500 and 100 ms for the alpha and gamma band were respectively used to perform multivariate functional ICA (with ZCA whitening [296]). This choice aimed to uphold a consistent ratio of neural fluctuations in each window while mitigating the risk of numerical instabilities in the estimation of the covariance function in the functional ICA model. We regressed out the data using B-spline basis functions keeping towards 0 the RMSE in the approximation. For the reconstruction of $\psi_1^j(\cdot)$ across all domain, an overlapping Gaussian window with width factor of 4 was used, and the hop size was determined as half the window size.

Pre-processing of motion capture data

Motion capture data was initially pre-processed in QTM 2023 (Qualisys AB, Sweden) for marker labeling, gap-filling and trajectory smoothing (10 Hz low pass Butterworth filter). Marker trajectories were exported and movement velocity was then calculated as the first-order time-derivatives of the marker positions. Velocity data was then normalized across the three axis to produce the magnitude velocity of each marker. We then applied the Minimum Covariance Determinant estimator [258] to the pre-processed spatial data, obtaining location estimates for each marker. These estimates were then median-averaged to calculate the movement velocity rate per participant and condition. Data from two participants were excluded from the MEM analyses due to technical issues with tracking.

Acknowledgements: We are grateful to Vadim Nikulin for his suggestions on the data pre-processing and to C. Gener and B. van Kerrebroeck for their support during data collection. **Author contributions:** Conceptualization, M.V. and M.L.; methodology, M.V.;

software, M.V.; validation, M.V. and A.A.; formal analysis, M.V., N.M., A.A. and M.L.; data curation, M.V.; writing—original draft preparation, M.V.; writing—review and editing, N.M., A.A., T.F. and M.L.; supervision, A.A., T.F. and M.L.; funding acquisition, A.A. and M. L. **Data accessibility:** Anonymized data has been deposited at an OSF repository (https://osf.io/c3t78). **Conflict of interest:** The authors declare no conflict of interest.

6 | Conclusions and research perspectives

§ 6.1. Contributions in mathematical and applied statistics

The research presented in this thesis represents a notable step forward in the field of functional ICA, building upon and extending the findings of previous studies [176, 197, 231, 305]. One of the main contributions lies in the utilization of Sobolev norms to regulate the smoothness level of the ICs estimates within the subspaces generated by the covariance/kurtosis eigenfunctions. Concurrently, we have developed an extension of functional ICA for wide functional data in the presence of spatial dependencies using wavelet expansions and customized shrinkage techniques for denoising. Overall, the proposed smoothing/denoising techniques have demonstrated efficacy in functional classification and representation of functional anomalies. By controlling the noise of the estimates, we effectively enhance the irreducibility of the components, a quality that has been described critical in estimating true independent components [111]. The success of the separation in ICs strongly depends on both the characteristics of the data and the whitening method employed. In this regard, we have generalized typical multivariate whitening procedures to infinite-dimensional function spaces, offering versatility to optimally explore functional data beyond second-order correlations. In functional classification, previous research has shown that methodologies achieving effective variance reduction per group, such as partial least squares [3, 6, 77, 78], variable selection [33, 34], or grup-wise smoothing as shown in §2.8, tend to yield superior outcomes. Guided by our theoretical insights (§2.7), classification strategies should probably align towards the principles of Feldman-Hájek Theorem and its derivations to attain acceptable levels of accuracy. Nevertheless, in unsupervised settings things become more complicated. We foresee that a rigorous stochastic study of the definition of Hilbertian independence, IC separability and moment operators could pave the way for understanding relationships among linear/non-linear independent structures in the data critical to unveil significant low-dimensional projections. Our research on functional ICA has further motivated us to accommodate our methodologies to more complex structures, encompassing different kinds of multivariate functional data (as for example we do in Chapter 5) as well as on manifold structures [56, 228]. After the results here obtained, it also would be interesting to explore avenues for inference with functional ICA similarly as in [1, 2, 260].

Regarding the asymptotic behavior of the proposed functional IC models, there are various points to take into consideration. Functional ICA is commonly defined in the RKHS generated by the covariance operator (i.e., the covariance operator's closed range), and therefore based on functional PCA estimates. While the asymptotic convergence of functional PCA has been extensively studied, see for example [38, 42, 73, 114, 123], addressing the smoothed case demands a more nuanced approach. Following [276] or [124], which extended Silverman's results on smoothed functional PCA consistency by incorporating additional Tikhonov regularization, most remarkably, Lakraj and Ruymgaart [167] derived its asymptotic properties using results on the perturbed eigensystem of a sample smoothed covariance operator to further investigate the consistency and asymptotic distributions for the first smoothed principal components and corresponding eigenvalues. Nevertheless, in smoothed functional ICA models, finite sample bounds (intermediate regimes) seem to be a more realistic choice of study, as all covariance eigenvalues exceeding n are consistently zero, thus hindering a straightforward application of typical asymptotic results. Further, Picard's law is likely to hold under particular regularity conditions. In this vein, Virta et al. [305] suggested that asymptotic convergence of functional ICA estimates could be obtained by leveraging the rank of $\mathscr{P}_\gamma$ together with the sample size via the operator norm $||\Gamma_j(F_n) - \Gamma_j(F)||_{\text{HS}} = \mathcal{O}_p(h_n) \quad (j = 1, \cdots, q)$, where $\Gamma(F) = (\psi_j \otimes \psi_j)\Psi(\mathscr{P}_\gamma X)$, F_n is the empirical distribution function of a sample of size n from the distribution F, and h_n is some sufficient rate of convergence. Alternatively, one can also build upon prior consistency results in functional canonical correlation (see, e.g., [59, 66, 102]), or other sibling techniques [12, 32, 261]. Parallel to this, Li et al. [176] studied the Fisher consistency of the kurtosis operator, similar to [130, 234], for the multivariate case. This is closely related to the approach we have employed to study the consistency of the proposed whitening procedures (§2.8). We hope to provide more extensive and rigorous results in the future.

Finally, we wish to emphasize that our methods, including all whitening procedures and functional ICA models, have been implemented in the R package pfica [297], which is available on CRAN.

§ 6.2. Some remarks on pre-processing neuroscientific data during embodied interactions

One of the most challenging objectives of this dissertation was to design an effective method for pre-processing neuroscientific data collected from non-invasive devices during complex motor tasks. This involved conducting numerous tests in the laboratory, and performing thorough examinations of the recorded data across multiple motor conditions to determine the extent of our capabilities. Undoubtedly, pre-processing data collected from neurophysiological monitoring techniques is currently one of the primary challenges confronting the development of embodied music neuroscience.

In Chapter 4 and Chapter 5, we provided an algorithm for unsupervised removal of ROEs in pupil signals based on non-linear polynomial optimization, and a wavelet-based spatial functional ICA for detecting and characterizing artifacts in EEG signals. Both methods yielded satisfactory results, enhancing to a great extent the estimation quality of brain signal and the applicability of subsequent statistical techniques. Results in simulations and real-data analyses speak for themselves: the use of smoothing/denoising techniques is indispensable for deriving reliable neuroscientific results. However, there are some points of discussion. During the preprocessing of the EEG data from Study II (Chapter 5), we observed a notable improvement in the application of the method described in Chapter 3 when applied independently across multiple frequency bands, thereby assuming artifacts can be stereotiped frequency-wise. Considering the highly complex mixture of artifacts in the signal during motor tasks, effectively estimating certain artifacts in a single spatial ICA decomposition proved to be nearly unfeasible. Conducting spatial ICA in frequency sublayers and integrating it with a PCA reductions/wavelet techniques offered more precise artifact estimates. However, more thorough study on the filtering methods and frequency ranges is needed to improve, validate and generalize the procedure. We wonder whether it is possible to integrate current developments into rigorous theoretical precepts that allow us to design a functional ICA covering all these aspects discussed. Future comparative and validation studies can certainly help to this endeavor, further simplifying the laborious (and often subjective) task of manually selecting artifacts. We believe the methods developed in this dissertation provide robust starting tools for researchers aiming to analyze brain signals under movement conditions. The proposed methodologies have however raised additional questions regarding their suitability for real-time estimation of brain signals, particularly for neurofeedback optimization oriented towards action-control and emotion regulation. A primary concern revolves around the adequacy and minimization of computational time needed for pre-processing in real-time scenarios, ensuring the usability of the neurofeedback system.

From a more fundamental perspective, our concern lies in understanding how concurrent repeated motor actions (e.g., walking or tapping) modulate the EEG signal, and in distinguishing genuine brain activity from artifact-induced activity. Preliminary analysis indicates that certain oscillatory features, temporally associated with these types of motor actions, may be either artifactual or represent basic neuroanatomical responses, potentially rendering them spurious or prone to overinterpretations. This is because EEG signals not only capture muscular activity originating from areas beyond the immediate recording area but also reflect volume conduction effects, such as produced by low-frequency axonal diffusion involved in the functionality of pre-motor areas [301]. This activity could be mixed with typical artifactual volume conduction effects making even more complex to accuratelly separate both processes in a spatio-temporal sense. Therefore, a rigorous study is expedient, possibly exploiting joint analyses with mocap data and the assessment of the EEG signal with more sophisticated spatio-temporal/source localization techniques [185, 226], or approaches that can provide cross-frequency insights [107, 215, 319]. Nevertheless, and to be realistic, we anticipate that invasive techniques will play a critical role in addressing some of these questions in the coming years.

§ 6.3. Understanding brain behaviour during embodied musical interactions: preliminary contributions to the field of embodied music neuroscience

Efforts undertaken in this investigation have been primarily aimed at uncovering neuroscientific evidence supporting the postulates of embodied music cognition, a paradigm spearheaded by Prof. Marc Leman over the past 20 years of his career. Often, embodiment is thought to be best investigated in terms of synchronization [47, 169, 171, 172, 293]. However, here we describe a different approach to investigating the neuroscience of embodiment, focusing on a broader spectrum of effects that can be measured when engaging with music. We all know that music is more than just a metronome, and that listening to music is driven by inherently emotional motivations. Therefore, investigating emotional effects mediated through music, along with the associated motor responses evoked by music listening and music making,

in our opinion, is better suited for studying the neurology of embodiment in these scenarios. As a result, a paradigm was developed, viewing emotion as an isomorphism of embodiment. This idea came up after long discussions with Prof. Thomas H. Fritz on the relevance of both emotional and voluntary motor systems in shaping embodied interactions (see [120, 295], also §4.1, 4.3, 5.1). It was natural to investigate CNS arousals to elucidate the neurological underpinnings of embodied music cognition, here understood as a construct of complex brain-CNS-body interactions, seemingly as postulated in the theory of constructed emotion by Barrett [24].

In our initial study (Chapter 4), we were among the first to propose analyzing pupil dilation at specific frequencies, building on prior research relating pupil signals to putative neurotransmitter activity during motor tasks. This way of modeling the data, often referred to as a blinded inference methodology [283], was bolstered by the principles of the generalized CNS arousal hypothesis (see §4.1, 5.1). With this, we aimed at quantifying emotional response during embodied interactions and examining potential relations with agency/control. We found that cholinergic-related pupil activity reflected the perceived level of emotionality as well as, possibly, the level of physical task demand, and that noradrenergic pupil-related activity consistently scaled down during conditions varying in emotionality of motor task, which is often interpreted as a signature of agency. Another intriguing aspect of the results was the comparable modulation of low-frequency pupil activity observed during conditions where only bodily sway or singing was permitted, but not during a combination of both. This suggests that slow pupil frequencies are also sensitive to different motor afferents.

The findings from this initial study prompted a second investigation (Chapter 5), where we intentionally maintained the basic experimental design to validate our previous findings. Indeed, Study I generated questions around the role of physical task demand in relation to the perceived level of emotionality. Following some recent studies [54, 199], we were further interested in how musical interactions in immersive virtual reality could have an impact on the level of emotionality. Specifically, we examined interactions with a virtual agent (VA) conducting music in different expressive fashions, with the expectation that neural engagement would decrease as a result of an embodiment process. To prove this, we used EEG recordings and formulated a turbulence model that combines dominant spatial representations and functional data. We chose this modeling approach since neuromodulatory activity is often described to be spatially diffuse [21], therefore the examination of a turbulent behaviour in the cortical field was a potential way to measure GA function. Our analyses provided evidence that turbulence down-modulation of alpha band reflected physical arousal, a factor undoubtedly related to the perceived level of emotionality as observed in Study I, which may easily be misunderstood as a signature of emotionality. Instead, down-modulation of turbulence in the high gamma band consistently corresponded to emotionality conditional to level of agency, here induced by the processing of expressive gestures from the VA.

Put together, our findings suggest that embodiment arises from complex bidirectional interactions between the CNS/brain and body behavior in response to environmental cues. Modulation of higher frequencies of pupil and EEG signals reflect this phenomenon, evidenced through a negative scaling or down-modulation of neural activity during embodied interactions. Nevertheless, embodiment is contingent to the quality of the interaction being able to ensure optimal brain functionality by allocating resources efficiently, which may include, say, delegating tasks to the body (under minimal motor processing) or other external agents (social,virtual, etc.) that ultimately facilitate neural processing and a fluent experience of music. This enhancement seems to occur particularly when interactions are emotionally engaging, emphasizing that while movement synchronization to music may be significant in these processes, it is not a sufficient condition for embodiment to be experimentally validated. Current notion of embodiment echoes key principles discussed in Leman's 2016 book [171], highlighting the essential role of expressivity in musical interactions, but now reflected through neurological parameters. Future research aims to uncover the physiological underpinnings of our findings, elucidating the role of particular brain regions, as well as sensory receptors and their corresponding neural pathways, in shaping embodied music interactions. Looking ahead, the insights gained from our research have the potential to facilitate further studies in this field and pave the way for optimizing neurofeedback applications across clinical, educational, and recreational settings.

Bibliography

[1] Acal, C. and Aguilera, A. (2023). Basis expansion approaches for functional analysis of variance with repeated measures. *Advances in Data Analysis and Classification*, **17**:291–321.

[2] Acal, C., Aguilera, A. M., Sarra, A., Evangelista, A., Battista, T. D., and Palermi, S. (2021). Functional ANOVA approaches for detecting changes in air pollution during the covid-19 pandemic. *Stochastic Environmental Research and Risk Assessment*, **36**(4):1–19.

