Extraction of multivariate components in brain signals obtained during general anesthesia

Mariia Fedotenkova

To cite this version:

Mariia Fedotenkova. Extraction of multivariate components in brain signals obtained during general anesthesia. Computer Science [cs]. Université de Lorraine, 2016. English. NNT : 2016LORR0189 . tel-01527901v2

HAL Id: tel-01527901
https://hal.science/tel-01527901v2

Submitted on 13 Dec 2016

UNIVERSITÉ
DE LORRAINE

École doctorale IAEM Lorraine

Extraction de composants multivariés des signaux cérébraux obtenus pendant l'anesthésie générale.

(Extraction of multivariate components in brain signals obtained during general anesthesia)

THÈSE

présentée et soutenue publiquement le 2 décembre 2016

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention informatique)

par

Mariia FEDOTENKOVA

Composition du jury

Directeur :	Axel HUTT, Directeur de Recherche, INRIA Nancy Grand-Est.
Rapporteurs :	Frédéric CHAZAL, Directeur de Recherche, INRIA Saclay Gordon PIPA, Professeur, Université de Osnabrueck
Examinateur :	Anne BOYER, Professeur, Université de Lorraine

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

Mis en page avec la classe thesul.

Abstract

Nowadays, surgical operations are impossible to imagine without general anesthesia, which involves loss of consciousness, immobility, amnesia and analgesia. Understanding mechanisms underlying each of these effects guarantees well-controlled medical treatment. This thesis focuses on analgesia effect of general anesthesia, more specifically, on patients reaction to nociceptive stimuli. We also study differences in the reaction between different anesthetic drugs. The study was conducted on dataset consisting of 230 EEG signals: pre- and post-incision recordings obtained form 115 patients, who received desflurane and propofol.

The first stage of the study comprise power spectral analysis, which is a widespread approach in signal processing. Spectral information was described by fitting the background activity, that exposes $1/f$ behavior, to power spectral density estimates of the EEG signals and measuring power contained in delta and alpha bands relatively to the power of background activity.

A further improvement was done by expanding spectra with time information due to observed non-stationary nature of EEG signals. To obtain time-frequency representations of the signals we apply three different methods: scalogram (based on continuous wavelet transform), conventional spectrogram, and spectrogram reassignment. The latter allows to ameliorate readability of a time-frequency representation by reassigning energy contained in spectrogram to more precise positions.

Subsequently, obtained spectrograms were used as phase space reconstruction in recurrence analysis and its quantification by complexity measure. Recurrence analysis allows to describe and visualize recurrent dynamics of a system and discover structural patterns contained in the data. Here, recurrence plots were used as rewriting grammar to turn an original signal into a symbolic sequence, where each symbol represents a certain state of the system. After computing three different complexity measures of resulting symbolic sequences they are used as features for classification.

Finally, combining features obtained with power spectral analysis and recurrence symbolic analysis, we perform classification of the data using two classification methods: linear discriminant analysis and support vector machines. Classification was carried out on two-class problem, distinguishing between pre-/post-incision EEG signals, as well as between two different anesthetic drugs, desflurane and propofol.

Keywords: multivariate analysis of signals, computational neuroscience, time-frequency analysis, anesthesia, recurrence analysis

Résumé

De nos jours, les opérations chirurgicales sont impossibles à imaginer sans anesthésie générale, qui implique la perte de conscience, l'immobilité, l'amnésie et l'analgésie. La compréhension des mécanismes sous-jacents de chacun de ces effets garantit un traitement médical bien contrôlé. Cette thèse se concentre sur l'effet analgésique de l'anesthésie générale, précisément, sur la réaction du patient aux stimuli nociceptifs. Nous étudions également les différences des réactions entre différents médicaments anesthésiques. L'étude a été effectuée sur un ensemble de données constituées de 230 signaux EEG : enregistrements pré- et post-incision obtenus sur 115 patients qui ont reçu du desflurane et du propofol.

La première phase de l'étude comprend l'analyse spectrale de puissance, qui est une méthode très répandue dans le traitement du signal. L'information spectrale a été décrite en ajustant l'activité de fond, qui exhibe un comportement $1/f$, aux estimations de la densité spectrale de puissance des signaux d'EEG et en mesurant la puissance contenue dans des bandes delta et alpha par rapport à la puissance de l'activité de fond.

Une autre amélioration a été réalisée par l'expansion des spectres avec des informations de temps en raison de la nature non stationnaire observée dans les signaux EEG. Pour obtenir les représentations temps-fréquence des signaux nous appliquons trois méthodes différentes : scalogramme (basé sur la transformée en ondelettes continue), spectrogramme classique, et réaffectation de spectrogramme. Celle-ci permet d'améliorer la lisibilité d'une représentation temps-fréquence en réaffectant l'énergie contenue dans le spectrogramme à des positions plus précises.

Par la suite, les spectrogrammes obtenus ont été utilisés pour la reconstruction de l'espace de phase, pour l'analyse récurrence et pour sa quantification par une mesure de complexité. L'analyse de récurrence permet de décrire et visualiser les dynamiques récurrentes d'un système et de découvrir des motifs structurels contenus dans les données. Ici, les diagrammes de récurrence ont été utilisés comme réécriture de grammaire pour transformer le signal original en une séquence symbolique, où chaque symbole représente un certain état du système. Trois mesures de complexité différentes sont alors calculées à partir de ces séquences symboliques afin de les utiliser comme éléments de classification.

Enfin, en combinant les caractéristiques obtenues avec l'analyse spectrale de puissance et avec l'analyse symbolique de récurrence, nous effectuons la classification des données en utilisant deux méthodes de classification : l'analyse discriminante linéaire et les machines à vecteurs de support. La classification a été effectuée sur des problèmes à deux classes, la distinction entre les signaux EEG pré- / post-incision, ainsi qu'entre les deux différents médicaments anesthésiques, desflurane et propofol.

Mots-clés: analyse des signaux multivariées, neuroscience computationnelle, analyse temps-fréquence, anesthésie, analyse de récurrence

Acknowledgments

I would like to take this chance to express my gratitude to people who helped me in my pursue of a PhD degree.

First of all, I would like to thank my thesis advisor, Axel Hutt, for this opportunity and for his wise guidance. I also would like to thank Peter beim Graben for his help throughout my work and his great ideas. Special thanks to Jamie Sleigh for his professional opinion and for providing the data.

I also want to express my gratitude to members of the jury: Frédéric Chazal, Gordon Pipa, and Anne Boyer for accepting to attend my defense. I appreciate their time and their feedback on my work.

I would also like to take time to thank Benoit for his help and support. Finally, I want to thank the people who supported me during these three years: my loving family, my friends, and all the great people I have met here in Nancy.

Contents

List of Figures

List of Tables

Glossary

AMI : Average Mutual Information
AWGN : Additive White Gaussian Noise

BIS : Bispectral Index

CNS : Central Nervous System
COI : Cone of Influence
CV : Cross-validation
CWT : Continuous Wavelet Transform

EEG : Electroencephalogram
ERP : Event-Related Potential

FFT : Fast Fourier Transform
FNN : False Nearest Neighbors
FT : Fourier Transform

GABA : γ-Aminobutyric Acid

ICA : Independent Component Analysis

KS : Kolmogorov-Smirnov

LDA : Linear Discriminant Analysis
LOOCV : Leave-one-out cross-validation
LZ : Lempel-Ziv

NMDA : N-methyl-D-aspartate

OU : Ornstein–Uhlenbeck

PCA : Principal Component Analysis
PSA : Power Spectral Analysis
PSD : Power Spectral Density
PSO : Particle Swarm Optimization

QDA : Quadratic Discriminant Analysis

RBF : Radial Basis Function
RP : Recurrence Plot
RQA : Recurrence Quantification Analysis

SRSA : Symbolic Recurrence Structure Analysis
STFT : Short-Time Fourier Transform
SVM : Support Vector Machines

TF : Time-Frequency
TFR : Time-Frequency Representation

WVD : Wigner-Ville Distribution

Résumé Étendu

Introduction

En 2004, il y a eu 26 millions de chirurgies aux états unis. La plupart des opérations chirurgicales sont conduites sous anesthésie, ce qui nécessite une observation constante afin d'éviter toute complications. Malgré la grande variété de moniteurs existant pour mesurer la concentration d'hypnotique, le niveau d'oxygène dans le sang, la fréquence cardiaque, l'électroencéphalogramme (EEG) et d'autres indices biologiques, aucune opération ne peut avoir lieu sans un anesthésiologiste qui peut prédire le comportement du patient pendant l'opération. L'anesthésiologiste doit être présent durant toute l'opération en observant le patient afin de prédire toute possibilité de niveau insuffisant d'anesthésie, d'éveil peropératoire ou de réponse aux stimuli nociceptifs. Cette tâche requiert beaucoup d'expérience et de pratique et malgré la présence des moniteurs lors de la chirurgie, la décision finale revient toujours à l'anesthésiste qui peut faire des prédictions basées sur des observations telles que le mouvement des yeux, les mouvements involontaires ou les grimaces. Même si le champ de l'anesthésiologie a considérablement avancé durant les dernières années, les mécanismes biologiques précis ne sont pas complètement découverts. Comprendre chaque facette de l'anesthésie : amnésie, immobilité, perte de conscience et analgésie, pourrai faciliter le travail des anesthésiologistes et assurer des opérations bien contrôlées. Ce travail contribue à ce champ de recherche en étudiant l'analgésie dans le cas d'anesthésie provoquée par deux différents types d'hypnotique : le propofol et le desflurane.

Motivation de la Thèse

Ce travail est motivé par les travaux de Sleigh et al. [133]. Dans cette publication, les auteurs analysent des signaux d'EEG obtenu sur des patients subissant une opération sous anesthésie générale. Plusieurs mesures sont proposées à partir de la puissance du spectre d'EEG afin d'identifier la différence entre deux conditions expérimentales : pré- et post-incision. Dans ce manuscrit, nous cherchons à étendre cette étude en proposant des mesures EEG différentes qui pourraient caractériser la réaction du patient à l'incision et qui étudient l'EEG comme un système dynamique. En plus d'extraire les marqueurs de la douleur, nous étudions aussi comment ces marqueurs varient en fonction de l'hypnotique utilisé lors de l'opération. Enfin, la dernière étape de ce travail est de classer les données en fonction des marqueurs développés afin de tester leur pertinence.

Objectifs de la Thèse

Cette étude se concentre sur l'analyse de signaux d'EEG enregistrés pendant une chirurgie sous anesthésie générale. Les données sont prélevées auprès de 115 patients. Pour chacun d'eux, deux signaux à un seul canal sont enregistrés, un avant le début de l'incision et l'autre pendant l'incision. Chaque signal dure 30 secondes. Les patients présentés dans cette étude ont reçus deux types d'hypnotiques : 60 patients ont reçus un hypnotique intraveineux (propofol) et 55 patients ont reçus un hypnotique volatile par inhalation (desflurane).