[3] Aguilera, A. M., Aguilera-Morillo, M., and Preda, C. (2016). Penalized versions of functional PLS regression. *Chemometrics and Intelligent Laboratory Systems*, **154**:80–92.

[4] Aguilera, A. M. and Aguilera-Morillo, M. C. (2013). Penalized PCA approaches for B-spline expansions of smooth functional data. *Applied Mathematics and Computation*, **219**(14):7805–7819.

[5] Aguilera, A. M., Escabias, M., Ocaña, F. A., and Valderrama, M. J. (2015). Functional wavelet-based modelling of dependence between lupus and stress. *Methodology and Computing in Applied Probability*, **17**(4):1015–1028.

[6] Aguilera, A. M., Escabias, M., Preda, C., and Saporta, G. (2010). Using basis expansions for estimating functional PLS regression: Applications with chemometric data. *Chemometrics and Intelligent Laboratory Systems*, **104**(2):289–305.

[7] Aguilera, A. M., Gutiérrez, R., Ocaña, F. A., and Valderrama, M. J. (1995). Computational approaches to estimation in the principal component analysis of a stochastic process. *Applied Stochastic Models and Data Analysis*, **11**(4):279–299.

[8] Aguilera, A. M., Gutiérrez, R., and Valderrama, M. J. (1996). Approximation of estimators in the PCA of a stochastic process using B-splines. *Communications in Statistics - Simulation and Computation*, **25**(3):671–690.

[9] Aguilera, A. M., Ocaña, F. A., and Valderrama, M. J. (1997). An approximated principal component prediction model for continuous-time stochastic processes. *Applied Stochastic Models and Data Analysis*, **13**:61–72.

[10] Akhtar, M. T., Mitsuhashi, W., and James, C. J. (2012). Employing spatially constrained ICA and wavelet denoising, for automatic removal of artifacts from multichannel EEG data. *Signal Processing*, **92**(2):401–416.

[11] Albergaria, C., Silva, N. T., Pritchett, D. L., and Carey, M. R. (2018). Locomotor activity modulates associative learning in mouse cerebellum. *Nature Neuroscience*, **21**(5):725–735.

[12] Álvarez Liébana, J., Bosq, D., and Ruiz-Medina, M. (2017). Asymptotic properties of a component-wise ARH(1) plug-in predictor. *Journal of Multivariate Analysis*, **155**:12–34.

[13] Amato, U., Antoniadis, A., Feis, I. D., and Gijbels, I. (2022). Penalized wavelet estimation and robust denoising for irregular spaced data. *Computational Statistics*, **37**(4):1621–1651.

[14] Angulo, J. and Ruiz-Medina, M. (1997). On the orthogonal representation of generalized random fields. *Statistics & Probability Letters*, **31**(3):145–153.

[15] Arribas-Gil, A. and Romo, J. (2014). Shape outlier detection and visualization for functional data: the outliergram. *Biostatistics*, **15**(4):603–619.

[16] Aston-Jones, G. and Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience*, **28**:403–450.

[17] Auksztulewicz, R. and Friston, K. J. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex*, **80**:125–140.

[18] Averkamp, B. R. and Houdré, C. (2003). Wavelet thresholding for non-necessarily Gaussian noise: idealism. *Annals of Statistics*, **31**(1):110–151.

[19] Bach, F. R. and Jordan, M. I. (2002). Kernel independent component analysis. *Journal of Machine Learning Research*, **3**:1–48.

[20] Balakrishnan, A. V. (1976). *Applied Functional Analysis*. Springer-Verlag, New York.

[21] Ballinger, E. C., Ananth, M. R., Talmage, D. A., and Role, L. W. (2016). Basal forebrain cholinergic circuits and signaling in cognition and cognitive decline. *Neuron*, **91**(6):1199–1218.

[22] Bang, D., Luo, Y., Barbosa, L. S., Batten, S. R., Hadj-Amar, B., Twomey, T., Melville, N., White, J. P., Torres, A., Celaya, X., Ramaiah, P., McClure, S. M., Brewer, G. A., Bina, R. W., Lohrenz, T., Casas, B., Chiu, P. H., Vannucci, M., Kishida, K. T., Witcher, M. R., and Montague, P. R. (2023). Noradrenaline tracks emotional modulation of attention in human amygdala. *Current Biology*, **33**(22):5003–5010.e6.

[23] Bangert, M. and Altenmüller, E. (2003). Mapping perception to action in piano practice: a longitudinal DC-EEG study. *BMC Neuroscience*, **4**(26):1–14.

[24] Barrett, L. F. (2016). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, **12**(1):1–23.

[25] Bedoyan, E., Reddy, J. W., Kalmykov, A., Cohen-Karni, T., and Chamanzar, M. (2023). Adaptive frequency-domain filtering for neural signal preprocessing. *NeuroImage*, **284**:120429.

[26] Bell, A. J. and Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Research*, **37**(23):3327–3338.

[27] Benarroch, E. E. (2009). The locus ceruleus norepinephrine system. *Neurology*, **73**(20):1699–1704.

[28] — (2021). Cholinergic Transmission. In *Neuroscience for Clinicians: Basic Processes, Circuits, Disease Mechanisms, and Therapeutic Implications*. Oxford University Press.

[29] Bentley, P., Vuilleumier, P., Thiel, C. M., Driver, J., and Dolan, R. J. (2003a). Cholinergic enhancement modulates neural correlates of selective attention and emotional processing. *NeuroImage*, **20**:58–70.

[30] — (2003b). Effects of attention and emotion on repetition priming and their modulation by cholinergic enhancement. *Journal of Neurophysiology*, **90**(2):1171–1181.

[31] Berkooz, G., Holmes, P., and Lumley, J. L. (1993). The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics*, **25**:539–575.

[32] Berrendero, J. R., Bueno-Larraz, B., and Cuevas, A. (2020). On Mahalanobis distance in functional settings. *Journal of Machine Learning Research*, **21**(1):1–33.

[33] Berrendero, J. R., Cuevas, A., and Torrecilla, J. L. (2016). Variable selection in functional data classification: A maxima-hunting proposal. *Statistica Sinica*, **26**(2):619–638.

[34] — (2018). On the use of reproducing kernel Hilbert spaces in functional classification. *Journal of the American Statistical Association*, **113**(523):1210–1218.
[35] Berridge, C. W. (2008). Noradrenergic modulation of arousal. *Brain Research Reviews*, **58**(1):1–17.
[36] Besse, P. and Ramsay, J. O. (1986). Principal components analysis of sampled functions. *Psychometrika*, **51**(2):285–311.
[37] Bickel, P. J., Kur, G., and Nadler, B. (2018). Projection pursuit in high dimensions. *Proceedings of the National Academy of Sciences*, **115**(37):9151–9156.
[38] Blanchard, G., Bousquet, O., and Zwald, L. (2007). Statistical properties of kernel principal component analysis. *Machine Learning*, **66**:259–294.
[39] Blood, A. J. and Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences*, **98**(20):11818–11823.
[40] Bogachev, V. I. (1998). *Gaussian Measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, New York.
[41] Bombeke, K., Duthoo, W., Mueller, S. C., Hopf, J.-M., and Boehler, C. N. (2016). Pupil size directly modulates the feedforward response in human primary visual cortex independently of attention. *NeuroImage*, **127**:67–73.
[42] Bosq, D. (2000). *Linear Processes in Function Spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York.
[43] Bowling, D. L., Gahr, J., Ancochea, P. G., Hoeschele, M., Canoine, V., Fusani, L., and Fitch, W. T. (2022). Endogenous oxytocin, cortisol, and testosterone in response to group singing. *Hormones and behavior*, **139**:105105.
[44] Breton-Provencher, V. and Sur, M. (2019). Active control of arousal by a locus coeruleus gabaergic circuit. *Nature Neuroscience*, **22**(2):218–228.
[45] Bruni, V., Cioppa, L. D., and Vitulano, D. (2020). An automatic and parameter-free information-based method for sparse representation in wavelet bases. *Mathematics and Computers in Simulation*, **176**:73–95.
[46] Brych, M., Murali, S., and Händel, B. (2021). How the motor aspect of speaking influences the blink rate. *PLoS One*, **16**(10):1–19.
[47] Buhmann, J., Moens, B., Van Dyck, E., Dotov, D., and Leman, M. (2018). Optimizing beat synchronized running to music. *PLoS One*, **13**(12):e0208702.
[48] Buzsáki, G. (2006). *Rhythms of the Brain*. Oxford University Press, New York.
[49] Buzsáki, G., Anastassiou, C., and Koch, C. (2012). The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, **13**:407–420.
[50] Buzsáki, G. and Mizuseki, K. (2014). The log-dynamic brain: how skewed distributions affect network operations. *Nature Reviews Neuroscience*, **15**:264–278.
[51] Cai, Z., Sohrabpour, A., Jiang, H., Ye, S., Joseph, B., Brinkmann, B. H., Worrell, G. A., and He, B. (2021). Noninvasive high-frequency oscillations riding spikes delineates epileptogenic sources. *Proceedings of the National Academy of Sciences*, **118**(17):e2011130118.
[52] Calderon, D. P., Kilinc, M., Maritan, A., Banavar, J. R., and Pfaff, D. W. (2016). Generalized CNS arousal: An elementary force within the vertebrate nervous system. *Neuroscience & Biobehavioral Reviews*, **68**:167–176.
[53] Cameron, R. H. and Martin, W. T. (1944). On transformations of Wiener integrals under translations. *Annals of Mathematics*, **45**(2):386–396.
[54] Campo, A., Michałko, A., Kerrebroeck, B. V., Stajic, B., Pokric, M., and Leman, M. (2023). The assessment of presence and performance in an AR environment for motor imitation learning: A case-study on violinists. *Computers in Human Behavior*, **146**:107810.
[55] Candia-Rivera, D., Catrambone, V., Thayer, J. F., Gentili, C., and Valenza, G. (2022). Cardiac sympathetic-vagal activity initiates a functional brain–body response to emotional arousal. *Proceedings of the National Academy of Sciences*, **119**(21):e2119599119.
[56] Caponera, A. (2020). SPHARMA approximations for stationary functional time series on the sphere. *Statistical Inference for Stochastic Processes*, **24**:609–634.
[57] Cardoso, J.-F. (2022). Independent component analysis in the light of information geometry. *Information Geometry*, **7**:28–38.
[58] Chang, A., Livingstone, S. R., Bosnyak, D. J., and Trainor, L. J. (2017). Body sway reflects leadership in joint music performance. *Proceedings of the National Academy of Sciences*, **114**(21):E4134–E4141.
[59] Chen, H., Chen, D.-R., and Zhao, Y. (2021). Estimations of singular functions of kernel cross-covariance operators. *Journal of Approximation Theory*, **266**:105576.
[60] Cheung, Y.-M. and Xu, L. (1999). An empirical method to select dominant independent components in ICA for time series analysis. In *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No.99CH36339)*, volume 6, pages 3883–3887 vol.6.
[61] Christensen, W. F. and Zabriskie, B. N. (2021). When your permutation test is doomed to fail. *American Statistician*, **76**(1):53–63.
[62] Cirelli, L. K., Jurewicz, Z. B., and Trehub, S. E. (2019). Effects of maternal singing style on mother–infant arousal and behavior. *Journal of Cognitive Neuroscience*, **32**(7):1213–1220.
[63] Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, **51**(4):661–703.
[64] Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, **36**(3):287–314.
[65] Cuevas, A. and Fraiman, R. (2009). On depth measures and dual statistics. a methodology for dealing with general data. *Journal of Multivariate Analysis*, **100**(4):753–766.
[66] Cupidon, J., Gilliam, D. S., Eubank, R., and Ruymgaart, F. (2007). The delta method for analytic functions of random operators with application to functional data. *Bernoulli*, **13**(4):1179–1194.
[67] Currie, I. D. and Durban, M. (2002). Flexible smoothing with P-Splines: a unified approach. *Statistical Modelling*, **2**(4):333–349.
[68] Da Prato, G. and Zabczyk, J. (2014). *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press.
[69] Daffertshofer, A., Ton, R., Kringelbach, M. L., Woolrich, M. W., and Deco, G. (2018). Distinct criticality of phase and amplitude dynamics in the resting brain. *NeuroImage*, **180**:442–447.
[70] Dahl, M. J., Mather, M., and Werkle-Bergner, M. (2022). Noradrenergic modulation of rhythmic neural activity shapes selective attention. *Trends in Cognitive Sciences*, **26**(1):38–52.
[71] Daubechies, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics.
[72] Daubechies, I., Roussos, E., Takerkart, S., Benharrosh, M., Golden, C. J., D'ardenne, K., Richter, W., Cohen, J. D., and Haxby, J. V. (2009). Independent component analysis for brain fMRI does not select for independence. *Proceedings of*