Notre tâche principale est de classifier les signaux de chaque condition expérimentales à savoir pré-/post-incision et desflurane/propofol. En plus de cette classification notre but est d'identifier les signatures de chaque condition, marqueurs qui possèdent un pouvoir discriminant plus ou moins fort. Pour cela, nous utilisons un outil très classique pour le traitement du signal biologique, l'analyse de densité spectrale de puissance. Nous faisons l'hypothèse que la puissance du spectre et plus particulièrement la puissance de certaines composante oscillantes : les rythmes δ et α, 1–4 Hz et 8–12 Hz, respectivement sont des marqueurs pertinents.

Nous proposons aussi une nouvelle méthode pour l'analyse de signaux oscillants, qui, au contraire de l'analyse de puissance, tire parti de la dynamique temporelle du signal [21, 22]. Cette méthode, basée sur la représentation symbolique récurrente du signal, permet d'extraire les états quasi stables de la dynamique du signal. Afin d'ajuster cette méthode à la nature oscillante de l'EEG nous proposons de combiner cette méthode avec la représentation temps fréquence du signal. Nous soutenons qu'une telle représentation symbolique peut servir de représentation à gros grains du signal alternative et peut être quantifiée par plusieurs mesures de complexités.

Biologie

Les mécanismes biologiques de l'anesthésie sont encore mal connus. Les médicaments hypnotiques ciblent les neurones et les synapses et diminuent leur activité. Les cibles de ces médicaments sont des récepteurs de neurotransmetteurs et des canaux ioniques. Les mécanismes d'inhibitions peuvent être direct ou indirect. Les mécanismes directs connus impliquent le récepteur $\mathrm{GABA_A}$ impliqué dans les mécanismes d'inhibition des potentiels synaptiques ou le récepteur NAMDA. Les sites de liaison du propofol aux récepteurs $\mathrm{GABA_A}$ qui provoquent l'activation du récepteur (et donc favorisent les flux inhibiteurs au niveau de la synapse) ont été caractérisés. Il est cependant rare que les cibles des hypnotiques soient si bien définies, d'autant plus qu'il est certain que le nombre de cible est multiple.

Il est donc difficile de prévoir le comportement d'un patient sous anesthésie ou même de prévoir la quantité d'hypnotique nécessaire pour provoquer la perte de conscience. L'expertise de l'anesthésiste avec l'aide de son moniteur sont donc nécessaires pour chaque chirurgie.

Un des signaux utilisable par l'anesthésiste afin de surveiller le patient sont les signaux enregistrés à l'aide d'électrodes sur le front du patient. Ces signaux d'EEG sont soit étudiés directement à l'oeuil nu par l'anesthésiste soit (le plus souvent) prétraités pour en extraire une information simple et directement compréhensible par le praticien. Cette information est souvent représentée par un indice variant de 0 à 100. Un des indices le plus populaire est l'indice BIS (pour index bispectral) qui représente le niveau d'inconscience du patient. L'EEG étant par nature un signal très bruité et très variable de patient en patient, cet indice seul n'est en général pas suffisant pour déterminer si oui ou non la quantité d'hypnotique injecté est suffisante.

Même si les mécanismes de la douleur et de son soulagement à l'aide de médicaments analgésiques comme les opioïdes sont mieux compris, la douleur ressentie par le patient lors de l'anesthésie est difficile à évaluer. Le patient étant hypnotisé et paralysé, il ne ressent pas la douleur. Cependant, des signes extérieurs montrent que certains mécanismes réflexes sont toujours présents car le patient peut avoir des mouvements involontaires (grimaces) et les signes hémodynamiques (pression artérielle) changent lors du début de la chirurgie (incision). L'anesthésiste peut alors réguler l'analgésique afin d'éviter un stress trop grand chez le patient afin de faciliter la chirurgie et de minimiser les risques de douleurs post-opératoires.

Nous nous intéressons dans cette thèse à l'intérêt que peut avoir l'étude de l'EEG lors de la chirurgie afin de mieux comprendre les phénomènes corticaux liés à la douleur d'une part et aider l'anesthésiste à doser l'analgésique avant et après incision d'autre part. Les données que nous utilisons sont issues d'une étude médicale menée lors de condition de chirurgie réelles et ont déjà été étudiées dans l'optique d'améliorer l'analgésie du patient. Nous proposons ici une étude plus systématique des marqueurs de la douleur et de l'anesthésie en général.

Données

Données Expérimentales

Les données utilisées dans ce travail sont enregistrées lors d'une chirurgie sous anesthésie générale avec de nombreux patients. Le jeu de données [133] contient les données temporelles de 115 patients. Les patients ont un âge compris entre 18 et 50 ans, ont le statu de la société américaine d'anesthésiologie I–III et subissent une opération non indispensable et non cardiaque sous anesthésie générale. Les types de chirurgie vont de la chirurgie abdominale et orthopédique jusqu'à la laparoscopie. L'anesthésie est induite par 1 à 2 μg/kg de propofol, fentanyl et relaxant musculaire. Le maintient de l'anesthésie est assuré par du desflurane pour 55 patients et par du propofol pour 60 patients. L'anesthésie est contrôlée par l'index bispectral (BIS : Aspect Medical Systems Inc.) ciblé pour un indice entre 40 et 55.

Le jeu de données est composé de deux signaux EEG par patient enregistrés 2 min avant et 2 min après l'incision cutanée et le début de la chirurgie. Les signaux d'EEG sont enregistrés sur le lobe frontal avec deux électrodes FP2 et FT7. Chaque signal dure 30 secondes avec un taux d'échantillonnage effectif de 128 Hz. Le signal est ensuite filtré numériquement entre 1 et 41 Hz avec un filtre de Butterworth d'ordre 9 afin d'enlever le bruit de l'alimentation.

Données Artificielles

Afin de valider les méthodes proposées dans cette étude, nous les appliquons d'abord à des données artificielles : oscillations passagères et solution du système chaotique de Lorenz.

Le signal d'oscillation transitoire est une superposition linéaire de 3 signaux qui ont des oscillations passagères bruitées à une fréquence spécifique [123]. Ces fréquences sont 1.0 Hz, 2.25 Hz et 6.3 Hz. La fréquence d'échantillonnage est de 50 Hz et le signal dure 70 s.

Le système de Lorenz est un système bien étudié d'équation différentielles à 3 dimensions qui exhibe un comportement chaotique. La solution du système de Lorenz [101] possède 2 ailes qui sont approchées avec une séquence non prévisible. Ces ailes sont caractéristiques des trajectoires métastable.

Méthodes

Ce travail est composé de deux parties principales : la construction de marqueurs et classification. L'étape de construction de marqueur est basée sur l'analyse temps-fréquence et l'analyse des structures récurrentes.

L'analyse de puissance spectrale consiste à calculer différentes mesures de puissance. La première étape de cette analyse est de calculer la puissance de la densité spectrale des signaux. Cela a été fait en utilisant le périodogramme modifié de Welch. Après avoir obtenu les densité de puissance spectrales (DPS) des signaux EEG nous nous intéressons à la puissance totale contenue dans les signaux et la puissance contenue dans les bandes α et δ [133]. Ces rythmes EEG sont en effet caractéristiques de la période de maintient de l'anesthésie : lorsque l'anesthésie progresse, on observe une croissance des activités à haute fréquence et à basse amplitude (α) et basse fréquence et haute amplitude (δ), particulièrement sur les lobes frontaux [130].

La puissance contenue dans certaine bandes peut être obtenue depuis la densité de la puissance spectrale en l'intégrant sur une plage donné. Le premier marqueur est la puissance totale contenue dans le signal. Pour cela, il faut intégrer l'estimation de DPS du signal de 0 à la fréquence de Nyquist. Cependant, puisque le signal a été auparavant filtré avec un filtre passe bande, nous intégrons uniquement entre 1 et 41 Hz.

Les deux marqueurs suivants sont la puissance des oscillation δ et α. Pour obtenir la puissance de la bande *delta*, il faut intégrer l'estimateur de PSD dans la fenêtre de fréquence 1–4 Hz. Cependant, avant d'obtenir les valeurs de la puissance de la bande, il faut séparer l'activité de bande étroite du bruit de la bande sous-jacente qui est caractérisé par une progression $1/f$ du spectre de puissance [40]. Dans ce manuscrit, nous cherchons à ajuster un modèle de l'activité d'arrière plan à la DPS de l'EEG afin de le soustraire. En faisant cela on obtient une estimation de la DPS qui peut être intégrée afin d'obtenir la puissance contenue dans les pics α et δ de l'estimation de DPS sans l'activité de font.

Le spectre de puissance est modélisé avec le spectre de puissance d'un processus de Ornstein-Uhlenbeck :

$$P_{OU}(f) = \frac{2D}{\gamma^2 + (2\pi f)^2} \ , \tag{1}$$

$2D$ et γ^2 sont deux paramètres de contrôle. Ces paramètres ainsi que la puissance des bandes α et δ et la puissance totale forment les marqueurs caractérisant les propriétés stationnaires du signal.

L'analyse des structures symboliques récurrentes consiste d'abord à calculer la matrice de récurrence du signal et ensuite à utiliser cette matrice en tant que grammaire pour encoder le signal en une chaîne de symboles où tous les états récurrents sont marqués avec le même symbole.

Cette méthode est basée sur le concept des diagrammes de récurrence (DR). Cette méthode relativement simple permet de visualiser les trajectoires de phases multi-dimensionnelles d'un système non linéaire sur un plan bi-dimensionnel. Les DR peuvent être obtenus en traçant la matrice de récurrence :

$$R_{ij} = \Theta\left(\varepsilon - ||\mathbf{x}_i - \mathbf{x}_j||\right), \qquad i, j = 1, 2, \ldots, N \ . \tag{2}$$

Ici, $\mathbf{x}_i \in \mathbb{R}^d$ est l'état du système complexe dans l'espace des phases de dimension d à l'instance temporelle i, $||\cdot||$ dénote une métrique, Θ est la fonction Heaviside, et ε est la distance seuil.

La matrice de récurrence peut ensuite être utilisée comme un ensemble de règles de réécriture afin de représenter une série temporelle comme une séquence de symboles où chaque symbole marque une activité distincte du système. Les chaînes de caractères sont d'abord initialisée avec les indices temporels discrets comme symbole et ensuite réécrites en utilisant les valeurs de la matrice récurrente. Si deux symboles sont récurrents, il fusionnent.

La valeur de ε est cruciale pour la méthode car elle affecte beaucoup le résultat de la matrice de récurrence et donc de la chaîne de caractères symbolique. Il existe plusieurs estimateurs de la valeur de seuil optimale [106], cependant, la plupart d'entre eux sont des règles empiriques. Nous utilisons ici un estimateur de ε basé sur des hypothèses sur la dynamique du système comme dans [21]. L'idée principale de cette méthode est de choisir le ε optimal de façon à maximiser une *fonction utilitaire* qui représente cette hypothèse. Dans ce travail, nous supposons que le système passe autant de temps dans chaque domaine récurent que dans les états transitoires.