the National Academy of Sciences, **106**(26):10415–10422.
[73] Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, **12**(1):136–154.
[74] de Gee, J. W., Colizoli, O., Kloosterman, N. A., Knapen, T. H. J., Nieuwenhuis, S., and Donner, T. H. (2017). Dynamic modulation of decision biases by brainstem arousal systems. *eLife*, **6**:e23232.
[75] Deco, G., Garcia, S. L., Perl, Y. S., Sporns, O., and Kringelbach, M. L. (2023). The effect of turbulence in brain dynamics information transfer measured with magnetoencephalography. *Communications Physics*, **6**(74):1–8.
[76] Deco, G. and Kringelbach, M. L. (2020). Turbulent-like dynamics in the human brain. *Cell Reports*, **33**(10):108471.
[77] Delaigle, A. and Hall, P. (2012a). Achieving near perfect classification for functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **74**(2):267–286.
[78] — (2012b). Methodology and theory for partial least squares applied to functional data. *Annals of Statistics*, **40**(1):322–352.
[79] Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, **34**(4):1443–1449.
[80] Demiral, S. B., Liu, C. K., Benveniste, H., Tomasi, D., and Volkow, N. D. (2023). Activation of brain arousal networks coincident with eye blinks during resting state. *Cerebral Cortex*, **33**(11):6792–6802.
[81] Demos, A. P., Chaffin, R., and Logan, T. (2018). Musicians body sway embodies musical structure and expression: A recurrence-based approach. *Musicae Scientiae*, **22**:244–263.
[82] Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**(3):425–455.
[83] — (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, **90**(432):1200–1224.
[84] Duzel, E. and Guitart-Masip, M. (2013). Not so uncertain at last: Locus coeruleus and decision making. *Neuron*, **79**(1):9–11.
[85] Eilers, P. and Marx, B. (2021). *Practical Smoothing: The Joys of P-splines.* Cambridge University Press.
[86] Engl, H., Hanke, M., and Neubauer, A. (2000). *Regularization of Inverse Problems.* Kluwer Academic Publishers, Boston.
[87] Epifanio, I. and Ventura-Campos, N. (2014). Hippocampal shape analysis in alzheimer's disease using functional data analysis. *Statistics in Medicine*, **33**(5):867–880.
[88] Escrichs, A., Perl, Y. S., Uribe, C., Camara, E., Türker, B., Pyatigorskaya, N., López-González, A., Pallavicini, C., Panda, R., Annen, J., Grosseries, O., Laureys, S., Naccache, L., Sitt, J. D., Laufs, H., Tagliazucchi, E., Kringelbach, M. L., and Deco, G. (2022). Unifying turbulent dynamics framework distinguishes different brain states. *Communications Bioliology*, **5**(638):638.
[89] Evarist, G. (1975). The addition formula for the eigenfunctions of the Laplacian. *Advances in Mathematics*, **18**(1):102–107.
[90] Everitt, B. J. and Robbins, T. W. (1997). Central cholinergic systems and cognition. *Annual Review of Psychology*, **48**:649–684.
[91] Fayomi, A., Pantazis, Y., Tsagris, M., and Wood, A. T. A. (2024). Cauchy robust principal component analysis with applications to high-dimensional data sets. *Statistics and Computing*, **34**(26):1–14.
[92] Ferrando, L., Ventura-Campos, N., and Epifanio, I. (2020). Detecting and visualizing differences in brain structures with SPHARM and functional data analysis. *NeuroImage*, **222**:117209.
[93] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis. Theory and Practice.* Springer-Verlag, New York.
[94] Filipowicz, A., Glaze, C. M., Kable, J. W., and Gold, J. I. (2020). Pupil diameter encodes the idiosyncratic, cognitive complexity of belief updating. *eLife*, **9**:e57872.
[95] Fink, L., Simola, J., Tavano, A., and et al. (2023). From pre-processing to advanced dynamic modeling of pupil data. *Behavior Research Methods*, **56**:1376–1412.
[96] Fitch, W. T. (2005). The evolution of music in comparative perspective. *Annals of the New York Academy of Sciences*, **1060**:29–49.
[97] Foxe, J. J. and Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology*, **2**:154.
[98] Friedman, J. H. (1987). Exploratory projection pursuit. *Journal of the American Statistical Association*, **82**(397):249–266.
[99] Friston, K. J. and Kiebel, S. J. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B*, **364**(1521):1211–1221.
[100] Frühholz, S., Trost, W., and Grandjean, D. (2014). The role of the medial temporal limbic system in processing emotions in voice and music. *Progress in Neurobiology*, **123**:1–17.
[101] Fujioka, T., Ross, B., and Trainor, L. J. (2015). Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. *Journal of Neuroscience*, **35**(45):15187–15198.
[102] Fukumizu, K., Bach, F. R., and Gretton, A. (2007). Statistical consistency of kernel canonical correlation analysis. *Journal of Machine Learning Research*, **8**(14):361–383.
[103] Gao, S., Proekt, A., Proekt, A., Renier, N., Calderon, D. P., Calderon, D. P., and Pfaff, D. W. (2019). Activating an anterior nucleus gigantocellularis subpopulation triggers emergence from pharmacologically-induced coma in rodents. *Nature Communications*, **10**:2897.
[104] Garey, J. D., Goodwillie, A. M., Frohlich, J., Morgan, M. A., Gustafsson, J.-Å., Smithies, O., Korach, K. S., Ogawa, S., and Pfaff, D. W. (2003). Genetic contributions to generalized arousal of brain and behavior. *Proceedings of the National Academy of Sciences*, **100**(19):11019–11022.
[105] Genzer, S., Ong, D. C., Zaki, J., and Perry, A. (2018). Mu rhythm suppression over sensorimotor regions is associated with greater empathic accuracy. *Social Cognitive and Affective Neuroscience*, **17**(19):788–801.
[106] Giordano, B. L., Whiting, C., Kriegeskorte, N., Kotz, S. A., Gross, J., and Belin, P. (2021). The representational dynamics of perceived voice emotions evolve from categories to dimensions. *Nature Human Behaviour*, **5**:1203–1213.
[107] Gong, R., Wegscheider, M., Mühlberg, C., Gast, R., Fricke, C., Rumpf, J.-J., Nikulin, V. V., Knösche, T. R., and Classen, J. (2020). Spatiotemporal features of $\beta - gamma$ phase-amplitude coupling in parkinson's disease derived from scalp eeg. *Brain*, **144**(2):487–503.
[108] Gordon, E. M., Chauvin, R. J., Van, A. N., Rajesh, A., Nielsen, A., Newbold, D. J., Lynch, C. J., Seider, N. A., Krimmel, S. R., Scheidter, K. M., Monk, J., Miller, R. L., Metoki, A., Montez, D. F., Zheng, A., Elbau, I., Madison, T., Nishino, T., Myers, M. J., Kaplan, S., Badke D'Andrea, C., Demeter, D. V., Feigelis, M., Ramirez, J. S. B., Xu, T., Barch, D. M., Smyser, C. D., Rogers, C. E., Zimmermann, J., Botteron, K. N., Pruett, J. R., Willie, J. T., Brunner, P., Shimony, J. S., Kay, B. P., Marek, S., Norris, S. A., Gratton, C., Sylvester, C. M., Power, J. D., Liston, C., Greene, D. J., Roland, J. L., Petersen, S. E., Raichle, M. E., Laumann, T. O., Fair, D. A., and Dosenbach, N. U. F. (2023). A somato-cognitive action network alternates with effector regions in motor cortex. *Nature*, **617**(7960):351–359.
[109] Gray, P. M., Krause, B. L., Atema, J., Payne, R. L., Krumhansl, C., and Baptista, L. F. (2001). The music of nature and the nature of music. *Science*, **291**(5501):52–54.

[110] Grenander, U. (1950). Stochastic processes and statistical inference. *Arkiv för Matematik*, **1**:195–277.
[111] Gutch, H. W. and Theis, F. J. (2012). To Infinity and Beyond: On ICA over Hilbert spaces. In Theis, F., Cichocki, A., Yeredor, A., and Zibulevsky, M., editors, *Latent Variable Analysis and Signal Separation*, pages 180–187.
[112] Haegens, S., Nácher, V., Luna, R., Romo, R., and Jensen, O. (2011). α-oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proceedings of the National Academy of Sciences*, **108**(48):19377–19382.
[113] Halgren, M., Ulbert, I., Bastuji, H., Fabó, D., Erőss, L., Rey, M., Devinsky, O., Doyle, W. K., Mak-McCully, R., Halgren, E., Wittner, L., Chauvel, P., Heit, G., Eskandar, E., Mandell, A., and Cash, S. S. (2019). The generation and propagation of the human alpha rhythm. *Proceedings of the National Academy of Sciences*, **116**(47):23772–23782.
[114] Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **68**(1):109–126.
[115] Harris, K. D. and Thiele, A. (2011). Cortical state and attention. *Nature Reviews Neuroscience*, **12**:509–523.
[116] Haufe, S., Dähne, S., and Nikulin, V. V. (2014). Dimensionality reduction for the analysis of brain oscillations. *NeuroImage*, **101**:583–597.
[117] Hayashi, N., Someya, N., and Fukuba, Y. (2010). Effect of intensity of dynamic exercise on pupil diameter in humans. *Journal of Physiological Anthropology*, **29**(3):119–122.
[118] Herman, L. M. (2017). The multiple functions of male song within the humpback whale (Megaptera novaeangliae) mating system: review, evaluation, and synthesis. *Biological Reviews*, **92**(3):1795–1818.
[119] Hofmann, S. M., Klotzsche, F., Mariola, A., Nikulin, V. V., Villringer, A., and Gaebler, M. (2020). Decoding subjective emotional arousal from EEG during an immersive virtual reality experience. *eLife*, **10**:e64812.
[120] Holstege, G., Bandler, R., and Saper, C. (1996). *The Emotional Motor System*. Progress in brain research. Elsevier.
[121] Holstege, G. and Subramanian, H. H. (2015). Two different motor systems are needed to generate human speech. *Journal of Comparative Neurology*, **524**(8):1558–1577.
[122] — (2016). Two different motor systems are needed to generate human speech. *Journal of Comparative Neurology*, **524**(8):1558–1577.
[123] Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer, New York.
[124] Hosseini-Nasab, S. M. E. (2014). A modification of silverman's method for smoothed functional principal components analysis. *Journal of Statistical Theory and Practice*, **8**(2):400–413.
[125] Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley, Chichester.
[126] Hudson, A. E., Calderon, D. P., Pfaff, D. W., and Proekt, A. (2014). Recovery of consciousness is mediated by a network of discrete metastable activity states. *Proceedings of the National Academy of Sciences*, **111**(25):9283–9288.
[127] Hyvärinen , A. and Oja, E. (1997). A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Computation*, **9**(7):1483–1492.
[128] Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. John Wiley & Sons, Ltd, New York.
[129] Ibragimov, I. A. (2014). On the ghurye–olkin–zinger theorem. *Journal of Mathematical Sciences*, **199**(2):174–183.
[130] Ilmonen, P., Nordhausen, K., Oja, H., and Ollila, E. (2012). On asymptotics of ICA estimators and their performance indices. *arXiv*, (1212.3953).
[131] Ilmonen, P. and Paindaveine, D. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *Annals of Statistics*, **39**(5):2448–2476.
[132] Itoh, K., Konoike, N., Nejime, M., Iwaoki, H., Igarashi, H., Hirata, S., and Nakamura, K. (2022). Cerebral cortical processing time is elongated in human brain evolution. *Scientific Reports*, **12**(1):1103.
[133] Iwasaki, M., Poulsen, T. M., Oka, K., and Hessler, N. A. (2013). Singing-related activity in anterior forebrain of male zebra finches reflects courtship motivation for target females. *PLoS One*, **8**(11):e81725.
[134] Jacques, J. and Preda, C. (2014). Model-based clustering for multivariate functional data. *Computational Statistics & Data Analysis*, **71**:92–106.
[135] Jaffe, P. I. and Brainard, M. S. (2020). Acetylcholine acts on songbird premotor circuitry to invigorate vocal output. *eLife*, **9**:e53288.
[136] Janata, P., Birk, J. L., Horn, J. D. V., Leman, M., Tillmann, B., and Bharucha, J. J. (2002). The Cortical Topography of Tonal Structures Underlying Western Music. *Science*, **298**(5601):2167–2170.
[137] Jonmohamadi, Y. and Muthukumaraswamy, S. D. (2017). Multi-band component analysis for EEG artifact removal and source reconstruction with application to gamma-band activity. *Biomedical Physics & Engineering Express*, **4**:035007.
[138] Joshi, S. and Gold, J. I. (2020). Pupil size as a window on neural substrates of cognition. *Trends in Cognitive Sciences*, **24**(6):466–480.
[139] Joshi, S., Li, Y., Kalwani, R. M., and Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, **89**(1):221–234.
[140] Jover, M., Cellier, M., and Scola, C. (2019). Infants' motor activity during a mother–infant interaction alternating silent and singing phases. *Journal of Motor Learning and Development*, **7**(3):426–436.
[141] Jung, T.-P., Humphries, C., Lee, T.-W., Makeig, S., McKeown, M., Iragui, V., and Sejnowski, T. (1998). Removing electroencephalographic artifacts: comparison between ICA and PCA. In *Neural Networks for Signal Processing VIII. Proceedings of the 1998 IEEE Signal Processing Society Workshop (Cat. No.98TH8378)*, pages 63–72.
[142] Jürgens, U. (2009). The neural control of vocalization in mammals: a review. *Journal of voice*, **23**(1):1–10.
[143] Kaiser, J. F. (1990). On a simple algorithm to calculate the 'energy' of a signal. *International Conference on Acoustics, Speech, and Signal Processing*, **1**:381–384.
[144] Karhunen, K. (1946). Zur spektraltheorie stochastiche prozesse. *Annales Academiae Scientiarum Fennicae. Ser. A. I. Math.-Phys.*, **34**.
[145] — (1947). Uber lineare methoden in der wahrscheinlichkeitsrechnung. *Annales Academiae Scientarum Fennicae. Ser. A. I. Math.-Phys.*, **37**:1–79.
[146] Kawashima, T., Shibusawa, S., and Amano, K. (2022). Frequency- and phase-dependent effects of auditory entrainment on attentional blink. *European Journal of Neuroscience*, **56**(4):4411–4424.
[147] Kessy, A., Lewin, A., and Strimmer, K. (2018). Optimal whitening and decorrelation. *American Statistician*, **72**(4):309–314.
[148] Kilinc, M., Calderon, D. P., Tabansky, I., Martin, E. M., and Pfaff, D. W. (2023). *Elementary Central Nervous System Arousal*, pages 1–33. Springer New York, New York, NY.
[149] Kim, J.-H., Kim, D.-W., and Im, C.-H. (2017). Brain areas responsible for vigilance: An EEG source imaging study. *Brain*

Topography, **30**(3):343–351.
[150] Kleber, B., Veit, R., Birbaumer, N., Gruzelier, J., and Lotze, M. (2010). The brain of opera singers: experience-dependent changes in functional activation. *Cerebral cortex*, **20**(5):1144–1152.
[151] Kleber, B., Zeitouni, A., Friberg, A., and Zatorre, R. J. (2013). Experience-dependent modulation of feedback integration during singing: Role of the right anterior insula. *The Journal of Neuroscience*, **33**(14):6070–6080.
[152] Knapen, T., de Gee, J. W., Brascamp, J., Nuiten, S., Hoppenbrouwers, S., and Theeuwes, J. (2016). Cognitive and ocular factors jointly determine pupil responses under equiluminance. *PLoS One*, **11**(5):1–13.
[153] Knapen, T. and Gee, J. W. D. (2016). Firdeconvolution.
[154] Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience*, **15**:170–180.
[155] Koelsch, S., Fritz, T., v. Cramon, D. Y., Müller, K., and Friederici, A. D. (2006a). Investigating emotion with music: An fMRI study. *Human Brain Mapping*, **27**.
[156] Koelsch, S., Fritz, T., v. Cramon, D. Y., Müller, K., and Friederici, A. D. (2006b). Investigating emotion with music: An fmri study. *Human Brain Mapping*, **27**(3):239–250.
[157] Koelsch, S., Vuust, P., and Friston, K. J. (2019). Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences*, **23**(1):63–77.
[158] Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. CRC Press.
[159] Koner, S. and Staicu, A.-M. (2023). Second-generation functional data. *Annual Review of Statistics and Its Application*, **10**(1):547–572.
[160] Koo, T. K. and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, **15**(2):155–163.
[161] Kosambi, D. D. (1943). Statistics in function space. *The Journal of the Indian Mathematical Society*, **7**:76–88.
[162] Krishnamurthy, K., Nassar, M. R., Sarode, S., and Gold, J. I. (2017). Arousal-related adjustments of perceptual biases optimize perception in dynamic environments. *Nature Human Behaviour*, **1**:0107.
[163] Kupresanin, A., Shin, H., King, D., and Eubank, R. (2010). An RKHS framework for functional data analysis. *Journal of Statistical Planning and Inference*, **140**(12):3627–3637.
[164] Kuwamizu, R., Yamazaki, Y., Aoike, N., Ochi, G., Suwabe, K., and Soya, H. (2022). Pupil-linked arousal with very light exercise: pattern of pupil dilation during graded exercise. *The Journal of Physiological Sciences*, **72**(23):1–9.
[165] Laha, R. and Rohatgi, V. (1979). *Probability Theory*. Dover, New York.
[166] Lahav, A., Saltzman, E., and Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *The Journal of Neuroscience*, **27**(2):308–314.
[167] Lakraj, G. P. and Ruymgaart, F. (2017). Some Asymptotic Theory for Silverman's Smoothed Functional Principal Components in an Abstract Hilbert Space. *Journal of Multivariate Analysis*, **155**:122–132.
[168] Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception*, **17**(4):481–509.
[169] — (2008). *Embodied music cognition and mediation technology*. MIT Press.
[170] — (2016). *The expressive moment: How interaction (with music) shapes human empowerment*. MIT Press.
[171] Leman, M. and Maes, P.-J. (2015). The role of embodiment in the perception of music. *Empirical Musicology Review*, **9**(3-4):236–246.
[172] Leman, Marc (2021). Co-regulated timing in music ensembles: a Bayesian listener perspective. *Journal of New Music Research*, **50**(2):121–132.
[173] Lense, M. D., Shultz, S., Astésano, C., and Jones, W. (2022). Music of infant-directed singing entrains infants' social visual behavior. *Proceedings of the National Academy of Sciences*, **119**(45):e2116967119.
[174] Leobacher, G. and Prochno, J. (2021). Statistical independence in mathematics–the key to a Gaussian law. *Mathematische Semesterberichte*, **68**:69–104.
[175] Lévêque, Y. and Schön, D. (2015). Modulation of the motor cortex during singing-voice perception. *Neuropsychologia*, **70**:58–63.
[176] Li, B., Bever, G. V., Oja, H., Sabolová, R., and Critchley, F. (2019). Functional independent component analysis: an extension of the fourth-orderblind identification. Technical report, Université de Namur: Namur, Belgium.
[177] Liebe, T., Kaufmann, J., Hämmerer, D., Betts, M. J., and Walter, M. (2022). In vivo tractography of human locus coeruleus—relation to 7T resting state fMRI, psychological measures and single subject validity. *Molecular Psychiatry*, **27**:4984–4993.
[178] Liebisch, A. P., Eggert, T., Shindy, A., Valentini, E., Irving, S., Stankewitz, A., and Schulz, E. (2020). A novel tool for the removal of muscle artefacts from EEG: Improving data quality in the gamma frequency range. *Journal of Neuroscience Methods*, **358**:109217.
[179] Lifshits, M. (2012). *Lectures on Gaussian Processes*. Springer Berlin, Heidelberg.
[180] Ling, N. and Vieu, P. (2018). Nonparametric modelling for functional data: selected survey and tracks for future. *Statistics*, **52**(4):934–949.
[181] Liu, X., Pfaff, D. W., Calderon, D. P., Tabansky, I., Wang, X., Wang, Y., and ming Kow, L. (2016). Development of electrophysiological properties of nucleus gigantocellularis neurons correlated with increased CNS arousal. *Developmental Neuroscience*, **38**(4):295–310.
[182] Liu, Y., Liang, X. S., and Weisberg, R. H. (2007). Rectification of the bias in the wavelet power spectrum. *Journal of Atmospheric and Oceanic Technology*, **24**:2093–2102.
[183] Lohani, S., Moberly, A. H., Benisty, H., Landa, B., Jing, M., Li, Y., Higley, M. J., and Cardin, J. A. (2022). Spatiotemporally heterogeneous coordination of cholinergic and neocortical activity. *Nature Neuroscience*, **25**(12):1706–1713.
[184] Loève, M. (1948). *Fonctions aléatoires du second ordre*.
[185] Mahjoory, K., Nikulin, V. V., Botrel, L., Linkenkaer-Hansen, K., Fato, M. M., and Haufe, S. (2017). Consistency of EEG source localization and connectivity estimates. *NeuroImage*, **152**:590–601.
[186] Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**:674–693.
[187] Marron, J. S. and Dryden, I. L. (2022). *Object Oriented Data Analysis*. Chapman and Hall/CRC.
[188] Martin, E. M., Devidze, N., Shelley, D. N., Westberg, L., Fontaine, C., and Pfaff, D. W. (2011). Molecular and neuroanatomical characterization of single neurons in the mouse medullary gigantocellular reticular nucleus. *Journal of Comparative Neurology*, **519**(13):2574–2593.
[189] Martin, E. M. and Pfaff, D. W. (2013). *Elementary CNS Arousal*. Springer New York, New York, NY.
[190] Martin, J. T., Whittaker, A. H., and Johnston, S. J. (2020). Component processes in free-viewing visual search: Insights

from fixation-aligned pupillary response averaging. *Journal of Vision*, **20**(7):5.
[191] Mas, A. (2006). Weak convergence in the functional autoregressive model. *Journal of Multivariate Analysis*, **98**:1231–1261.
[192] McDougal, D. H. and Gamlin, P. D. (2015). Autonomic control of the eye. *Comprehensive Physiology*, **5**(1):439–73.
[193] McGaugh, J. D. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience*, **27**:1–28.
[194] McGinley, M. J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C. R., Tolias, A. S., Cardin, J. A., and McCormick, D. A. (2015). Waking state: Rapid variations modulate neural and behavioral responses. *Neuron*, **87**(6):1143–1161.
[195] McGowan, A. L., Chandler, M. C., Brascamp, J. W., and Pontifex, M. B. (2019). Pupillometric indices of locus-coeruleus activation are not modulated following single bouts of exercise. *International Journal of Psychophysiology*, **140**:41–52.
[196] Megemont, M., Mcburney-lin, J., and Yang, H. (2022). Pupil diameter is not an accurate real-time readout of locus coeruleus activity. *eLife*, **11**:e70510.
[197] Mehta, N. and Gray, A. (2009). Funcica for time series pattern discovery. In *Proceedings of the 2009 SIAM International Conference on Data Mining*, pages 73–84, Sparks, NV, USA.
[198] Mesters, G. and Zwiernik, P. (2023). Non-independent components analysis. *arXiv*, (2206.13668).
[199] Michałko, A., Di Stefano, N., Campo, A., and Leman, M. (2024). Enhancing human-human musical interaction through kinesthetic haptic feedback using wearable exoskeletons: theoretical foundations, validation scenarios, and limitations. *Frontiers in Psychology*, **15**:1327992.
[200] Montefusco-Siegmund, R., Schwalm, M., Jubal, E. R., Devia, C., Egaña, J. I., and Maldonado, P. E. (2022). Alpha EEG activity and pupil diameter coupling during inactive wakefulness in humans. *eNeuro*, **9**(2):ENEURO.0060–21.2022.
[201] Móri, T., Rohatgi, V. K., and Székely, G. J. (1994). On Multivariate Skewness and Kurtosis. *Theory of Probability and its Applications*, **38**(3):547–551.
[202] Moura, N., Fonseca, P., Goethel, M., Oliveira-Silva, P., Vilas-Boas, J. P., and Serra, S. (2023a). The impact of visual display of human motion on observers' perception of music performance. *PLoS One*, **18**(3):1–15.
[203] Moura, N. and Serra, S. (2024). Saxophone Players' Self-Perceptions About Body Movement in Music Performing and Learning: An Interview Study. *Music Perception*, **41**(3):199–216.
[204] Moura, N., Vidal, M., Aguilera, A., Vilas-Boas, J., Serra, S., and Leman, M. (2023b). Knee flexion of saxophone players anticipates tonal context of music. *npj Science of Learning*, **8**(22).
[205] Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., and Balsters, J. H. (2014). Pupil diameter covaries with bold activity in human locus coeruleus. *Human Brain Mapping*, **35**(8):4140–4154.
[206] Muthukumaraswamy, S. D. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. *Frontiers in Human Neuroscience*, **7**:138.
[207] Naber, M., Alvarez, G. A., and Nakayama, K. (2013). Tracking the allocation of attention using human pupillary oscillations. *Frontiers in Psychology*, **4**:919.
[208] Naber, M., Hommel, B., and Colzato, L. S. (2015). Improved human visuomotor performance and pupil constriction after choline supplementation in a placebo-controlled double-blind study. *Scientific Reports*, **5**:13188.
[209] Naber, M. and Murphy, P. R. (2019). Pupillometric investigation into the speed-accuracy trade-off in a visuo-motor aiming task. *Psychophysiology*, **57**(3):e13499.
[210] Nair, A., Teo, Y. Y., Augustine, G. J., and Graf, M. (2023). A functional logic for neurotransmitter corelease in the cholinergic forebrain pathway. *Proceedings of the National Academy of Sciences*, **120**(28):e2218830120.
[211] Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B. S., and Gold, J. I. (2012). Rational regulation of learning dynamics by pupil–linked arousal systems. *Nature Neuroscience*, **15**:1040–1046.
[212] Navarro-García, M., Guerrero, V., and Durban, M. (2023). On constrained smoothing and out-of-range prediction using P-splines: A conic optimization approach. *Applied Mathematics and Computation*, **441**:127679.
[213] Nelson, A. and Mooney, R. (2016). The basal forebrain and motor cortex provide convergent yet distinct movement-related inputs to the auditory cortex. *Neuron*, **90**(3):635–648.
[214] Nikulin, V. V., Nolte, G., and Curio, G. (2011). A novel method for reliable and fast extraction of neuronal EEG/MEG oscillations on the basis of spatio-spectral decomposition. *NeuroImage*, **55**:1528–1535.
[215] — (2012). Cross-frequency decomposition: A novel technique for studying interactions between neuronal oscillations with different frequencies. *Clinical Neurophysiology*, **123**(7):1353–1360.
[216] Noei, S., Zouridis, I. S., Logothetis, N. K., Panzeri, S., and Totah, N. K. (2022). Distinct ensembles in the noradrenergic locus coeruleus are associated with diverse cortical states. *Proceedings of the National Academy of Sciences*, **119**(18):e2116507119.
[217] Nordhausen, K. and Oja, H. (2018). Independent component analysis: A statistical perspective. *WIREs Computational Statistics*, **10**(5):e1440.
[218] Nordhausen, K. and Virta, J. (2019). An overview of properties and extensions of FOBI. *Knowledge-Based Systems*, **173**:113–116.
[219] Nordström, K. (1999). The life and work of Gustav Elfving. *Statistical Science*, **14**(2):174–196.
[220] Norman-Haignere, S. V., Feather, J., Boebinger, D., Brunner, P., Ritaccio, A. L., McDermott, J. H., Schalk, G., and Kanwisher, N. G. (2022). A neural population selective for song in human auditory cortex. *Current Biology*, **32**(6):1470–1484.e12.
[221] Nusseck, M. and Wanderley, M. M. (2009). Music and motion: How music-related ancillary body movements contribute to the experience of music. *Music Perception*, **26**(4):335–353.
[222] Ocaña, F. A., Aguilera, A. M., and Escabias, M. (2007). Computational considerations in functional principal component analysis. *Computational Statistics*, **22**:449–465.
[223] Ocaña, F. A., Aguilera, A. M., and Valderrama, M. J. (1999). Functional principal component analysis by choice of norm. *Journal of Multivariate Analysis*, **71**(2):262–276.
[224] Ocaña, F. A. (1995). *Alternativas geometricas en el ACP de una V.A. Hilbertiana*. PhD thesis, University of Granada.
[225] Oja, E., Kiviluoto, K., and Malaroiu, S. (2000). Independent component analysis for financial time series. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, pages 111–116.
[226] Ojeda, A., Kreutz-Delgado, K., and Mishra, J. (2021). Bridging M/EEG source imaging and independent component analysis frameworks using biologically inspired sparsity priors. *Neural Computation*, **33**(9):2408–2438.
[227] Olszewska, A. M., Gaca, M., Droź dziel, D., Widlarz, A., Herman, A. M., and Marchewka, A. (2024). Understanding functional brain reorganization for naturalistic piano playing in novice pianists. *Journal of Neuroscience Research*, **102**(2):e25312.
[228] Ovalle–Muñoz, D. P. and Ruiz–Medina, M. D. (2024). LRD spectral analysis of multifractional functional time series on manifolds. *Test*.
[229] Owren, M. J., Amoss, R. T., and Rendall, D. (2011). Two organizing principles of vocal production: Implications for