Cette hypothèse peut être résumée par le principe d'entropie maximale, la fonction utilitaire sera donc l'entropie symbolique.

$$\varepsilon^* = \arg\max_{\varepsilon} h(\varepsilon) \ , \tag{3}$$

où $h(\varepsilon)$ est l'entropie normalisée :

$$h(\varepsilon) = -\frac{\sum_{k=1}^{M(\varepsilon)} p_k \log p_k}{M(\varepsilon)} \ . \tag{4}$$

où p_k est la fréquence relative du symbole $s_i = k$ et $M(\varepsilon)$ est la cardinalité de l'alphabet symbolique.

Nous présentons une méthode pour estimer les paramètres optimaux de la méthode à partir des données. Nous présentons aussi une nouvelle méthode pour la reconstruction de l'espace des phases basée sur la représentation en temps fréquence du signal qui est plus adaptée aux signaux oscillants.

La méthode de reconstruction proposée a été appliquée avec différents types de représentation temps-fréquence : spectrogramme, spectrogramme réaffecté, et scalogramme avec des fréquences à intervalle linéaire et non-linéaire. Finalement, l'efficacité de cette méthode est comparée à des méthodes standard de plongement.

Malgré un meilleur taux de détection des domaines récurrents des données synthétiques avec le spectrogramme réaffecté, il apparait que la meilleure méthode pour l'EEG est le scalogramme avec les fréquences à intervalle linéaire. Une des raisons est la propriété intrinsèque qu'ont les transformations d'ondes de mettre à l'échelle la puissance avec la fréquence exposant donc le même comportement $1/f$ que l'EEG.

Enfin, afin de construire les marqueurs il est nécessaire d'obtenir une mesure quantitative de la représentation symbolique des signaux. Dans ce but, les bon candidat sont les mesures de complexité. Nous utilisons 3 mesures différentes comme marqueur. La première est la taille de l'alphabet d'une chaîne de caractère. La deuxième est le nombre de mot extraits de la chaîne de caractères symboliques où un mot est un groupe de symbole consécutif de même valeur. Le troisième marqueur est la complexité de Lempel-Ziv [94]. Enfin, la distance de seuil optimale ε^* et la valeur correspondante d'entropie symbolique $h(\varepsilon)$ sont rajoutés dans la liste des marqueurs de façons à obtenir un total de 5 marqueurs.

La classification est faite avec deux classifieurs. Le premier est l'analyse linéaire discriminante qui produit des frontières de décision linéaires. Le second classifieur est une machine à vecteurs de supports avec un noyaux RBF (radial basis function) qui permet l'apprentissage de frontières de classe non-linéaires complexes.

Ces travaux ont permis de classer avec succès environ 60% des signaux pré- et post-incision et environ 70% des signaux desflurane et propofol. Il apparaît que la classification avec les marqueurs extraits de l'EEG en utilisant l'analyse de la puissance spectrale surclasse légèrement la classification utilisant les marqueurs extraits de l'analyse des symboles récurrents.

Ce chapitre est conclu avec une procédure hiérarchisation des marqueurs afin d'identifier les marqueurs les plus discriminants. Il apparaît que les trois marqueurs les plus significatifs font partie des deux groupes de marqueurs et que l'utilisation de seulement trois marqueurs est généralement efficace pour réduire le surapprentissage. La classification finale est faite avec les trois premiers marqueurs en utilisant une machine à vecteur de support avec un noyaux RBF.

Conclusion

L'anesthésie générale est un outil indispensable pour les pratiques cliniques modernes. Une anesthésie bien ajustée et contrôlée garantie une chirurgie sans encombres, un temps de réveil cours et moins de risques de douleurs post-opératoire et moins de stress pour le patient. Les aspects principaux de l'anesthésie sont l'amnésie, la paralysie, l'hypnose et l'analgésie. Plus de détails sur les mécanismes sous-jacents de chacune de ces composantes permettra d'améliorer les conditions des opérations. Notre travail contribue à ce champ de recherche en étudiants certains aspects de l'analgésie pendant l'anesthésie.

Cette étude porte sur l'EEG enregistrée sur des patients subissant une opération sous anesthésie générale. L'objectif de cette thèse était de classifier les signaux enregistrés pendant les phases pré-incision et post-incision de la chirurgie. Pendant la chirurgie, un groupe a reçu du propofol alors que l'autre a reçu du desflurane. Ces deux agents hypnotiques ont des propriétés différentes ce qui nous a permis de pouvoir classer les EEG des patients de chacun des groupes.

Perspectives

Nous ne sommes pas parvenus à un bon taux de classification pour différencier les signaux pré- et post-incision. Cela peut indiquer que la réponse des patients n'est pas visible sur l'EEG. L'une des améliorations possible est de changer le processus d'acquisition. Les signaux considérés sont enregistrés à 4 min d'intervalle et l'activité post-incision est enregistrée 2 min après l'incision. L'analyse pourrai être améliorée si les données étaient enregistrées en continu. Avec une telle approche l'incision peut être traitée comme un stimulus dans une analyse des potentiels évoqués. L'encodage symbolique basé sur les récurrences a déjà été appliqué avec succès à l'analyse de potentiels évoqués. De plus, augmenter le taux d'acquisition pourrait simplifier la tâche puisque les événements corticaux peuvent être de durée très brève (de l'ordre de la milliseconde) et avec des données sous échantillonnées de tel événements peuvent se perdre.

General Introduction

In 2014, the total number of surgeries performed was estimated to be as high as 26 million, only in the United States [1]. Most surgeries are conducted under general anesthesia, which requires constant monitoring in order to avoid any kind of complications. Despite the number of existing monitors, measuring anesthetic concentration, oxygen level in the blood, heart rate, electroencephalogram, and other biological indices, none of the surgeries can be conducted without an anesthesiologist, who can predict patients behavior during the operation. The anesthesiologist have to be present throughout the surgery carefully observing a patient and predicting any possibility of insufficient level of anesthesia, intraoperative awareness or response to nociceptive stimuli. This task requires a lot of expertise and practice, and despite of numerous monitors available during the surgery, the final decision remains with the anesthesiologist, who can make predictions based on such subjective observations as eye movement, involuntarily movements or grimaces.

Motivation of the Thesis

Since the first surgical anesthesia in 1846, followed by the creation of the first professional organization in 1893, up to present times the field of anesthesiology has advanced dramatically, covering such fields as pre- and postoperative monitoring of patients, available equipment, formal protocols and prescriptions. However, precise biological mechanisms underlying general anesthesia remain unknown. The situation is further complicated by the presence of a large number of various anesthetics, each of which produces their own effect. Understanding every anesthesia component: amnesia, immobility, loss of consciousness and analgesia, can facilitate the work of anesthesiologist and ensure well-controlled medical surgery. Our work contributes to the field by studying the analgesia effect of anesthesia provided by two different anesthetic drugs: propofol and desflurane.

This thesis is motivated by a previous work of Sleigh *et al.* [133]. In this paper the authors analyzed EEG dataset obtained from patients undergoing surgery under general anesthesia. Here, authors propose several measures of EEG spectral power in order to identify difference between two experimental conditions: pre- and post-incision. In this manuscript, we aim to extend the study by proposing additional EEG measures, which might characterize patient's reaction to nociception and which study EEG from dynamical system's point of view. In addition to extracting features of pain, we also study how these features differ depending on the anesthetic drug used during the surgery. This topic was previously outlined in [95]. Finally, the last step of this work is to perform classification of the data with given features in order to asses their effectiveness.

---

[1] The data taken from the American Hospital Association: `http://www.aha.org/research/reports/tw/chartbook/ch3.shtml`

Objective of the Thesis

Our study focuses on the analysis of EEG data recorded during surgery under general anesthesia. The data consist of 115 patients, for each patient two single-channel EEG signals were obtained: one before the start of the incision and another one during the incision. Each signal is 30 seconds long. Patients presented in the current work received two types of anesthetic agents: 60 patients received an intravenous agent (the propofol) and 55 patients received a volatile (inhalational) anesthetic (the desflurane).

Our main task is to perform classification of signals from distinct experimental conditions, namely pre-/post-incision and desflurane/propofol. In addition to classification, our goal is to identify signatures of each of the conditions, features which posses the most discriminative power. For that we apply a common tool in biological signal processing: power spectral analysis. We hypothesize that power information, specifically power of certain oscillatory components: δ- and α-rhythms (1–4 Hz and 8–12 Hz, respectively) are relevant features.

We also propose a novel method for the analysis of oscillatory signals, which, in contrast to power analysis, takes advantage of temporal dynamics of the signal [21, 22]. This method, based on recurrence symbolic representation of the signal, allows to extract quasi-stable states in the signal's dynamics. In order to adjust this method to oscillatory nature of the EEG, we propose to combine it with time-frequency representation of the signal. We argue that such symbolic representation can serve as an alternative coarse-grained representation of the signal and can be quantified by various complexity measures.

Structure of the Thesis

Chapter 1 presents a general overview of biological mechanisms of general anesthesia and its analgesia effect. It also covers various techniques used in anesthesia monitoring with a special focus on EEG-based methods. It discusses differences in anesthesia maintained with propofol and desflurane. The previous work related to our study is also presented in this chapter.

Chapter 2 is dedicated to feature construction phase of our work. It is divided into two main parts based on the method used for generating features. First part describes power spectral analysis. The second part focuses on recurrence symbolic analysis and the steps involved in it, such as evaluation of optimal symbolic encoding of the signal, different methods of phase space reconstruction (four time-frequency representations and delayed embeddings), and complexity measures. Finally, the obtained results are discussed at the end of the chapter.

Chapter 3 focuses on the final step of our work, namely the classification of obtained features. The chapter describes two classifiers used in this manuscript: linear discriminant analysis and support-vector machines. The chapter is concluded with feature ranking, a procedure carried out in order to identify the features which play the most significant role in the process of classification. The chapter also presents the interpretation of the obtained results.

1

Analgesia Effect of Anesthesia

Contents

General anesthesia is a reversible state of analgesia, muscle relaxation and hypnosis induced in order to perform operation [8]. Modern anesthesia originates from 1846 when William Thomas Green Morton used diethyl ether to remove a neck tumor. Anesthesia is now induced via an intravenous anesthetic (in our case propofol) followed by the administration of a muscle relaxant. The trachea is then intubated to maintain lung ventilation. Ventilation is done with oxygen, nitrous oxide and volatile anesthetic (like desflurane or isoflurane) to maintain anesthesia. Anesthesia maintenance is necessary to avoid spontaneous recovering which can happen 5 to 10 minutes after the intravenous injection [131]. Intravenous maintenance is also possible by continuing the injection of propofol. During the surgery a wide variety of drugs are available to the anesthesiologist or the surgeon to stabilize the patient. One of them is the opioid derivatives used to improve the analgesic effect of the anesthetic drug. This chapter discuss mechanisms of anesthesia and various changes in a body caused by it.

1.1 Mechanisms of Anesthesia

The detailed mechanisms of anesthesia are still not well understood. The anesthetic agent leads to a general decrease of activity of neurons in the brain and in the spinal cord. This decrease of activity is due to interactions with ion channels regulating synaptic transmission or membrane potential [8].

For instance, one of the most reported effects of anesthetic agents is enhancement of $\mathrm{GABA_A}$ (*gamma*-Aminobutyric acid type A) receptors which are the receptors of the main inhibitory neurotransmitter of the central nervous system (CNS). NMDA (N-methyl-D-aspartate) receptors channels are also one of the main suspected target of anesthetic drugs [11]. The anesthetic drug ketamine was shown to block NMDA channels and every anesthetic was shown to inhibit NMDA channels directly or indirectly. But it was shown that many more receptors and ion channels could be affected by these drugs in a direct or indirect way [100]. Another mechanisms was suggested by the hydrophobicity of the anesthetic agents. They are able to solubilize in the membrane lipidic bilayer and change its properties (hypothesis involve the lateral pressure, or interacting with lipid rafts). This change in the lipidic membrane would slightly change the conformation of membrane's proteins and channels thus diminishing the activity of the neurons.

At a macroscopic level, the decrease of neural activity leads to weaker correlations and synchrony in the neural populations. This can be noticed by the switch from fast frequency oscillations in gamma frequency (30–70 Hz) to slower theta and delta oscillations (4–7 Hz and 1–2 Hz respectively). As consciousness is a state which is not really explained by neurobiologists, the loss of it during anesthesia is not well understood. Main theories include neurone-specific theory and process-specific theory. In the neuron-specific view, the consciousness is believed to depend on the activity of a specific set or populations of neurons while in the process theory, it is believed that consciousness emerges from the synergies between different process in the cortex [11].

Propofol is one of the most commonly used intravenous anesthetic agent because of its short acting time. Its hypnotic effect is rapid and its redistribution time is short which gives a fast recovery for the patient. However, it has the disadvantage to decrease blood pressure and to have a bad analgesic effect. It is, thus, often administrated in conjunction with an opioid. Propofol was shown to affect the $\mathrm{GABA_A}$ receptor and the glutamate action at the NMDA receptors. It was also shown to affect both CNS and spinal cord [100].

Desflurane is gradually replacing isoflurane for the maintenance of general anesthesia. It is a positive modulator of $\mathrm{GABA_A}$ and glycine receptors. It is less soluble in the blood than other volatile anesthetics. This poor solubility permit rapid pulmonary uptake and elimination which is beneficial for efficient control of the depth of the anesthesia as well as the rapid emergence of a patient.

1.2 Analgesia

In our work, we focus particularly on analgesia effect of anesthesia. In this section we will focus on the mechanisms of pain and how the pain is controlled during anesthesia.

1.2.1 The Pathways of Pain

In normal conditions, noxious or pre-noxious stimuli are detected by dedicated sensors on the skin, in the bones, in the vascular system, etc. This sensors are called nociceptors, they are usually classified by the speed of their fiber and the type of stimuli they can detect. There are two types of fiber connecting the sensory neuron to the spinal chord. The A-fibers have myelinated axon and conduct impulses at the speed of 2 to 30 $\mathrm{ms^{-1}}$. The C-fibers are not myelinated and have a speed of 2 ms^{-1}. A-fibers react on sharp pain like mechanical stimuli, whereas C-fibers react on thermal stimuli [33].

Their fibers are connected to inter-neurons in the spinal chord (on the dorsal horns) and, thus, their activation is projected to the motor neurons of the spinal chord and to the central nervous system. In CNS, complex pathways involving several brain areas produce the unpleasant sensation of pain with affective and motor responses. Central nervous system is integrating information about the stimuli. Its role is to identify the source, the intensity and the nature of the noxious stimulus in order to trigger an appropriate behavioral response. Thalamus, hypothalamus, limbic system, cerebral cortex, cingulate cortex, basal ganglia and cerebellum are all involved in pain processing and reaction [100].

1.2.2 Analgesia During Anesthesia

If anesthetics remove the consciousness and memory of pain, they do not completely block the pain transmission pathways toward the central nervous system. Hemodynamic monitoring show that the body is still reacting to pain during noxious stimulation like incision or intubation. These operations create a stress response characterized by an increase of heart rate and blood pressure, pupil widening and contraction of skin vessels [100].

All the components involved in noxious stimuli processing have a complex behavior subject to plasticity. This plasticity is the source of post-operative acute or chronic pain if the analgesia is not handled with care. For instance, acute pain can come from hyperalgesia, an exaggerated pain response at the site of the injury (primary hyperalgesia) or outside the site (secondary hyperalgesia) [28]. Even though the mechanisms of hyperalgesia are well understood, their prevention during anesthesia is not consensual [105].

Nevertheless analgesia control is mandatory to reduce the stress and perturbations during surgery and to reduce chances of post operative pain. The removal of stress response is generally induced by opioids during anesthesia. In our case the fentanyl was used even though fentanyl does not block every hemodynamic and hormonal response to surgery [116].

Analgesia is difficult to asses during anesthesia. One way is to check cardiovascular reaction to noxious stimuli, another way is to use the analgesic effect as reference. Indeed, another advantage of opioid is that they reduce the amount of anesthetic drug to use to produce hypnosis [100]. Often the anesthesiologist has to rely on their judgment based on their knowledge and experience to decide if yes or no the patients needs more analgesic.

In the next section we will investigate more precisely the monitoring tools available to the anesthesiologist and we will investigate the expected benefits of EEG monitoring for anesthesia and analgesia.

1.3 Anesthesia and EEG

The main tool available for the anesthesiologist is the monitor which uses several biological signals from the patient to give valuable information about their state.

1.3.1 Anesthesia Monitoring

The goal of anesthesia monitoring is twofold. First, it helps the anesthesiologist to asses the depth of anesthesia. And second, during the operation, monitoring helps to detect changes in patient's condition in order to adapt the drug usage or to adapt the surgery. The anesthesia depth is classified according to the awareness of the patient with 6 level [58]:

1. conscious awareness with complain of pain perception;
2. conscious awareness with explicit recall but without pain;
3. conscious awareness (ability to respond to simple verbal commands);
4. subconscious with explicit recall and pain;
5. subconscious with implicit memory of operation;
6. unconscious.

The anesthesiologist's goal is to reach the last level relying on different monitoring strategies. The most basic monitoring used in anesthesia is the hemodynamic monitoring using arterial blood pressure, central venous pressure and cardiac output.

The other means for monitoring patient's response to stimuli are the evoked potentials. Evoked potentials are electrical potentials produced in response to a stimulation. It is possible to monitor them using electrodes on different point on the sensory-motor pathways in spinal chord, peripheral nerve or scalp EEG. As the amplitude of evoked potentials is small compared to standard EEG it is necessary to compute the average. Evoked potentials are usually used to monitor damages during surgeries involving spinal chord [89].

The last tool available for monitoring patient's state is EEG which is the main focus of our work. During surgery, the electroencephalogram is captured with few electrodes located on the forehead. These electrodes are able to capture the sum of the variations of the potentials of the cortical neurons. EEG is mainly use to asses the depth of anesthesia in order to help the dosage of anesthetic.

The main indexes used for anesthesia depth monitoring are the Bispectral index (BIS) and entropy. The BIS index is derived from a proprietary algorithm which process the signal from one single forehead electrode [131]. The index is computed using three parameters: the beta ratio (the slow wave content of the signal derived from fast Fourier transform), the bi-coherence of all frequency pairs, and the amount of burst suppression patterns in the filtered EEG [100].

Another possible measure of anesthesia depth can be entropy index, which is a measure of the order in the EEG signal. Time and space frequencies are separated using a fast Fourier transform and an entropy measure is applied to each integer frequency [100].

Other indexes are provided by different monitor constructors, for instance: Danmeter cerebral state index, Narcotrend, Sedline, and Snap II [74].

In addition to the depth of anesthesia, other things could be monitored by EEG. It is still in the domain of research because it often requires more electrodes which are not practical during surgery. For instance using a 64-channel EEG a more precise classification can be established. In their experiment Purdon *et al.* manged to characterize the loss of consciousness by an increase of frequency and loss of spatially coherent occipital alpha oscillation and appearance of frontal spatially coherent alpha oscillation [120].

1.3.2 Drugs Effect on EEG

Generally the main reproach that doctors have against EEG monitoring is that EEG signals and EEG-based indexes are very dependent on the anesthetic drug, on the analgesic drug and, in general, on many drugs that can be given to the patient during the surgery. For instance, in [89] the authors report the effect of 15 drugs on the somatosensory evoked potentials and report increase or decrease of latency and amplitude for each of them.

For anesthetic drugs the EEG patterns are quite different as well. For instance, many studies tried to explain the spindle patterns found in propofol anesthesia [45] which might be enhanced by opioids [95].

As our dataset contains two type of anesthetics (propofol and desflurane), knowing the differences induced in EEG pattern by these drugs can be useful.

Spindle activity was characterized in the stage II of slow wave sleep as waxing and waning oscillations from 7 to 14 Hz appearing with a period of 1–3 seconds. The spindles contribute to the alpha activity during anesthesia. The spindle activity is generally high during anesthesia but not necessarily, depending on the patients [62]. Propofol induce stronger spindle-like waves than desflurane [95]. Propofol gives frontal alpha oscillations with peak frequency around 11 Hz and power 1.1 dB [6]. This might be the result of the GABA_A receptor agonist effect of propofol or the antinociceptive effect of the opioids commonly added to propofol anesthesia [95].

Burst suppression patterns are commonly found in deep anesthesia. It is an alternation of high activity (burst) with very low activity of the brain (suppression). The frequency of burst is a useful marker of anesthesia depth. In [7] the authors compare the effect on several anesthetic drugs on the burst suppression patterns and find that isoflurane provokes significantly longer burst duration than propofol (1.4 s against 0.4 s). The amplitude of the activity peaks is higher as well. In [68] the authors show that isoflurane and desflurane have the same effect on burst suppression patterns. They, however, insist on the variability with the age and baseline delta EEG of the patients.

1.3.3 Incision Effect on EEG

Several works have studied the effect of incision or any other nociceptive stimulus on an anesthetized patients or animals using EEG. It is of primary interest for anesthesia research, as it might lead to well established methods to monitor the patient pain during anesthesia together with good marker of unconsciousness, which would lead to a general improvement of the surgery conditions for patients and savings in anesthetic and analgesic drugs [120].