nonhuman and human primates. *American Journal of Primatology*, **73**(6):530–544.
[230] Pachitariu, M., Lyamzin, D. R., Sahani, M., and Lesica, N. A. (2015). State-dependent population coding in primary auditory cortex. *The Journal of Neuroscience*, **35**(5):2058–2073.
[231] Peña, D., Prieto, F. J., and Rendón, C. (2014). Independent components techniques based on kurtosis for functional data analysis. Working paper 14–10, Universidad Carlos III de Madrid.
[232] Percival, D. B. and Mofjeld, H. O. (1997). Analysis of subtidal coastal sea level fluctuations using wavelets. *Journal of the American Statistical Association*, **92**(439):868–880.
[233] Peysakhovich, V., Causse, M., Scannella, S., and Dehais, F. (2015). Frequency analysis of a task-evoked pupillary response: Luminance-independent measure of mental effort. *International Journal of Psychophysiology*, **97**(1):30–37.
[234] Peña, D., Prieto, F. J., and Viladomat, J. (2010). Eigenvectors of a kurtosis matrix as interesting directions to reveal cluster structure. *Journal of Multivariate Analysis*, **101**(9):1995–2007.
[235] Pfaff, D. W. (2009). *Brain Arousal and Information Theory: Neural and Genetic Mechanisms*. Harvard University Press.
[236] Pfaff, D. W., Martin, E. M., and Faber, D. S. (2012). Origins of arousal: roles for medullary reticular neurons. *Trends in Neurosciences*, **35**(8):468–476.
[237] Pfaff, D. W., Ribeiro, A. C., Matthews, J., and ming Kow, L. (2008). Concepts and mechanisms of generalized central nervous system arousal. *Annals of the New York Academy of Sciences*, **1129**:11–25.
[238] Picciotto, M. R., Higley, M. J., and Mineur, Y. S. (2012). Acetylcholine as a neuromodulator: Cholinergic signaling shapes nervous system function and behavior. *Neuron*, **76**(1):116–129.
[239] Pinto, L., Goard, M. J., Estandian, D., Xu, M., Kwan, A., Lee, S.-H., Harrison, T. C., Feng, G., and Dan, Y. (2013). Fast modulation of visual perception by basal forebrain cholinergic neurons. *Nature Neuroscience*, **16**:1857–1863.
[240] Pozeg, P., Palluel, E., Ronchi, R., Solcà, M., Al-Khodairy, A., Jordan, X., Kassouha, A., and Blanke, O. (2017). Virtual reality improves embodiment and neuropathic pain caused by spinal cord injury. *Neurology*, **89**(18):1894–1903.
[241] Proekt, A., Banavar, J. R., Maritan, A., and Pfaff, D. W. (2012). Scale invariance in the dynamics of spontaneous behavior. *Proceedings of the National Academy of Sciences*, **109**(26):10564–10569.
[242] Qi, X. and Zhao, H. (2011). Some Theoretical Properties of Silverman's Method for Smoothed Functional Principal Component Analysis. *Journal of Multivariate Analysis*, **102**:742–767.
[243] Quinkert, A. W., Vimal, V. P., Weil, Z. M., Reeke, G. N., Schiff, N. D., Banavar, J. R., and Pfaff, D. W. (2011). Quantitative descriptions of generalized arousal, an elementary function of the vertebrate brain. *Proceedings of the National Academy of Sciences*, **108**(supp.3):15617–15623.
[244] R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
[245] Ramsay, J. O. (1982). When the data are functions. *Psychometrika*, **47**:379–396.
[246] Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, **53**(3):539–572.
[247] Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. Springer-Verlag, New York.
[248] — (2005). *Functional Data Analysis*. Springer, New York.
[249] Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, **14**(1):1–17.
[250] Reimer, J., Froudarakis, E., Cadwell, C. R., Yatsenko, D., Denfield, G. H., and Tolias, A. S. (2014). Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron*, **84**(2):355–362.
[251] Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., and Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, **7**:13289.
[252] Reyhani, N., Ylipaavalniemi, J., Vigário, R., and Oja, E. (2012). Consistency and asymptotic normality of FastICA and bootstrap FastICA. *Signal Processing*, **92**(8):1767–1778.
[253] Ribeiro, A. B., Gabriel, R., Garcia, B., Cuccio, C., Aqeel, W., Moreno, A., Landeen, C., Hurley, A., Kavey, N., and Pfaff, D. D. (2022). Temporal relations between peripheral and central arousals in good and poor sleepers. *Proceedings of the National Academy of Sciences*, **119**(25):e2201143119.
[254] Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, **53**(1):233–243.
[255] Rosa, M. D., Sangalli, L. M., and Vantini, S. (2014). Principal differential analysis of the aneurisk65 data set. *Advances in Data Analysis and Classification*, **8**:287–302.
[256] Rosas, F. E., Mediano, P. A. M., Luppi, A. I., Varley, T. F., Lizier, J. T., Stramaglia, S., Jensen, H. J., and Marinazzo, D. (2022). Disentangling high-order mechanisms and high-order behaviours in complex systems. *Nature Physics*, **18**:476–477.
[257] Ross, J. M., Comstock, D. C., Iversen, J. R., Makeig, S., and Balasubramaniam, R. (2021). Cortical mu rhythms during action and passive music listening. *Journal of neurophysiology*, **127**(1):213–224.
[258] Rousseeuw, P. J. and van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, **41**(3):212–223.
[259] Roy, S., Banerjee, A., Roy, C., Nag, S., Sanyal, S., Sengupta, R., and Ghosh, D. (2021). Brain response to color stimuli: an EEG study with nonlinear approach. *Cognitive Neurodynamics*, **15**(6):1023–1053.
[260] Ruiz-Medina, M. (2016). Functional analysis of variance for Hilbert-valued multivariate fixed effect models. *Statistics*, **50**(3):689–715.
[261] Ruiz-Medina, M. D., Miranda, D., and Espejo, R. M. (2019). Dynamical multiple regression in function spaces, under kernel regressors, with ARH(1) errors. *TEST*, **28**(3):943–968.
[262] Sakata, S. (2016). State-dependent and cell type-specific temporal processing in auditory thalamocortical circuit. *Scientific Reports*, **6**:18873.
[263] Saporta, G. (1985). Data analysis for numerical and categorical individual time-series. *Applied Stochastic Models and Data Analysis*, **1**(2):109–119.
[264] Sara, S. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, **10**:211–223.
[265] Sasaki, A., Sotnikova, T. D., Gainetdinov, R. R., and Jarvis, E. D. (2006). Social context-dependent singing-regulated dopamine. *The Journal of Neuroscience*, **26**(35):9010–9014.
[266] Sasaki, H., Gutmann, M., Shouno, H., and Hyvarinen, A. (2014). Estimating Dependency Structures for non-Gaussian Components with Linear and Energy Correlations. In Kaski, S. and Corander, J., editors, *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, volume 33 of *Proceedings of Machine Learning Research*, pages 868–876, Reykjavik, Iceland. PMLR.
[267] Schaworonkow, N. and Nikulin, V. V. (2018). Spatial neuronal synchronization and the waveform of oscillations: Implica-

tions for EEG and MEG. *PLoS Computational Biology*, **15**(5):e1007055.
[268] Scherer, K. R., Sundberg, J., Fantini, B., Trznadel, S., and Eyben, F. (2017). The expression of emotion in the singing voice: Acoustic patterns in vocal performance. *The Journal of the Acoustical Society of America*, **142**(4):1805.
[269] Schmidt, G., Malik, M., Barthel, P., Schneider, R., Ulm, K., Rolnitzky, L., Camm, A. J., Bigger, J. T., and Schömig, A. (1999). Heart-rate turbulence after ventricular premature beats as a predictor of mortality after acute myocardial infarction. *The Lancet*, **353**(9162):1390–1396.
[270] Schwabe, L., Merz, C. J., Walter, B., Vaitl, D., Wolf, O. T., and Stark, R. (2011). Emotional modulation of the attentional blink: The neural structures involved in capturing and holding attention. *Neuropsychologia*, **49**(3):416–425.
[271] Schwartz, J. W., Sanchez, M. M., and Gouzoules, H. (2022). Vocal expression of emotional arousal across two call types in young rhesus macaques. *Animal Behaviour*, **190**:125–138.
[272] Shang, H. L. and Hyndman, R. J. (2018). *fds: Functional Data Sets*. R package version 1.8.
[273] Shen, H., Jegelka, S., and Gretton, A. (2009). Fast kernel-based independent component analysis. *IEEE Transactions on Signal Processing*, **57**(9):3498–3511.
[274] Shigeta, T. T., Morris, T. P., D.L. Hartwell, C. H., Kucyi, A., Bex, P. J., Kramer, A. F., and Hillman, C. H. (2021). Acute exercise effects on inhibitory control and the pupillary response in young adults. *International Journal of Psychophysiology*, **170**:218–228.
[275] Shin, H. (2008). An extension of Fisher's discriminant analysis for stochastic processes. *Journal of Multivariate Analysis*, **99**(6):1191–1216.
[276] Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Annals of Statistics*, **24**(1):1–24.
[277] Siman-Tov, T., Gordon, C. R., Avisdris, N., Shany, O., Lerner, A., Shuster, O., Granot, R. Y., and Hendler, T. (2022). The rediscovered motor-related area 55b emerges as a core hub of music perception. *Communications Biology*, **5**(1):1104.
[278] Siqi-Liu, A., Harris, A., Atkinson, A. P., and Reed, C. L. (2018). Dissociable processing of emotional and neutral body movements revealed by μ-alpha and beta rhythms. *Social Cognitive and Affective Neuroscience*, **13**(12):1269–1279.
[279] Skitovich, V. (1962). Linear combinations of independent random variables and the normal distribution. *Selected Translations in Mathematical Statistics and Probability*, **2**:211–229.
[280] Spiech, C., Danielsen, A., Laeng, B., and Endestad, T. (2024). Oscillatory attention in groove. *Cortex*, **174**:137–148.
[281] Spiech, C., Sioros, G., Endestad, T., et al. (2022). Pupil drift rate indexes groove ratings. *Scientific Reports*, **12**(1):11620.
[282] Staib, M. and Frühholz, S. (2021). Cortical voice processing is grounded in elementary sound analyses for vocalization relevant sound patterns. *Progress in Neurobiology*, **200**:101982.
[283] Starns, J. J., Cataldo, A. M., and Rotello, C. M. (2019). Blinded inference: an opportunity for mathematical modelers to lead the way in research reform. *Computational Brain & Behavior*, **2**:223–228.
[284] Stepniak, C. (2007). Bernstein's examples on independent events. *The College Mathematics Journal*, **38**(2):140–142.
[285] Stitt, I., Zhou, Z. C., Radtke-Schuller, S., and Fröhlich, F. (2018). Arousal dependent modulation of thalamo-cortical functional interaction. *Nature Communications*, **9**(1):2455.
[286] Sweeney, K. T., McLoone, S. F., and Ward, T. E. (2013). The use of ensemble empirical mode decomposition with canonical correlation analysis as a novel artifact removal technique. *IEEE Transactions on Biomedical Engineering*, **60**(1):97–105.
[287] Tabansky, I., Liang, Y., Frankfurt, M., Daniels, M. A., Harrigan, M., Stern, S. A., Milner, T. A., Leshan, R. L., Rama, R., Moll, T., Friedman, J. M., Stern, J. N., and Pfaff, D. W. (2018). Molecular profiling of reticular gigantocellularis neurons indicates that eNOS modulates environmentally dependent levels of arousal. *Proceedings of the National Academy of Sciences*, **115**(29):E6900–E6909.
[288] Theis, F. (2004a). Uniqueness of complex and multidimensional independent component analysis. *Signal Processing*, **84**(5):951–956.
[289] Theis, F. J. (2004b). A new concept for separability problems in blind source separation. *Neural Computation*, **16**(9):1827–1850.
[290] Torrence, C. and Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, **79**(1):61–78.
[291] Valderrama, M. and Aguilera, A. (1997). A normality criterion for random vectors based on independence. *Statistics & Probability Letters*, **33**(2):159–165.
[292] Valderrama, M. J., Ocaña, F. A., Aguilera, A. M., and Ocaña-Peinado, F. M. (2010). Forecasting pollen concentration by a two-step functional model. *Biometrics*, **66**(2):578–585.
[293] Van Noorden, Leon and De Bruyn, Leen and Van Noorden, Raven and Leman, Marc (2017). Embodied social synchronization in children's musical development. In Lesaffre, Micheline and Maes, Pieter-Jan and Leman, Marc, editor, *The Routledge companion to embodied music interaction*, Routledge Music Companions, pages 195–204. Routledge.
[294] Vazey, E. M., Moorman, D. E., and Aston-Jones, G. (2018). Phasic locus coeruleus activity regulates cortical encoding of salience information. *Proceedings of the National Academy of Sciences*, **115**(40):E9439–E9448.
[295] Venkatraman, A., Edlow, B. L., and Immordino-Yang, M. H. (2017). The brainstem in emotion: A review. *Frontiers in Neuroanatomy*, **11**:15.
[296] Vidal, M. and Aguilera, A. M. (2022). Novel whitening approaches in functional settings. *Stat*, **12**(1):e516.
[297] — (2023a). *pfica: Independent Components Analysis Techniques for Functional Data*. R package version 0.1.3.
[298] — (2023b). Wavelet-based sparse optimization via fixed-point iteration scheme in high-dimensional data analysis. In Carillo, S., Conti, C., Mansutti, D., Pitolli, F., and Spitaleri, R. M., editors, *Proceedings of the 21st International Association for Mathematics and Computers in Simulation World Congress*, page 304. IMACS.
[299] — (2024). Wavelet thresholding on independent subspace factorizations of spatially indexed wide functional data for robust estimation of cortical activity. *Under review*.
[300] Vidal, M., Leman, M., and Aguilera, A. M. (2024a). Functional independent component analysis by choice of norm: a framework for near-perfect classification. *Under review*.
[301] Vidal, M., Moura, N., Aguilera, A. M., Fritz, T. H., and Leman, M. (2024b). Geometric-based turbulence analysis of EEG signals for modeling emotional arousal during active immersive virtual interactions. *Under review*.
[302] Vidal, M., Onderdijk, E., K., Aguilera, A. M., Six, J., Maes, P.-J., Fritz, T. H., and Leman, M. (2023). Cholinergic-related pupil activity reflects level of emotionality during motor performance. *European Journal of Neuroscience*, **n/a**:1–14.
[303] Vidal, M., Rosso, M., and Aguilera, A. M. (2021). Bi-smoothed functional independent component analysis for EEG artifact removal. *Mathematics*, **9**(11):1243.
[304] Vinck, M. A., Batista-Brito, R. A., Knoblich, U., and Cardin, J. A. (2015). Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. *Neuron*, **86**(3):740–754.
[305] Virta, J., Li, B., Nordhausen, K., and Oja, H. (2020). Independent component analysis for multivariate functional data.