Classification of patient reaction on noxious stimuli during a surgery is still not clear as the underlying mechanism are not well understood. There is still no consensus for reliable markers of pain during anesthesia as it seems to depend a lot on the patient: age, condition and natural variability of metabolism [62].

Nevertheless two different kind of responses have been observed by experimental setups with both patients and animals. In [84] 46 patients were put under isoflurane anesthesia and monitored with 17 scalp electrodes according to the 10–20 location system. The recording started 6 min before incision and stopped 14 min after anesthesia. To study the influence of the level of the anesthesia on the response, the patients were separated in two groups one with 0.6% of isoflurane and one with 1.2% of isoflurane. In each group half of the patients were incised and the other half were not. The results are showing a strong increase in δ-power and a decrease in α-power. The increase of δ-power is stronger with low concentration group.

Theses results are interesting because they are in contradiction with what was traditionally found in nociceptive stimuli influence on EEG during anesthesia. Indeed, previous studies reported arousal pattern during sensory stimulation similar to the emergence from anesthesia: an increase in the α-power and lower amplitude characteristic of EEG resynchronization. This study show an opposite effect, called reverse arousal or paradoxical arousal which looks like a deeper anesthesia state with slow waves of high amplitude.

These finding have been reproduced on sheep, cats and humans [113] and with different kinds of anesthetic [133]. It seems that these two patterns depend on the level of anesthesia but are still present in deep anesthesia.

In [83] the effect of extradural analgesia on the EEG response to incision was studied. The signals were recorded during 30 minutes after incision. They have shown that if analgesia is strong enough, the hemodynamic variables as well as the EEG does not vary during incision. Without analgesia, EEG is displaying the paradoxical arousal pattern and arterial pressure is rising. These two responses are predominant in the frontal lobe, it is thus expected that single or bi-electrode monitoring on the forehead would be able to capture it.

In [63], the EEG bicoherence was shown to be a good indicator of the analgesia of the patient. The data were recorded before incision, 5min after incision, and after administration of fentanyl intravenously (opioid). There were two groups: one with isoflurane and one with sevoflurane. Two index derived from bicoherence were shown to be good markers of incision with both anesthetic. The pBIC-low is the maximal value of bicoherence between 2 to 6 Hz and pBIC-high, from 7 to 13 Hz. Both these markers are decreasing during incision and return to pre-incisional values after the fentanyl administration.

On the contrary, the BIS index and SEF (spectral edge frequency) were not able to separate the classes due to the difference between arousal and paradoxical arousal patterns. Hagihira has shown similar results with isoflurane and even reported some patients displaying a mixture of paradoxical arousal and arousal [62].

With the same data set as the present work, Sleigh *et al.* [133] were able to design a marker of post-incision with the spindle amplitudes in the alpha range. They use $1/f$ regression on power spectrum, a method that we will discuss in the following chapters. Other markers like BIS, δ-peak and others were not able to separate pre- and post-incision with significance.

1.4 Conclusion

Anesthesia is a complex phenomena which is very far from being understood. Experiments on patients or animals models are now the only way to improve medical practice of the anesthesiologist. EEG is a promising tool for anesthesia monitoring. It is already successfully used for monitoring the depth of anesthesia but anesthesiologist would like to monitor analgesia level as well. Generally bad analgesia is noticed by the anesthesiologist only when noxious-stimulation-like incision or intubation are starting, a rise in arterial pressure, for instance. It is challenging but interesting task to predict an inappropriate level of analgesia with a simple EEG measure.

One of the main obstacles is presence of the two different kinds of EEG patterns in response to noxious stimuli. The first one is an increase in high frequency power (classical arousal) and the second is an increase in low frequency components (paradoxical arousal). These effects are diametrically opposite which does not facilitate the task. However, some interesting leads were studied such as the spindle amplitude [133] or the bicoherence peak in low or high frequencies [63].

Nonetheless, EEG is a difficult tool to use because of its variability. Many artifacts can compromise the signal [24]: noise from electrical components, bad impedance of the electrodes, movement of the patients or liquid spilling during the surgery. Adjunctive drugs used with the anesthetic can also have effects on EEG, effects which are not well studied. EEG can also be altered by factors other than various drug effects, which are linked with the patient state such as hypothermia and ischemia [74]. Patient variability can also be genetic, as for instance, 5 to 10 percent of the population genetically have low-amplitude EEG [110]. Thus, it is necessary to have powerful mathematical tools able to extract useful information to improve decision making during surgery.

2

Feature construction

Contents

In this chapter we discuss the analysis methods used in the course of the current study. The proposed methods are used as a feature construction tool for further classification. Our study focuses on the analysis of incisional EEG data recorded during general anesthesia, we propose methods reflecting the specificity of the signals and also demonstrate their efficiency on synthetic data. The first group of features is acquired with power spectral analysis, which is an indispensable tool in signal processing. With this approach, we obtain five features capturing power of oscillatory components in the signal and the background activity of the EEG. However, the spectral analysis focuses on stationary nature of the signal, we extend feature construction process with methods from dynamical system analysis. More specifically, we utilize modification of recurrence analysis: recurrence-based symbolic encodings. It transforms a signal into a symbolic string, and the features obtained on this stage comprise various complexity measures of the string. We propose a modification to the method to make it more adjusted to the analysis of oscillatory signal by utilizing time-frequency embedding. Before proceeding to the analysis of the real data, we demonstrate the efficiency of this modification on synthetic data.

2.1 Data Description

Current work focuses on the analysis of EEG signals obtained during general anesthesia. General anesthesia is an everyday practice in hospital which involves certain risks. In order to minimize the risks and improve the supervision of patients during surgery, more and more hospitals monitor brain activity by registering EEG on the scalp of the patient. Typically, a single EEG electrode is fixed on the forehead of the patient. The methods proposed in this manuscript are novice, for that reason before we can apply them to the EEG data, we have to test them on the synthetic data. Synthetic data is presented with two types of signals: transient oscillations and Lorenz system, which are described below.

2.1.1 Experimental Data

The data investigated in the present work was registered during surgery conducted under general anesthesia in a large set of patients. The dataset [133] comprises time series of 115 patients. The patients were aged 18–50 years, were American Society of Anesthesiologists' physical status I–III, undergoing elective noncardiac surgery under relaxant general anesthesia. The performed surgery varied between patients, including abdominal, orthopedic, and laparoscopic surgery. Anesthesia was induced with 1–2 μg/kg propofol, fentanyl, and a muscle relaxant. The maintenance hypnotic was introduced and included either *desflurane* (number of patients: 55) and *propofol* (number of patients: 60). Anesthesia was monitored with bispectral index (BIS: Aspect Medical Systems Inc.) and aimed for BIS 40–55.

The dataset includes two EEG recordings for each patient, registered 2 minutes before and 2 minutes after skin incision in the beginning of surgery. The incision took place at least 15 minutes after anesthesia induction. The experimental setup is schematically illustrated on the Fig. 2.1. The EEG signals were registered using BIS Aspect XP monitor with the proprietary electrode strip on the standard recommended pre-frontal montage (electrodes FP2-FT7). Each signal is 30-second long with an effective sampling rate of 128 Hz. The signal was then digitally filtered between 1 and 41 Hz using a ninth order Butterworth filter in order to remove power-line noise.

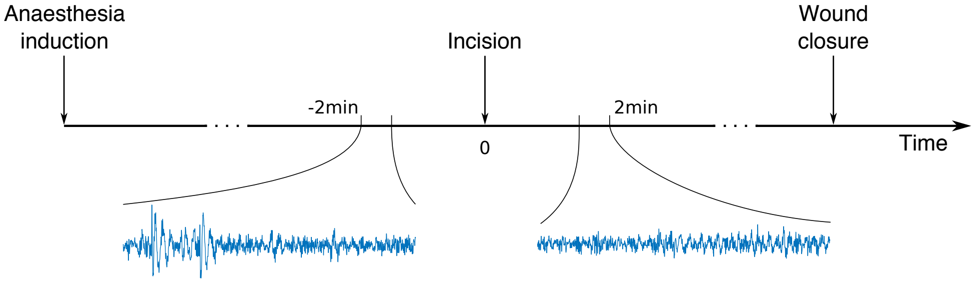
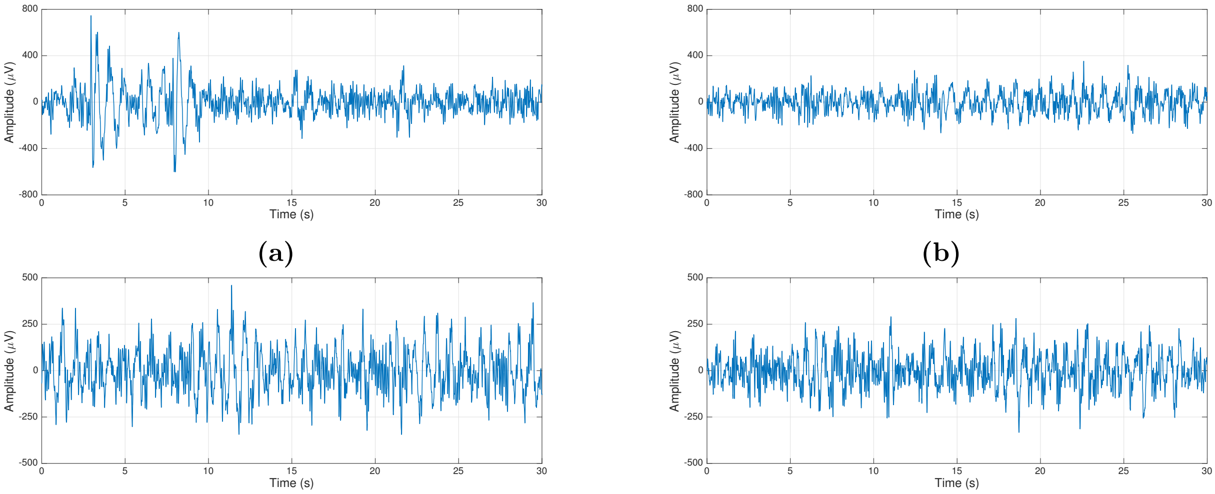


Fig. 2.1: Illustration of data acquisition process. For each patient signals were recorded during surgery under general anesthesia. Skin incision takes place at least 15 minutes after anesthesia induction. Two 30-second signal were recorded 2 minutes before and after the start of the incision.

Patients are divided into two groups based on the anesthetic drug they received: desflurane or propofol. For illustration, Fig. 2.2 shows examples of signals in each of the groups taken from pre- and post-incision periods.



(a) **(b)**

(c) **(d)**

Fig. 2.2: Example signals of the EEG data. (a) Desflurane pre-incision; (b) desflurane post-incision; (c) propofol pre-incision; (d) propofol post-incision.

2.1.2 Synthetic data

In order to validated methods proposed in the study they are first applied to synthetic data. Synthetic data under investigation includes transient oscillations and the solution of the chaotic Lorenz system. Examples of such signals are shown on the Figs. 2.3 and 2.4.