Journal of Multivariate Analysis, **176**:104568.
[306] Virta, J. and Nordhausen, K. (2017). On the optimal non-linearities for Gaussian mixtures in FastICA. In Tichavský, P., Babaie-Zadeh, M., Michel, O. J., and Thirion-Moreau, N., editors, *Latent Variable Analysis and Signal Separation*, pages 427–437. Springer International Publishing.
[307] Walden, A. T. (2001). Wavelet analysis of discrete time series. In Casacuberta, Carlesand Miró-Roig, R. M., Verdera, J., and Xambó-Descamps, S., editors, *European Congress of Mathematics*, pages 627–641, Basel. Birkhäuser Basel.
[308] Wang, C.-A. and Munoz, D. P. (2015). A circuit for pupil orienting responses: implications for cognitive modulation of pupil size. *Current Opinion in Neurobiology*, **33**:134–140.
[309] Wang, X. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological reviews*, **90**(3):1195–1268.
[310] Wang, Y., Siu, T.-S. C., and Cheung, H. (2023). Effect of music emotion on mu and beta oscillations. *Psychology of Music*, **51**(5):1489–1500.
[311] Washburn, A., Román, I., Huberth, M., Gang, N., Dauer, T., Reid, W., Nanou, C., Wright, M., and Fujioka, T. (2019). Musical role asymmetries in piano duet performance influence alpha-band neural oscillation and behavioral synchronization. *Frontiers in Neuroscience*, **13**:1088.
[312] Weber, D., Hertweck, S., Alwanni, H., Fiederer, L., Wang, X., Unruh, F., Fischbach, M., Latoschik, M., and Ball, T. (2021). A structured approach to test the signal quality of electroencephalography measurements during use of head-mounted displays for virtual reality applications. *Frontiers in Neuroscience*, **15**:733673.
[313] Weinstein, D., Launay, J., Pearce, E., Dunbar, R. I. M., and Stewart, L. (2016). Singing and social bonding: changes in connectivity and pain threshold as a function of group size. *Evolution and Human Behavior*, **37**(2):152–158.
[314] White, O. and French, R. (2017). Pupil diameter may reflect motor control and learning. *Journal of Motor Behavior*, **49**(2):141–149.
[315] Xu, G., Mihaylova, T., Li, D., Tian, F., Farrehi, P. M., Parent, J. M., Mashour, G. A., Wang, M. M., and Borjigin, J. (2023). Surge of neurophysiological coupling and connectivity of gamma oscillations in the dying human brain. *Proceedings of the National Academy of Sciences*, **120**(19):e2216268120.
[316] Yang, K., Tong, L., Shu, J., Zhuang, N., Yan, B., and Zeng, Y. (2020). High gamma band EEG closely related to emotion: Evidence from functional network. *Frontiers in Human Neuroscience*, **14**:89.
[317] Yoo, K., Ahn, J., and Lee, S.-H. (2021). The confounding effects of eye blinking on pupillometry, and their remedy. *PLoS One*, **16**(12):1–32.
[318] Zarate, J. M. (2013). The neural control of singing. *Frontiers in Human Neuroscience*, **7**:237.
[319] Zhang, J., Idaji, M. J., Villringer, A., and Nikulin, V. V. (2021). Neuronal biomarkers of parkinson's disease are present in healthy aging. *NeuroImage*, **243**:118512.
[320] Zhou, R., Liu, J., Kumar, S., and Palomar, D. P. (2020). Student's *t*-VAR modeling with missing data via stochastic EM and Gibbs sampling. *IEEE Transactions on Signal Processing*, **68**:6198–6211.
[321] Zhu, K. (2007). *Operator Theory in Function Spaces: Second Edition*, volume 138 of *Mathematical Surveys and Monographs*. American Mathematical Society, New York.
[322] Zima, M., Zima, M., Tichavský, P., Paul, K., and Krajca, V. (2012). Robust removal of short-duration artifacts in long neonatal EEG recordings using wavelet-enhanced ICA and adaptive combining of tentative reconstructions. *Physiological Measurement*, **33**(8):N39–N49.
[323] Zénon, A., Sidibé, M., and Olivier, E. (2014). Pupil size variations correlate with physical effort perception. *Frontiers in Behavioral Neuroscience*, **8**:286.

A Appendix

Data	n_k	Results (%) for the centroid classifiers: PC_1	PC_m	Whitening	IC_q	SIC_q
		Scenario I (Gaussian)				
Example 1	30	45.49 (3.295)	6.167 (10.43)	PCA	2.583 (3.698)	2.520 (5.193)
				PCA-cor	2.583 (3.698)	2.540 (5.869)
				ZCA	2.592 (3.709)	1.590 (2.632)
				ZCA-cor	2.592 (3.709)	1.955 (3.234)
				Cholesky	2.575 (3.702)	1.219 (2.083)
	50	45.47 (3.382)	3.985 (7.754)	PCA	0.970 (1.662)	0.588 (1.058)
				PCA-cor	0.970 (1.662)	0.885 (2.122)
				ZCA	0.965 (1.661)	0.531 (1.051)
				ZCA-cor	0.970 (1.671)	0.678 (1.066)
				Cholesky	0.965 (1.661)	0.490 (1.266)
Example 2	30	44.83 (4.103)	17.35 (15.10)	PCA	4.500 (4.971)	1.716 (2.318)
				PCA-cor	4.492 (4.960)	1.766 (2.411)
				ZCA	4.475 (4.969)	0.876 (1.431)
				ZCA-cor	4.475 (4.969)	0.903 (1.455)
				Chol	4.492 (4.966)	1.688 (2.342)
	50	45.98 (3.047)	11.35 (11.32)	PCA	1.750 (2.198)	0.758 (1.308)
				PCA-cor	1.755 (2.200)	0.969 (1.451)
				ZCA	1.755 (2.200)	0.397 (0.790)
				ZCA-cor	1.750 (2.203)	0.420 (0.787)
				Cholesky	1.740 (2.202)	0.942 (1.517)
Example 3	30	45.16 (3.810)	2.975 (4.474)	PCA	3.525 (4.031)	1.667 (2.472)
				PCA-cor	3.533 (4.026)	1.988 (4.355)
				ZCA	3.533 (4.026)	2.264 (4.476)
				ZCA-cor	3.525 (4.018)	2.004 (4.214)
				Cholesky	3.517 (4.023)	1.286 (2.117)
	50	45.54 (3.230)	1.910 (1.560)	PCA	1.530 (1.974)	0.941 (1.541)
				PCA-cor	1.535 (1.974)	1.129 (1.775)
				ZCA	1.535 (1.974)	1.043 (1.700)
				ZCA-cor	1.535 (1.974)	0.867 (1.505)
				Cholesky	1.530 (1.974)	0.663 (1.044)
		Scenario II (non-Gaussian)				
Example 1	30	45.40 (3.592)	6.742 (10.45)	PCA	3.917 (3.967)	3.909 (5.068)
				PCA-cor	3.917 (3.967)	4.167 (5.950)
				ZCA	3.900 (3.970)	3.438 (2.643)
				ZCA-cor	3.900 (3.970)	3.452 (2.602)
				Cholesky	3.908 (3.967)	3.563 (2.543)
	50	45.97 (3.271)	4.875 (9.073)	PCA	3.075 (3.547)	2.920 (3.190)
				PCA-cor	3.075 (3.547)	2.819 (2.889)
				ZCA	3.075 (3.547)	3.010 (3.448)
				ZCA-cor	3.075 (3.547)	3.071 (3.936)
				Cholesky	3.075 (3.547)	2.678 (2.864)
Example 2	30	45.208 (4.058)	7.45 (10.58)	PCA	8.250 (9.897)	8.349 (10.48)
				PCA-cor	8.250 (9.897)	8.063 (10.77)
				ZCA	8.250 (9.889)	7.155 (9.419)
				ZCA-cor	8.242 (9.892)	6.800 (9.561)
				Cholesky	8.233 (9.897)	7.343 (9.434)
	50	46.29 (2.748)	7.805 (11.97)	PCA	7.145 (10.48)	6.053 (8.575)
				PCA-cor	7.150 (10.50)	5.840 (8.591)
				ZCA	7.155 (10.50)	6.694 (10.56)
				ZCA-cor	7.155 (10.50)	6.640 (10.69)
				Cholesky	7.150 (10.48)	6.075 (8.519)
Example 3	30	45.15 (4.049)	10.92 (11.87)	PCA	6.858 (7.936)	5.951 (6.266)
				PCA-cor	6.858 (7.936)	6.516 (7.232)
				ZCA	6.867 (7.934)	7.060 (8.419)
				ZCA-cor	6.867 (7.934)	6.425 (7.626)
				Cholesky	6.858 (7.936)	5.654 (6.642)
	50	45.94 (3.092)	10.92 (11.69)	PCA	5.085 (6.098)	4.832 (5.616)
				PCA-cor	5.085 (6.098)	5.119 (6.249)
				ZCA	5.095 (6.120)	4.935 (6.120)
				ZCA-cor	5.100 (6.119)	4.873 (5.696)
				Cholesky	5.095 (6.095)	4.173 (4.228)

Table A.1: Simulation results for the mean and standard deviation (in parentheses) of the classification errors obtained with 200 repetitions of the experiment for different sample sizes and zero error variance. PC_1, first principal component; PC_m, principal component with lowest kurtosis coefficient; IC_q, qth independent component (minimal kurtosis); SIC_q, qth smoothed independent component; PCA, principal component analysis whitening; PCA-cor, principal component analysis correlated whitening; ZCA, zero-phase component analysis or Mahalanobis whitening; ZCA-cor, zero-phase component analysis or Mahalanobis correlated whitening; Cholesky, Cholesky whitening.

Method	Universal				Multiplicative scaling				ENID			
Wavelet	MRMSE	$ISNR_{var}$	$ISNR_{cor}$	NCD	MRMSE	$ISNR_{var}$	$ISNR_{cor}$	NCD	MRMSE	$ISNR_{var}$	$ISNR_{cor}$	NCD
	Space factorization via FastICA											
	Estimation of artifact 1											
d4	0.657841 (0.054892)	0.463164 (0.060811)	0.747474 (0.048975)	0.997745 (0.014261)	0.14744 (0.024254)	0.527069 (0.071056)	0.989449 (0.003292)	1.004141 (0.003844)	0.165179 (0.020554)	0.546836 (0.069524)	0.986186 (0.003157)	1.003312 (0.003796)
d8	0.654121 (0.049689)	0.456799 (0.057024)	0.75094 (0.042332)	0.997804 (0.010781)	0.146762 (0.022762)	0.516577 (0.066261)	0.989174 (0.003035)	1.003399 (0.00425)	0.169321 (0.017953)	0.534618 (0.064933)	0.984854 (0.002682)	1.003698 (0.003009)
d12	0.664732 (0.062922)	0.467959 (0.06702)	0.741682 (0.056257)	0.997843 (0.010855)	0.158773 (0.033939)	0.531458 (0.080123)	0.987699 (0.003974)	1.003011 (0.003537)	0.187063 (0.025776)	0.547876 (0.078217)	0.981829 (0.002979)	1.002668 (0.00314)
s4	0.674182 (0.069498)	0.465488 (0.065459)	0.733095 (0.065597)	0.995529 (0.010776)	0.149827 (0.029964)	0.533555 (0.08142)	0.989299 (0.003555)	1.004635 (0.003541)	0.165849 (0.025953)	0.552471 (0.078376)	0.986335 (0.003442)	1.00337 (0.004673)
s8	0.662008 (0.070549)	0.461941 (0.068915)	0.745201 (0.061283)	0.996788 (0.010722)	0.147735 (0.043849)	0.524252 (0.087342)	0.989401 (0.004757)	1.00347 (0.005009)	0.152865 (0.050574)	0.53012 (0.083029)	0.987959 (0.007949)	1.003197 (0.00403)
s12	0.689355 (0.162002)	0.478221 (0.138828)	0.737399 (0.06731)	0.995632 (0.011597)	0.168446 (0.109922)	0.550704 (0.21001)	0.985715 (0.017965)	1.002996 (0.006368)	0.17984 (0.14963)	0.554262 (0.17579)	0.976944 (0.070141)	1.001929 (0.003731)
l2	0.653949 (0.075971)	0.45909 (0.061999)	0.751058 (0.065643)	0.999594 (0.00668)	0.174939 (0.035771)	0.537958 (0.081046)	0.985109 (0.004112)	1.003979 (0.003385)	0.175247 (0.039638)	0.559213 (0.075559)	0.985085 (0.004854)	1.003903 (0.00421)
l4	0.658251 (0.089136)	0.455168 (0.048934)	0.750035 (0.059427)	0.996658 (0.011368)	0.150381 (0.054262)	0.519542 (0.068559)	0.987702 (0.01367)	1.002709 (0.004035)	0.168047 (0.050643)	0.53917 (0.0648)	0.984441 (0.013343)	1.003644 (0.004307)
l6	0.682652 (0.141768)	0.478555 (0.117798)	0.740986 (0.065049)	0.995317 (0.0117)	0.154036 (0.096781)	0.549401 (0.175358)	0.987827 (0.0174)	1.003927 (0.004826)	0.161814 (0.125562)	0.549477 (0.14955)	0.980589 (0.062262)	1.003116 (0.003972)
	Estimation of artifact 2 after removal of artifact 1											
d4	0.448573 (0.080736)	7.058343 (1.5539)	0.206537 (0.091623)	0.994474 (0.01221)	0.179966 (0.043525)	13.08566 (2.08737)	0.94488 (0.013193)	0.988835 (0.005719)	0.219096 (0.034826)	11.30297 (2.346056)	0.752741 (0.105738)	1.001098 (0.003419)
d8	0.478646 (0.114056)	6.231146 (1.848091)	0.190032 (0.086348)	0.994389 (0.011764)	0.165698 (0.051095)	12.07822 (2.01196)	0.954331 (0.028477)	0.990374 (0.004926)	0.229771 (0.058317)	10.01514 (2.374593)	0.685261 (0.132237)	1.002501 (0.003344)
d12	0.463715 (0.117373)	7.195422 (2.538593)	0.212163 (0.097225)	0.997188 (0.012264)	0.174244 (0.053313)	13.40127 (3.739797)	0.960158 (0.020007)	0.990068 (0.006701)	0.228199 (0.053029)	11.39044 (3.863455)	0.690974 (0.151923)	1.000551 (0.002689)
s4	0.481191 (0.094494)	6.876887 (2.964637)	0.182231 (0.08988)	0.996344 (0.011589)	0.166153 (0.049944)	13.08194 (3.678795)	0.938575 (0.020764)	0.988792 (0.007044)	0.221005 (0.059506)	11.25547 (4.291475)	0.736291 (0.101049)	1.001947 (0.004974)
s8	0.477981 (0.122649)	6.148436 (2.141086)	0.184581 (0.086663)	0.999877 (0.011393)	0.162497 (0.054183)	12.09902 (2.521667)	0.938797 (0.02968)	0.989764 (0.006364)	0.228402 (0.066671)	10.22325 (2.781534)	0.718982 (0.123985)	1.002124 (0.003841)
s12	0.450076 (0.121299)	6.908305 (1.750583)	0.195547 (0.102384)	0.9948 (0.011853)	0.171401 (0.051476)	12.83481 (2.106197)	0.941029 (0.024431)	0.990128 (0.004157)	0.222343 (0.0637)	11.16996 (2.443443)	0.758519 (0.130586)	1.001261 (0.003741)
l2	0.474286 (0.116032)	6.38784 (1.973903)	0.201856 (0.095818)	0.991828 (0.014245)	0.158886 (0.051663)	12.63274 (2.028879)	0.920712 (0.035078)	0.987492 (0.008565)	0.204492 (0.059059)	11.19462 (2.486581)	0.831017 (0.113073)	1.001821 (0.004013)
l4	0.481977 (0.108547)	6.526081 (1.519683)	0.177943 (0.105206)	0.995067 (0.011407)	0.16754 (0.047464)	12.85754 (1.828242)	0.938471 (0.025187)	0.989671 (0.005161)	0.222795 (0.06059)	10.80704 (2.158874)	0.713172 (0.107712)	1.002276 (0.003537)
l6	0.475801 (0.135504)	6.688158 (2.230017)	0.194956 (0.102018)	0.994838 (0.011626)	0.170529 (0.068397)	12.56913 (2.374871)	0.947924 (0.020882)	0.989334 (0.005779)	0.23031 (0.075192)	10.77806 (2.841559)	0.725982 (0.130428)	1.001989 (0.004048)
	Residual process											
d4	0.738246 (0.080425)	14.0968 (1.017869)	0.776452 (0.054885)	1.000071 (0.00061)	0.218931 (0.035644)	12.41901 (0.125124)	0.987847 (0.00525)	1.000045 (0.000278)	0.653094 (0.205282)	17.4102 (2.876577)	0.841971 (0.140961)	1.003297 (0.00401)
d8	0.765165 (0.098509)	14.39541 (1.194617)	0.75228 (0.081967)	1.000145 (0.000653)	0.211383 (0.044286)	12.41617 (0.140848)	0.986339 (0.00924)	1.000007 (0.000391)	0.585683 (0.198107)	16.2318 (2.572425)	0.869526 (0.124406)	1.005189 (0.009544)
d12	0.758087 (0.087336)	14.13698 (1.216202)	0.76204 (0.065671)	1.000044 (0.000344)	0.22903 (0.055803)	12.39145 (0.160723)	0.983142 (0.011898)	1.000076 (0.000479)	0.558872 (0.207715)	15.76073 (2.633102)	0.876353 (0.123663)	1.003598 (0.005857)
s4	0.777628 (0.08824)	14.1943 (1.17767)	0.749269 (0.062649)	1.000078 (0.000555)	0.2191 (0.042485)	12.42149 (0.136114)	0.984919 (0.008589)	0.999975 (0.000469)	0.659411 (0.212358)	17.63971 (3.49145)	0.836779 (0.149657)	1.003288 (0.003653)
s8	0.779329 (0.115211)	14.2918 (1.231617)	0.743803 (0.082314)	1.000183 (0.000737)	0.214661 (0.054557)	12.4172 (0.153326)	0.984814 (0.015237)	1.000042 (0.00038)	0.601071 (0.220126)	16.52397 (2.94092)	0.857555 (0.134843)	1.004735 (0.007739)
s12	0.77373 (0.187424)	14.03099 (1.601392)	0.760316 (0.088662)	1.000289 (0.001056)	0.235843 (0.1075)	12.35732 (0.304361)	0.97791 (0.041909)	1.000045 (0.000472)	0.589722 (0.229105)	16.08278 (3.070026)	0.866431 (0.12719)	1.005028 (0.007345)
l2	0.757961 (0.121074)	14.47602 (1.253644)	0.758431 (0.089847)	1.00242 (0.002298)	0.233153 (0.040547)	12.54336 (0.14477)	0.985532 (0.009437)	1.002417 (0.002354)	0.9726 (0.172043)	16.0039 (3.10405)	0.57697 (0.224512)	1.008963 (0.004623)
l4	0.765353 (0.119101)	14.41742 (1.142444)	0.752045 (0.085337)	1.00012 (0.000413)	0.21357 (0.053796)	12.43673 (0.1391)	0.986281 (0.018187)	0.999979 (0.000322)	0.664585 (0.20677)	17.60278 (3.086074)	0.834492 (0.143537)	1.003279 (0.004079)
l6	0.777206 (0.153319)	14.11338 (1.480863)	0.7524 (0.087891)	1.000212 (0.000855)	0.225024 (0.100507)	12.37257 (0.268305)	0.975033 (0.045931)	1.000096 (0.000488)	0.620489 (0.208674)	16.61105 (2.811296)	0.856105 (0.130395)	1.00448 (0.005421)
	Direct wavelet thresholding without space factorization											
	Residual process											
d4	5.224387 (6.7e−05)	0.001952 (3.9e−05)	0.000859 (1.9e−05)	1.007742 (0.005988)	0.350488 (0.032121)	12.63689 (0.112317)	0.959848 (0.008763)	1.016823 (0.000997)	4.655145 (0.014108)	0.975732 (0.023955)	0.027352 (0.003129)	1.002533 (0.003737)
d8	5.224411 (6.8e−05)	0.001385 (2e−05)	0.000621 (1.1e−05)	1.010126 (0.008669)	0.364778 (0.032564)	12.55015 (0.11611)	0.954875 (0.009424)	1.016263 (0.001566)	4.959671 (0.015523)	0.488918 (0.028622)	0.01033 (0.001262)	1.001518 (0.003497)
d12	5.224431 (6.8e−05)	0.001346 (2e−05)	0.000623 (1.1e−05)	1.007699 (0.006925)	0.392122 (0.033239)	12.47344 (0.117539)	0.945953 (0.010439)	1.016001 (0.001352)	5.092638 (0.073315)	0.367497 (0.475808)	0.008612 (0.015257)	1.001061 (0.003151)
s4	5.224387 (6.7e−05)	0.001952 (3.9e−05)	0.000859 (1.9e−05)	1.007742 (0.005988)	0.350488 (0.032121)	12.63689 (0.112317)	0.959848 (0.008763)	1.016823 (0.000997)	4.655145 (0.014108)	0.975732 (0.023955)	0.027352 (0.003129)	1.002533 (0.003737)
s8	5.22441 (6.8e−05)	0.001377 (2e−05)	0.000616 (1.1e−05)	1.009041 (0.008097)	0.361796 (0.031641)	12.55376 (0.118119)	0.95602 (0.008998)	1.016371 (0.000831)	4.27154 (0.02344)	1.628292 (0.037872)	0.056054 (0.005375)	1.003574 (0.005942)
s12	5.224425 (6.8e−05)	0.001339 (2e−05)	0.000614 (1e−05)	1.008357 (0.006263)	0.366345 (0.033106)	12.53038 (0.122384)	0.954428 (0.009598)	1.016139 (0.000905)	4.588933 (0.201799)	1.093925 (0.331744)	0.033573 (0.015406)	1.003898 (0.005626)
l2	5.224434 (6.8e−05)	0.008067 (0.000318)	0.003731 (0.000162)	1.011359 (0.008578)	0.356622 (0.035124)	12.67734 (0.120713)	0.957368 (0.010016)	1.01729 (0.001456)	4.530846 (0.204702)	1.253825 (0.363926)	0.033993 (0.014062)	1.001847 (0.002889)
l4	5.224387 (6.7e−05)	0.001952 (3.9e−05)	0.000859 (1.9e−05)	1.007742 (0.005988)	0.350488 (0.032121)	12.63689 (0.112317)	0.959848 (0.008763)	1.016823 (0.000997)	4.655145 (0.014108)	0.975732 (0.023955)	0.027352 (0.003129)	1.002533 (0.003737)
l6	5.224397 (6.8e−05)	0.001463 (2.5e−05)	0.000643 (1.1e−05)	1.008026 (0.006987)	0.366042 (0.031447)	12.57657 (0.120449)	0.954616 (0.009027)	1.016468 (0.000706)	4.217773 (0.024888)	1.73342 (0.037461)	0.059514 (0.005928)	1.001965 (0.002704)

Table A.2: Simulation 1. Results for the mean and standard deviation (in parentheses).

	Method	Multiplicative scaling				ENID			
Bumps	Wavelet	MRMSE	$ISNR_{var}$	$ISNR_{cor}$	NCD	MRMSE	$ISNR_{var}$	$ISNR_{cor}$	NCD
6	d4	0.159291 (0.069863)	0.358825 (0.144347)	0.984507 (0.016588)	1.002765 (0.004964)	0.193867 (0.063698)	0.379448 (0.144599)	0.979141 (0.01601)	1.005223 (0.010008)
	d8	0.163211 (0.070968)	0.356399 (0.143688)	0.983955 (0.017346)	1.005484 (0.011355)	0.199216 (0.06538)	0.372693 (0.148255)	0.978004 (0.016922)	1.00467 (0.013603)
	d12	0.164168 (0.071788)	0.355782 (0.142736)	0.983331 (0.01806)	1.005512 (0.013642)	0.195436 (0.061398)	0.367613 (0.147055)	0.978602 (0.016409)	1.004877 (0.013144)
	s4	0.161111 (0.070123)	0.358263 (0.143787)	0.984421 (0.016788)	1.005032 (0.008621)	0.194974 (0.064198)	0.378684 (0.143937)	0.978802 (0.0162)	1.005997 (0.012937)
	s8	0.162521 (0.071738)	0.356998 (0.143806)	0.984027 (0.017515)	1.00451 (0.014427)	0.191091 (0.064118)	0.372247 (0.144607)	0.980151 (0.016253)	1.003506 (0.014803)
	s12	0.166472 (0.069998)	0.356361 (0.144719)	0.983452 (0.017627)	1.003101 (0.014161)	0.192127 (0.06193)	0.368337 (0.146528)	0.979802 (0.016172)	1.002007 (0.008341)
	l4	0.159557 (0.069042)	0.357821 (0.144692)	0.984501 (0.016638)	1.003891 (0.010418)	0.193677 (0.06453)	0.378221 (0.144862)	0.979103 (0.016428)	1.00473 (0.010382)
	l6	0.159694 (0.070206)	0.356714 (0.143863)	0.984282 (0.016722)	1.001941 (0.006664)	0.187241 (0.059847)	0.370021 (0.146461)	0.981099 (0.015089)	1.0053 (0.012834)
	l14	0.167355 (0.069427)	0.355992 (0.143556)	0.984022 (0.017148)	1.002998 (0.01267)	0.205929 (0.05885)	0.369595 (0.147001)	0.976907 (0.015745)	1.000983 (0.013902)
12	d4	0.724685 (1.612114)	0.685095 (2.102155)	0.901374 (0.129116)	1.006923 (0.020861)	0.591157 (0.856144)	0.343955 (0.512492)	0.890397 (0.134923)	1.010199 (0.018469)
	d8	0.551384 (0.987695)	0.457148 (1.194392)	0.910466 (0.116887)	1.006886 (0.018809)	0.627072 (1.205259)	0.350894 (0.604046)	0.90271 (0.115208)	1.006781 (0.022956)
	d12	0.703307 (1.532512)	1.11241 (4.914024)	0.90176 (0.123371)	1.001302 (0.01846)	0.60838 (1.009288)	0.34754 (0.585179)	0.900579 (0.108871)	1.002743 (0.019389)
	s4	0.641476 (1.154969)	0.55715 (1.348479)	0.920366 (0.092921)	1.006083 (0.018532)	0.732812 (1.409297)	0.41983 (0.732078)	0.918276 (0.079527)	1.006933 (0.013144)
	s8	0.55633 (0.873282)	0.458521 (1.149952)	0.893676 (0.155817)	1.002721 (0.017266)	0.621449 (1.037493)	0.363579 (0.627734)	0.898906 (0.127887)	1.004279 (0.018832)
	s12	0.513823 (0.767747)	0.364192 (0.867869)	0.911304 (0.110165)	1.004306 (0.017468)	0.56101 (0.90929)	0.314307 (0.465586)	0.902988 (0.109343)	1.0024 (0.0171)
	l4	0.448391 (0.605762)	0.335321 (0.819777)	0.903447 (0.134264)	1.004411 (0.016742)	0.488141 (0.6715)	0.307291 (0.498307)	0.897219 (0.130477)	1.005373 (0.019825)
	l6	0.483993 (0.900187)	0.346639 (0.913789)	0.912547 (0.112069)	1.003771 (0.02029)	0.532507 (1.039031)	0.29606 (0.468976)	0.904181 (0.110642)	1.003258 (0.014893)
	l14	0.617499 (1.408627)	1.014224 (4.871136)	0.907653 (0.114166)	1.002906 (0.022153)	0.504228 (0.650793)	0.290652 (0.426294)	0.899666 (0.110273)	1.000883 (0.018175)
18	d4.2	2.289305 (3.78737)	3.704562 (8.783911)	0.782068 (0.191837)	1.006454 (0.019372)	1.117433 (1.203409)	0.417781 (0.593175)	0.771339 (0.181056)	1.003642 (0.019137)
	d8	2.321636 (3.864667)	4.126944 (10.05845)	0.805676 (0.179522)	1.00599 (0.017931)	1.147826 (1.207275)	0.406517 (0.537533)	0.798205 (0.167418)	1.005686 (0.017435)
	d12	3.235971 (4.531725)	5.977425 (11.170955)	0.816142 (0.171087)	1.005847 (0.019492)	1.443686 (1.746928)	0.561799 (0.917904)	0.797775 (0.171935)	1.00421 (0.014973)
	s4	2.745307 (3.962096)	4.076504 (8.834003)	0.786004 (0.186007)	1.006435 (0.0184)	1.651523 (2.170198)	0.615613 (0.878697)	0.779281 (0.181533)	1.006058 (0.015548)
	s8	2.575273 (3.951517)	3.86153 (9.080391)	0.822613 (0.153616)	1.003017 (0.018229)	1.68199 (2.230744)	0.625587 (0.922405)	0.810139 (0.141282)	1.006552 (0.014503)
	s12	2.638055 (4.057873)	4.638803 (10.229077)	0.801158 (0.185309)	1.007601 (0.01738)	1.338485 (1.56303)	0.470911 (0.674924)	0.810442 (0.156894)	1.005819 (0.016086)
	l4	2.995763 (4.36862)	5.353475 (11.095618)	0.79115 (0.177379)	1.005817 (0.01713)	1.573052 (2.083158)	0.638905 (1.008776)	0.795401 (0.159402)	1.008974 (0.016238)
	l6	2.695803 (4.228284)	4.760014 (10.92135)	0.805486 (0.174724)	1.006337 (0.018507)	1.470696 (2.008197)	0.537353 (0.833422)	0.810694 (0.154772)	1.004203 (0.016277)
	l14	2.310197 (3.785727)	3.618736 (8.676085)	0.795568 (0.188833)	1.004359 (0.019337)	1.196305 (1.219698)	0.407566 (0.518579)	0.791504 (0.174445)	1.004859 (0.017136)