Transient Oscillations

Transient oscillations are generated from a modified Lotka-Volterra model with $N = 3$ interactive elements [122, 123] defined as:

$$\dot{x}_i(t) = x_i(t)\left(\sigma_i - \sum_{j=1}^{N} \rho_{ij} x_j(t)\right), \quad i = 1, 2, 3 \tag{2.1}$$

serves as an abstract model of a heteroclinic sequence.

It has been shown to serve as a good model for the temporal structure of event-related brain potentials [22, 140]. Here $x_i(t) \geqslant 0$ is the activity rate of the element i, σ_i is the growth rate of the i-th element and ρ_{ij} is an entry of the $N \times N$ interaction matrix. In our setup, $\sigma_1 = 1$, $\sigma_2 = 1.2$ and $\sigma_3 = 1.6$, $\rho_{ii} = 1$, $\rho_{12} = 1.33$, $\rho_{13} = 1.125$, $\rho_{21} = 0.7$, $\rho_{23} = 1.25$, $\rho_{31} = 2.1$, and $\rho_{32} = 0.83$. The output signal $s(t)$ to be investigated is a linear superposition of transient oscillations with frequencies $f_1 = 1$ Hz, $f_2 = 2.25$ Hz, $f_3 = 6.3$ Hz, where at one time instance, only one of these three components is dominant, see more details below. We point out that these frequencies are chosen rather arbitrarily for an optimal illustration.

The activity rate x_i defines the amplitude of the signal component y_i with frequency f_i and the output signal is given by:

$$u(t) = \sum_{i=1}^{N} y_i(t) \sin 2\pi f_i t + \xi(t) , \qquad y_i(t) = \exp\left[-\frac{(x_i - \sigma_i)^2}{2\eta_i^2}\right] , \tag{2.2}$$

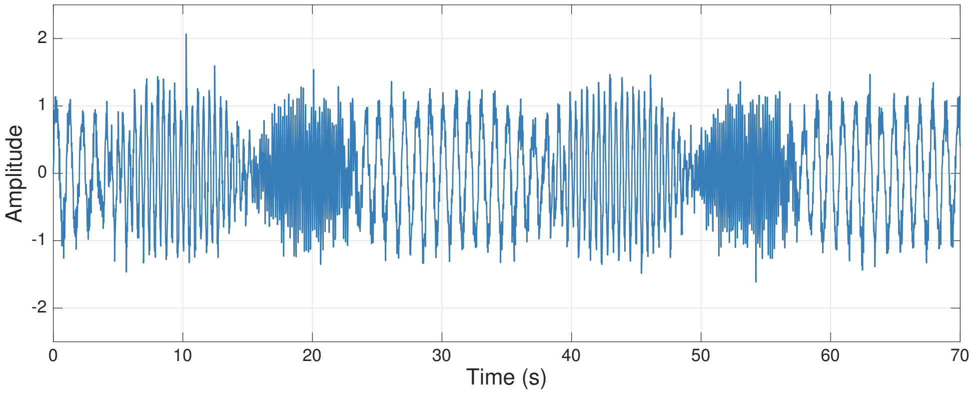


Fig. 2.3: An example of transient oscillations with frequencies $f_1 = 1$ Hz, $f_2 = 2.25$ Hz and $f_3 = 6.3\ Hz$ and the sequence $f_1 \cdot f_2 \cdot f_3 \cdot f_1 \cdot f_2 \cdot f_3 \cdot f_1$.

with $\eta_1 = 0.5, \eta_2 = 0.33, \eta_3 = 0.4$ and $\xi(t) \sim \mathcal{N}(0, 0.01)$ is additive white Gaussian noise (AWGN). By this construction, the signal components y_i oscillates in a certain time window outside of which they almost vanish. These windows of the three oscillation modes do not overlap and the transitions between them are rapid. The sampling rate of the dataset is 50 Hz.

Lorenz dataset

The *Lorenz system* [101] is a well-studied three-dimensional differential equation system:

$$\dot{x} = -\sigma x + \sigma y \ , \quad \dot{y} = \rho x - y - xz \ , \quad \dot{z} = -\beta z + xy \tag{2.3}$$

with $\sigma = 10,\ \rho = 28,\ \beta = 8/3$.

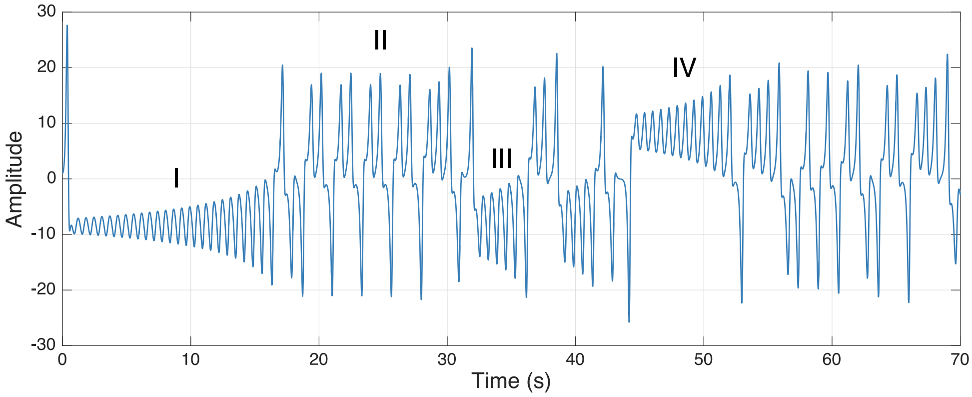


Fig. 2.4: Example solution of the Lorenz system along a single dimension, roman number mark states of interest in the signal.

The system was first introduced by Lorenz in application to weather phenomena [101], it can also serve as a model of a macroscopic-level brain signal, such as an chaotic EEG recording [16, 51, 132]. Its solutions show non-trivial transient dynamics and their wings represent metastable states, cf. [21]. In this work, we study the univariate time series $u(t) = y(t)$ with sampling rate equal to 50 Hz.

2.2 Power Spectral Analysis

We start with *power spectral analysis* (PSA), which captures distribution of power over frequency in finite-length signals. PSA found application in various domains including, for instance, speech analysis [104], seismology [32], and neuroscience [30].

The most important tool of PSA is Fourier transform (FT), which provides representation of a continuous signal $x(t)$ in the frequency domain, and it is defined as:

$$X(f) = \int_{-\infty}^{+\infty} x(t)\,\mathrm{e}^{-2\pi i f t}\,\mathrm{d}t \ , \tag{2.4}$$

here $i = \sqrt{-1}$.

For finite-energy signals, the above expression provides us with energy of the signal, according to Plancherel theorem [117]:

$$\int_{-\infty}^{+\infty} |X(f)|^2\,\mathrm{d}f = \int_{-\infty}^{+\infty} |x(t)|^2\,\mathrm{d}t \ . \tag{2.5}$$

Now, we can see that $|X(f)|^2$ gives us distribution of energy contained in the signal over frequency, i.e., the *energy spectral density* of the signal.

However, real-world signals usually posses infinite energy, in this case, instead of the energy spectral density, we talk about the *power spectral density* (PSD): a distribution of signal's average power. In addition, observed signal is commonly regarded as a realization of a random process. Taking the above into account, we can formulate power spectral density of a discrete-time signal.

For a given signal sampled at equal time intervals $x(t_n) = x(n\Delta t)$ with $n = 0, 1, \ldots, N-1$ and Δt being sampling time, the power spectral density is defined as:

$$P(f) = \lim_{N\to\infty} \mathbb{E}\left[\frac{1}{N}\left|\sum_{n=0}^{N-1} x(t_n)\,\mathrm{e}^{-2\pi i f t_n}\right|^2\right] \ , \tag{2.6}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operation.

Our goal is to obtain a closest estimate $\hat{P}(f)$ of the true PSD $P(f)$. In 1898, Schuster, while working on periodicities in meteorological time series, proposed the *periodogram* method. The periodogram can be obtained from Eq. (2.6) by neglecting the limit and the expectation operation:

$$\hat{P}(f_k) = \frac{\Delta t}{N}\left|\sum_{n=0}^{N-1} x(t_n)\,\mathrm{e}^{-2\pi i f_k t_n}\right|^2 \ , \tag{2.7}$$

where $f_k = k\Delta f$ with $k = 0, 1, \ldots, N-1$ are the equally spaced frequencies, at which periodogram is computed. The frequency spacing is defined through the sampling frequency f_s of a signal:

$$\Delta f = \frac{f_s}{N} = \frac{1}{N\Delta t} \ , \tag{2.8}$$

this is the smallest frequency that can be resolved, for that reason, Δf is also called frequency resolution.

The PSD of real-valued signals is symmetric about $N/2$ and periodic each N samples. It is common to use positive and negative frequencies by treating frequencies $k \geqslant N/2$ as negative ones and centering the PSD at $k = 0$, this representation is known as two-sided PSD. However, in most of the cases we are not interested in power distribution over negative frequencies, and we prefer working with one-sided version, which can be obtained by neglecting negative part of the PSD and multiplying the remaining part by two, except for frequencies at $k = 0$ and $k = N/2$. With this modifications periodogram in Eq. (2.7) represents distribution of power over frequencies from 0 to $f_s/2$ (Nyquist frequency).

However, periodogram suffers from high variance and spectral leakage (also known as end or edge effect), which appears due to finiteness of the data. The latter effect can be partially reduced with an appropriately chosen window function [111] (more information on the use of windows in spectral analysis can be found in [65]). In order to compensate the variance in the PSD estimate, one can average several PSD estimates of the same signal [134]. One of such methods is *Welch's modified periodogram.*

Welch's method [144] consists of dividing the signal into several (possibly overlapping) segments and averaging PSD estimates of each segment. Prior to PSD estimation, a smoothing window (also called taper) is applied to each segment in order to decrease edge effect. For a signal x_n divided into M segments Welch's PSD estimate can be written as:

$$\hat{P}(f_k) = \frac{1}{M} \sum_{m=1}^{M} \hat{P}_m(f_k) \ , \tag{2.9}$$

where $\hat{P}_m(f_k)$ is a modified periodogram of m-th segment:

$$\hat{P}_m(f_k) = \frac{\Delta t}{N_w U} \left| \sum_{n=0}^{N_w-1} x_m(t_n)\, w(t_n)\, \mathrm{e}^{-2\pi i f_k t_n} \right|^2 \ , \tag{2.10}$$

where N_w is the length of the window with its total power denoted by U:

$$U = \frac{1}{N_w} \sum_{n=0}^{N_w-1} |w(t_n)|^2 \ . \tag{2.11}$$

Welch's method reduces variance of the estimate, although reducing frequency resolution, as now it is defined by window's length. More specifically, N in Eq. (2.8) is replaced by N_w. It is a common practice to pad the window with zeros, so that its length equals a power of 2. It is done for two reasons: (i) modern FT implementations (specifically, fast Fourier transform or FFT) work more efficiently with signals whose length is a power of 2; (ii) it allows to increase resolution of the spectrum and make it smoother [112]. However, one have to keep in mind that actual frequency resolution is defined through the length of the original window, and finer estimate obtained with zero-padding should be seen as adding interpolation points in between real spectral frequencies.

In this work, we use Welch's modified periodogram with Gaussian window:

$$w(t_n) = \exp\left[-\frac{1}{2} \left(\alpha \frac{t_n}{(N_w - 1)/2} \right)^2 \right] \ , \tag{2.12}$$

where $t_n = n\Delta t,\ -(N_w - 1)/2 \leqslant n \leqslant (N_w - 1)/2$, N_w is the length of the window, and α is a parameter controlling the width of the window. These parameters are inversely related: larger α lead to narrower windows. Here, $\alpha = 2.5$ and overlap between consecutive segments is 50%. The length of the window is 128 time points, providing frequency resolution $\Delta f = 1$ Hz, afterwards the segments were zero-padded to 256 points to make PSD estimate smoother.

After obtaining PSD estimates of the EEG signals we are interested in total power contained in the signals and power in alpha(α)- and delta(δ)-band [133]. The reason of our interest is that these EEG rhythms are characteristic to maintenance period of anesthesia, the period, when surgery is normally conducted. Namely, as the state of general anesthesia progresses, we can observe increased high-frequency low-amplitude (alpha) and low-frequency high-amplitude (delta) activity, especially in frontal lobes [130].

Power contained in particular frequency band can be obtained from spectral estimate by integrating it over a given range. Our first feature is the total power contained in the signal, for that we need to integrate signal's PSD estimate from 0 Hz to Nyquist frequency, however, since the signal was previously bandpass filtered, we integrate only from 1 to 41 Hz.

Next two features comprise power of δ- and α-oscillations. To obtain delta-band power we need to integrate the PSD estimate in the frequency range of 1–4 Hz and for alpha-band this range is 8–12 Hz. However, prior to obtaining the values of band power, we need to separate narrow-band activity from the underlying broad-band noise, which reveals $1/f$ scaling of the power spectrum [40]. This broad band activity in EEG signals has different names in literature, few of such names include: amorphous or arrhythmic activity [40], background activity [15, 95], broad-band or $1/f$-noise [72, 133]. The nature of such scaling in brain activity has been extensively studied. The $1/f$ behavior of power spectrum can be associated with system's self-organized criticality [14, 50, 72], filtering of neuronal activity by extracellular space [18, 19] and even chaotic dynamics [42]. However, study of origins of power law behavior of the PSD is beyond the scope of this work. In this manuscript, we want to fit a model of background activity to the PSD of EEG data and subtract it afterwards. By doing that we obtain a resulting PSD estimate, which we can integrate to obtain power contained in α- and δ-peaks of the PSD estimate without background activity.

We utilize *Ornstein–Uhlenbeck (OU) process* to model EEG's background activity [54, 76, 92], which is used to generate correlated or colored Gaussian noise (as opposed to white noise). It has been proposed by Uhlenbeck and Ornstein to model Brownian motion [141] and is given by:

$$\frac{\mathrm{d}\nu}{\mathrm{d}t} = -\gamma\nu + \sqrt{2D}\xi(t) \ , \tag{2.13}$$

where $\xi(t)$ is Gaussian white noise with zero mean, $\sqrt{2D}$ defines the intensity of the noise, and $\gamma = 1/\tau$ is the inverse of the correlation time of the noise.

And the power spectral density of OU process [91]:

$$P_{OU}(f) = \frac{2D}{\gamma^2 + (2\pi f)^2} \ . \tag{2.14}$$

This model has two parameter which have to be fitted: $2D$ and γ^2, for simplicity we denote them as `2D` and `gamma2`, respectively. We want to fit the model to the PSD excluding α- and δ-peaks, however, we didn't exclude the points directly from the PSD, as it may produce artifacts, such as rapid change in values. Instead we used weight vector $\mathbf{u}$ for the fitting process. By using weights in curve fitting process we ensure that the curve will be fitted more accurately to the regions with higher weights than to the regions with lower weights. We initialize the vector with weights $u_i = 50$, then we assign the highest weights to frequencies above 20 Hz, i.e., $u_i = 200$, for $f \leqslant 20$ Hz, as we are sure that no narrow-band activity is expected there, next, we assign the smallest weights to α- and δ regions: $u_i = 1$, for $f \leqslant 4.5$ Hz and $8 \leqslant f \leqslant 12$ Hz. Finally the weight vector is normalized in the following way: $u_i = u_i / \sum_{i=0}^{N-1} u_i$. When performing curve fitting we want to find such parameters of the model, that give us the smallest value of objective function (also known as cost function), which measures the difference between the data and the fitted curve. Here, least-squares estimation plays the role of the objective function, it computes the (weighted) sum of squared error between the experimental and modeled data:

$$SSE(\theta) = \sum_{i=0}^{N-1} u_i \left| y_i - \hat{y}_i(\theta) \right|^2 \quad , \tag{2.15}$$

where y_i is the ith data point, $\hat{y}_i(\theta)$ is the model prediction for ith data point, θ is the model's parameters, u_i is the ith value of the weight vector, and N is the total number of data samples.

We performed curve fitting in two steps combining global and local search algorithms. Local search methods converge fast to the optimal solution when a good guess for initial parameters exists. Unfortunately, there is a high risk of converging to local minimum of optimization function if this function has several minima [125] (which is the case in nonlinear least squares fitting). The global search methods require no initial guess of the parameters, however they also do not guarantee global convergence, but they can be used as a good estimate of initial conditions. In this work, we start by using Particle Swarm Optimization (PSO), a global stochastic search method introduced by Kennedy and Eberhart while working on model of social behavior, such as bird flocking or fish schooling [80, 119]. We use PSO 10 time, which gives us 10 different sets of parameters, then we choose the parameters with the smallest value of objective function, and we use the parameters as initial parameters in local optimization. Local search method for this study was presented with Levenberg-Marquardt [96, 103] algorithm, a modification of gradient descent algorithm based on updating solution in the direction of the steepest decrease of the gradient of the objective function.

After obtaining optimal parameters of the background activity model for each signal, we substitute the curve from the PSD estimates, and compute α-band, δ-band, and total power. This gives us three features of the EEG signals. In addition we also include $2D$ and γ^2 from Eq. (2.14) as characteristic to EEG background activity.

To summarize this section, Table 2.1 provide a list of all features generated on this stage of the analysis along with notations we use throughout the manuscript.

Table 2.1: Features generated during the power spectral analysis. First column shows the notation used to refer to the feature further in the manuscript. Second column provides the description of the feature.

Notation	Description
`totPow`	total power contained in the signal's PSD after removing the PSD of background activity
`alphaPow`	power contained in α-band computed after removing the PSD of background activity
`deltaPow`	power contained in δ-band computed after removing the PSD of background activity
`2D`	parameter of the model of EEG background activity, correspond to $2D$ in Eq. (2.14)
`gamma2`	parameter of the model of EEG background activity, correspond to γ^2 in Eq. (2.14)

The examples of PSD estimates of EEG signals with fitted background activity PSD, as well as distributions of the features are presented in the Results section (cf. Section 2.6.1).

2.3 Symbolic Recurrence Structure Analysis

The next step in our work is to analyze non-stationary behavior in the data. For this purpose we chose recurrence-based methods. More precisely, we utilize *symbolic recurrence structure analysis* (SRSA), which allows to identify recurrent and stationary domains in system's dynamics and rewrite the signal as a sequence of symbols. Symbolic representation then allows for analysis with information-based methods, such as complexity measure.

Recurrence is a fundamental property of nonlinear dynamical systems, which was first formulated by Poincaré in [118]. It was further illustrated in *recurrence plot* (RP) technique proposed by Eckmann *et al.* [41]. This relatively simple method allows to visualize multidimensional phase trajectories of a nonlinear system on a two-dimensional plane. The RP can be obtained by plotting the recurrence matrix:

$$R_{ij} = \Theta\left(\varepsilon - ||\mathbf{x}_i - \mathbf{x}_j||\right), \qquad i,j = 1,2,\ldots,N \ . \tag{2.16}$$

Here $\mathbf{x}_i \in \mathbb{R}^d$ is the state of the complex system in the phase space of dimension d at a time instance i, $||\cdot||$ denotes a metric, Θ is the Heaviside step function, and ε is a threshold distance.

It can be seen from Eq. (2.16), that if two points in the phase space are relatively close, the corresponding element of the recurrence matrix $R_{ij} = 1$ and will be represented by a black dot on the RP, in contrast, white dots on the RP represent the absence of recurrence, i.e., $R_{ij} = 0$, an example of a recurrence plot can be seen on Fig. 2.6c. By visual inspection of RPs an expert can draw conclusions on the studied system's dynamics and behavior. For example, isolated black dots may indicate rare or unique event, which is not repeated in time or it fluctuates; parallel diagonal lines may indicate periodical behavior; horizontal or vertical lines mean that system is evolving very slowly or not evolving at all; dots uniformly distributed over the RP may be a sign of a stochastic stationary process.

Visual analysis of RPs may be challenging and require a lot of expertise, the search for more quantitative and formal approach gave rise to Recurrence Quantification Analysis (RQA) developed by Zbilut and Webber [143, 148] and then further extended by Marwan *et al.* in [107]. RQA allows to identify the nature of a system under study by calculating a set of single-valued measures. It can be adopted to time-dependent study by using sliding analysis window. This analysis was adopted in such domains as biology [3, 10, 97], physics [31, 87] and economics [5, 108].

RQA is a powerful tool, which permits to identify non-stationarity, drifts, transitions to chaos and to order. However, we are interested in specific types of activity, which can be extracted from RPs. This activity can be partitioned into three broad groups: (i) stable states, meaning that the system evolves close to a single state within a long time interval, (ii) recurrent states, which are the states to which trajectory returns after some time and stays there for a short intervals of time, and (iii) transient states, which evolve fast, are unique for the system and do not repeat within the time of observation. The first two states have a subtle difference between each other and both are denoted on a recurrence plot with a black-colored dot, i.e., $R_{ij} = 1$, however, stable states occupy a region of the RP, which does not necessarily repeat, on another hand, the second type of activity is always present on the RP more than once. The last type of activity can be seen as non-recurrent regimes which occurs during system's transitions from one stable state to another.

To identify the different kinds of behavior described above, we use symbolic recurrence plots. By transforming original signal to a symbolic sequence, we can assign a particular symbol to each type of dynamics and afterwards quantify it using information theory. Previous studies showed successful application of different implementations of this approach. For instance, order patterns recurrence plots [59, 128], where recurrence plots were obtained from order patterns, were used to provide a measure of coupling strength between two systems robust against artifacts and successfully applied to EEG analysis. Another variation of the approach, that consist of constructing RPs from words in symbolic sequences, found applications in such domains as performance analysis [39] or behavior analysis [44]. The symbol-based RPs provide advantage of coarse-graining, i.e., focusing on larger time scales, in addition, such RPs are transitive, which means if $R_{ij} = 1$ and $R_{jk} = 1$, then $R_{ik} = 1$.