Table A.3: Simulation 2. Results for the mean and standard deviation (in parentheses).

	Method	Multiplicative scaling				ENID			
Bumps	Wavelet	MRMSE	$ISNR_{var}$	$ISNR_{cor}$	NCD	MRMSE	$ISNR_{var}$	$ISNR_{cor}$	NCD
6	d4	0.278977 (0.075478)	0.310165 (0.153355)	0.956232 (0.022263)	1.003926 (0.009127)	0.295637 (0.066051)	0.332025 (0.158492)	0.951295 (0.020952)	1.002317 (0.005758)
	d8	0.381498 (0.710254)	0.834804 (3.74196)	0.955144 (0.023296)	1.002774 (0.005705)	0.366557 (0.475057)	0.361546 (0.325496)	0.94925 (0.02421)	1.003113 (0.007367)
	d12	0.286015 (0.078918)	0.308136 (0.152218)	0.953715 (0.023619)	1.003407 (0.007132)	0.300079 (0.071558)	0.320898 (0.161124)	0.949205 (0.022617)	1.002044 (0.005747)
	s4	0.366373 (0.630382)	0.43806 (0.899154)	0.95649 (0.022628)	1.001635 (0.006477)	0.395209 (0.71569)	0.393364 (0.442472)	0.95164 (0.02111)	1.00322 (0.008703)
	s8	0.36444 (0.585368)	0.431208 (0.869336)	0.954911 (0.023046)	1.001013 (0.006736)	0.374136 (0.572328)	0.402068 (0.562952)	0.951229 (0.022525)	1.00259 (0.005957)
	s12	0.284629 (0.07721)	0.308566 (0.152475)	0.954309 (0.022996)	0.999097 (0.008341)	0.295098 (0.071145)	0.320162 (0.154227)	0.951062 (0.021983)	0.99951 (0.008496)
	l4	0.279207 (0.075893)	0.310101 (0.153201)	0.956123 (0.022487)	1.002857 (0.00571)	0.295626 (0.066044)	0.33199 (0.158229)	0.951294 (0.020947)	1.00422 (0.005852)
	l6	0.3794 (0.718319)	0.453815 (1.026379)	0.956133 (0.023038)	1.002638 (0.006066)	0.390604 (0.704092)	0.377876 (0.397596)	0.951748 (0.022917)	1.003278 (0.005562)
	l14	0.360668 (0.534311)	0.431371 (0.877391)	0.953921 (0.023169)	1.00117 (0.007913)	0.389645 (0.598784)	0.381469 (0.427141)	0.946951 (0.023886)	0.999426 (0.007593)
12	d4	0.53741 (1.482123)	1.057295 (6.257677)	0.943494 (0.037982)	1.001305 (0.011715)	0.483093 (0.947932)	0.243026 (0.410164)	0.936155 (0.036315)	1.00155 (0.013933)
	d8	0.576363 (1.289214)	0.432287 (1.299278)	0.944652 (0.03944)	1.002677 (0.017383)	0.626345 (1.430758)	0.28966 (0.545162)	0.935713 (0.039442)	1.000418 (0.018353)
	d12	0.324535 (0.1045)	0.167582 (0.0788)	0.944216 (0.038764)	0.999009 (0.02012)	0.354507 (0.096669)	0.176342 (0.081029)	0.933003 (0.038674)	1.000247 (0.020905)
	s4	0.325248 (0.101277)	0.170678 (0.07947)	0.944425 (0.037728)	1.002563 (0.016044)	0.346576 (0.092593)	0.184544 (0.083682)	0.937145 (0.03608)	1.003901 (0.017645)
	s8	0.321038 (0.103521)	0.168496 (0.078943)	0.945582 (0.038128)	0.998785 (0.017964)	0.34569 (0.098836)	0.180468 (0.082236)	0.936463 (0.037626)	0.999636 (0.014431)
	s12	0.628834 (1.524998)	0.773952 (3.139666)	0.943981 (0.039028)	0.99604 (0.01692)	0.575311 (1.161048)	0.292821 (0.583894)	0.935015 (0.039318)	0.99919 (0.014637)
	l4	0.324999 (0.101587)	0.170399 (0.079182)	0.944444 (0.037812)	1.00159 (0.017377)	0.34633 (0.092664)	0.184288 (0.08348)	0.937204 (0.036112)	1.005251 (0.017332)
	l6	0.322387 (0.10313)	0.16838 (0.079303)	0.945189 (0.038304)	0.996719 (0.017164)	0.345075 (0.09709)	0.180264 (0.083373)	0.936968 (0.037112)	0.99962 (0.017637)
	l14	0.327088 (0.104244)	0.16813 (0.078912)	0.9434 (0.038998)	0.997651 (0.014466)	0.356428 (0.101977)	0.177906 (0.082459)	0.931794 (0.040157)	0.997215 (0.018032)
18	d4	0.905044 (2.261311)	1.605199 (7.186142)	0.919071 (0.071803)	1.008833 (0.020804)	0.780203 (1.442554)	0.272594 (0.518534)	0.909313 (0.066224)	1.005734 (0.022098)
	d8	0.610808 (1.355359)	0.300605 (1.093299)	0.91782 (0.081938)	1.007045 (0.021838)	0.678018 (1.422401)	0.213704 (0.464417)	0.897264 (0.09483)	1.005309 (0.022458)
	d12	0.458799 (0.410554)	0.165084 (0.199359)	0.920552 (0.074049)	1.008349 (0.023205)	0.513442 (0.46369)	0.163986 (0.170019)	0.904851 (0.07156)	1.002686 (0.021561)
	s4	0.866828 (2.039579)	1.394554 (6.129677)	0.918756 (0.074621)	1.007634 (0.021973)	0.76726 (1.359431)	0.287662 (0.599528)	0.909053 (0.069523)	1.007725 (0.020853)
	s8	0.449803 (0.401718)	0.16601 (0.198488)	0.924418 (0.070533)	1.00834 (0.021393)	0.50325 (0.453814)	0.166742 (0.168517)	0.910259 (0.065097)	1.006501 (0.021204)
	s12	0.852483 (1.972862)	0.479465 (1.544905)	0.920621 (0.072136)	1.004966 (0.020075)	0.925264 (2.032946)	0.283825 (0.581927)	0.901148 (0.073822)	1.004148 (0.020672)
	l4	0.661345 (1.423682)	0.510414 (2.411068)	0.918585 (0.072485)	1.004082 (0.018574)	0.640855 (1.017962)	0.235268 (0.476841)	0.908977 (0.067377)	1.00565 (0.02271)
	l6	0.451066 (0.402885)	0.167089 (0.201114)	0.924094 (0.070665)	1.003885 (0.022509)	0.501066 (0.460699)	0.165798 (0.167181)	0.911239 (0.067503)	1.008086 (0.019978)
	l14	1.024389 (2.455579)	1.058305 (4.1202)	0.919855 (0.075064)	1.004743 (0.020503)	0.928163 (1.838425)	0.305748 (0.640235)	0.902548 (0.072899)	1.00622 (0.019914)

Table A.4: Simulation 2 (with PCA reduction). Results for the mean and standard deviation (in parentheses).

Method	Multiplicative scaling				ENID			
Wavelet	MRMSE	$ISNR_{var}$	$ISNR_{cor}$	NCD	MRMSE	$ISNR_{var}$	$ISNR_{cor}$	NCD
	Estimation of artifact 3							
d4	0.32449 (0.110405)	19.666781 (16.022239)	0.375898 (0.345606)	0.996752 (0.011416)	0.307209 (0.015833)	35.087096 (59.891356)	0.161303 (0.074897)	1.001241 (0.009637)
d8	0.300474 (0.019488)	10.396631 (1.956171)	0.194692 (0.038094)	0.98356 (0.010629)	0.313112 (0.014667)	27.166834 (37.185624)	0.174632 (0.071993)	1.002353 (0.010394)
d12	0.297385 (0.021813)	10.592946 (1.744777)	0.194671 (0.037912)	0.983509 (0.012712)	0.318688 (0.017602)	37.031991 (57.92455)	0.176195 (0.067008)	0.998639 (0.009202)
s4	0.329951 (0.114453)	20.220228 (16.274081)	0.376508 (0.345268)	0.998346 (0.011645)	0.30677 (0.017629)	30.073986 (49.705099)	0.155408 (0.059579)	1.001806 (0.009501)
s8	0.30127 (0.020871)	10.380395 (1.510583)	0.195429 (0.040296)	0.987409 (0.013945)	0.315699 (0.021737)	33.968198 (48.685646)	0.163204 (0.041023)	0.999826 (0.008382)
s12	0.297059 (0.022694)	10.519902 (1.719476)	0.192203 (0.041188)	0.983148 (0.011443)	0.316246 (0.016719)	35.200289 (59.096552)	0.17844 (0.062609)	1.001565 (0.00913)
l4	0.330069 (0.112139)	20.686284 (16.420751)	0.441564 (0.368488)	0.997428 (0.01235)	0.311746 (0.025501)	44.478047 (68.038811)	0.154451 (0.058194)	1.001661 (0.011234)
l6	0.295269 (0.019324)	10.202649 (1.146886)	0.195629 (0.039621)	0.986915 (0.011474)	0.311885 (0.017601)	24.357743 (31.35387)	0.163433 (0.049073)	0.999569 (0.01018)
l14	0.295777 (0.038664)	10.93541 (5.039836)	0.235729 (0.200106)	0.98914 (0.014021)	0.312929 (0.010345)	24.031415 (29.78275)	0.172338 (0.037118)	1.002093 (0.009656)
	Residual process							
d4	0.3279 (0.109782)	0.826845 (0.465825)	0.450759 (0.402703)	1.000108 (0.000364)	1.124316 (0.022943)	27.945365 (58.116577)	0.318605 (0.072596)	1.008613 (0.001926)
d8	0.304247 (0.019222)	1.141014 (0.101401)	0.783703 (0.097536)	1.000022 (0.000333)	1.099346 (0.02275)	15.605888 (27.662141)	0.304468 (0.09202)	1.024867 (0.007719)
d12	0.301585 (0.021501)	1.108117 (0.10845)	0.773917 (0.107725)	0.999997 (0.000387)	1.096181 (0.025846)	21.588673 (49.320924)	0.280452 (0.076679)	1.015238 (0.007331)
s4	0.333332 (0.113818)	0.807657 (0.478003)	0.411115 (0.407524)	0.999946 (0.000404)	1.124426 (0.022933)	32.274049 (65.581258)	0.325612 (0.076183)	1.008835 (0.002014)
s8	0.305048 (0.020596)	1.128588 (0.131168)	0.790893 (0.109763)	0.999959 (0.000448)	1.098626 (0.027366)	24.567193 (59.515188)	0.296483 (0.080671)	1.019395 (0.00525)
s12	0.301153 (0.022367)	1.113911 (0.108619)	0.795174 (0.119217)	1.000086 (0.000428)	1.089849 (0.024621)	26.34881 (58.767608)	0.277901 (0.05723)	1.019431 (0.007893)
l4	0.333439 (0.111529)	0.802553 (0.477065)	0.373569 (0.388143)	0.99997 (0.000382)	1.128657 (0.023591)	42.141073 (89.750224)	0.322659 (0.075404)	1.008547 (0.002108)
l6	0.298853 (0.019067)	1.137067 (0.113262)	0.838727 (0.082222)	1.000004 (0.000374)	1.107698 (0.022136)	15.021812 (32.288008)	0.318269 (0.105439)	1.013479 (0.003051)
l14	0.300058 (0.038306)	1.0931 (0.176757)	0.772466 (0.205443)	0.999987 (0.000418)	1.084504 (0.020424)	9.642752 (0.466005)	0.279749 (0.086515)	1.023144 (0.007724)

Table A.5: Simulation 3. Results for the mean and standard deviation (in parentheses).