In this study, we deviate from previously described methods and use approach proposed by beim Graben and Hutt, which generates symbolic representation of a signal from the recurrence plots. Previous studies [21–23] have shown successful application of the method to the detection of transient and stable states in synthetic and event-related brain potential.

The main idea of the method is to partition phase trajectory into a set of balls with diameter ε around sampling point $\mathbf{x}_i$. And then merge intersecting balls into a new set by examining the system's RP: if two balls i and j are intersecting, then $R_{ij} = 1$. By continuing merging intersecting balls, we partition phase trajectory into n disjoint sets S_k. Let $\mathcal{P} = \{S_k \subset \mathbb{R}^d | 1 \leqslant k \leqslant n\}$ be the final partitioning of the phase trajectory, then desired sequence $s = s_i$ represents states visited by the system at each discrete time, i.e., $s_i = k$ when $\mathbf{x}_i \in S_k$. To obtain aforementioned partitioning, we start by mapping a phase trajectory $\mathbf{x}$ to an initial symbolic sequence, where each symbol correspond to a time index, $s_i = i$. If two states at times i and j are recurrent, i.e., $R_{ij} = 1$, and $i > j$, we can rewrite the symbol at time i with the symbol s_j. We now see that the recurrence matrix can be interpreted as a grammar in rewriting process, where higher indices are replaced by the lower ones if the states at these discrete times are recurrent. In addition, this method resolves cases when three states are recurrent, i.e., $R_{ij} = 1$ and $R_{ik} = 1$. The described rewriting rules can be formulated as follows:

$$\begin{aligned} &i \to j && \text{if } i > j \text{ and } R_{ij} = 1 \\ &\left.\begin{aligned} i \to k \\ j \to k \end{aligned}\right\} && \text{if } i > j > k \text{ and } R_{ij} = 1,\ R_{ik} = 1\ . \end{aligned} \tag{2.17}$$

The process of rewriting the initial symbolic partition accordingly should be performed twice in order to resolve both rules in Eq. (2.17). The subsequent step is to identify the transient states. Since we are not interested in each of these states in particular, we map sequentially increasing indices to the symbol $s_0 = 0$ to mark them as transitions from one quasi-stable state to another [21, 22]. Figure 2.5 schematically illustrates the process of phase trajectory's partitioning.

The method described allows the visual representation of a system's dynamics and is rather simple with only one parameter to estimate. This parameter is the threshold distance ε. As we can see from Eq. (2.16), this parameter is crucial to the method. If ε is too small no recurrence can be formed and each state will be detected as a distinct state. Conversely, if the value of ε is too large, then all the states will be observed as one single state.

Several estimates of the optimal threshold distance exist [106], however most of them can be regarded as "rules of thumb". We utilize the estimates of ε based on assumptions of the system's dynamics as it was done in [21]. The main idea of the method is to choose the optimal ε such that it maximizes a *utility function* which represents our assumption. This utility function reflects a

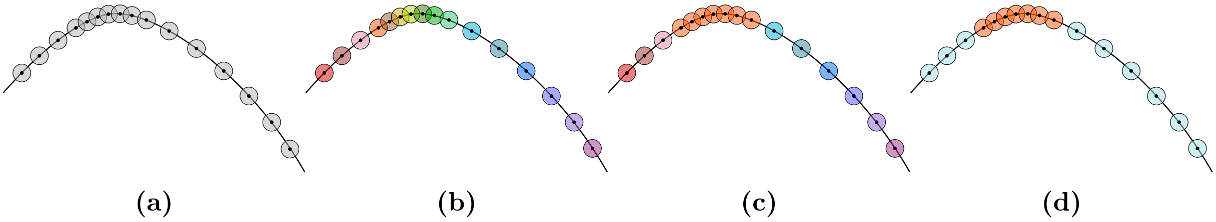


Fig. 2.5: Process of partitioning the system's phase trajectory into recurrence domains. (a) An initial partition of the phase trajectory (shown with black solid line) into a set of balls with diameter ε. (b) All balls are mapped onto a symbolic sequence $s_i = i$, corresponding to discrete time indices $i = 1, \ldots, N$. (c) Intersecting balls are merged together and assigned one symbol s_k. (d) Non-overlapping balls, representing transient states, are mapped to the symbol $s_0 = 0$ (Adopted from [20]).

given model of the temporal recurrence structure. In the present work, we assume that the system spends an equal amount of time in each of the recurrence domains and in the transient states. This hypothesis can be summarized by the principle of maximum entropy. In more details, to obtain the optimal threshold distance, first one creates a vector of ε values, and for each value in this vector obtains a symbolic sequence. Then, the optimal threshold distance ε^* can be chosen as the one which maximizes the utility function:

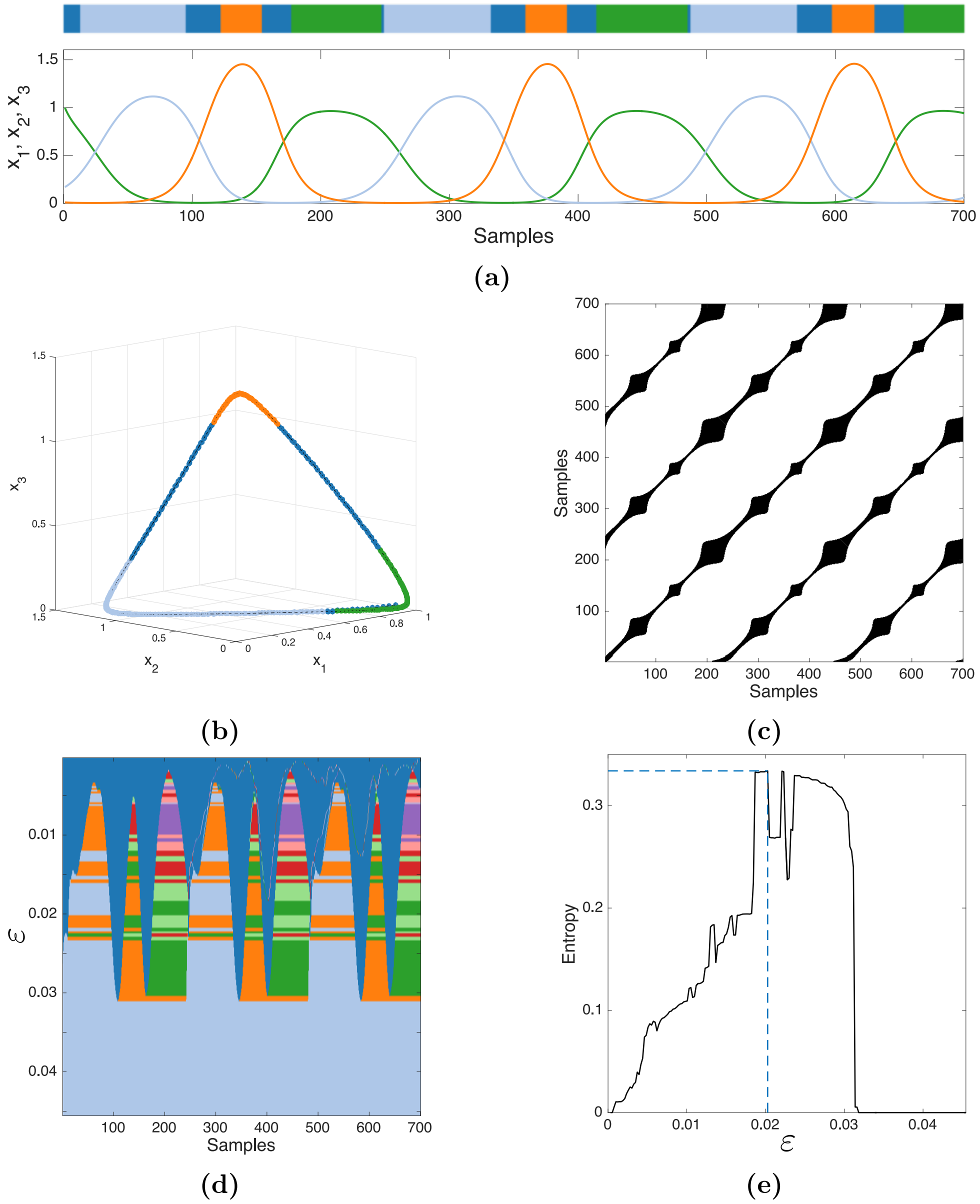
$$\varepsilon^* = \arg\max_{\varepsilon} h(\varepsilon) \quad , \tag{2.18}$$

where $h(\varepsilon)$ is the normalized entropy:

$$h(\varepsilon) = -\frac{\sum_{k=1}^{M(\varepsilon)} p_k \log p_k}{M(\varepsilon)} \quad . \tag{2.19}$$

Here p_k is the relative frequency of symbol $s_i = k$ and $M(\varepsilon)$ is the cardinality of the symbolic alphabet. The denominator in Eq. (2.19) ensures that a too small or too large alphabet does not bias the estimate. If ε is too small the resulting symbolic sequence will consist of many isolated non-recurrent symbols, which are distributed almost uniformly. This will lead to high value of entropy, Eq. (2.19) penalize it by large alphabet size. On the contrary, large ε produce a small number of recurrent states, which yield trivial partition with low entropy. Figure 2.6 illustrates the procedure of obtaining an optimal recurrence-based symbolic representation of a signal using Lotka-Volterra system given by Eq. (2.1) as an example.

Figure 2.6a shows the analyzed signal, consisting of three time series x_1, x_2, and x_3, where each time series represents certain type of system's dynamics. The system's trajectory in the phase space resembles a triangle, as shown on Fig. 2.6b. The system visits each vertex of the triangle successively, staying in each vertex for a period of time and then transitioning to the next one. In order to find ε^*, first we create an arbitrary vector of ε values. There is no strict requirements to the ε vector, except that the smallest value have to ensure presence of at least one recurrent point and the larges value have to be such that utility function returns to zero (the utility function should resemble the one presented in Fig. 2.6e). It also has to be densely spaced, for precision reasons. Next step is to compute RPs for each value of ε vector and obtain corresponding symbolic partitions, on Fig. 2.6d each row is a symbolic sequence obtained with ε-RP. As we can observe, the top of this panel shows only transient state (encoded in dark blue), as ε is too small to capture any recurrence domain. On the other hand, bottom of the same



(a)

(b) **(c)**

(d) **(e)**

Fig. 2.6: Recurrence-based symbolic dynamics of the Lotka-Volterra equations. (a) Solution of the Lotka-Volterra equations, presented with three time-series x_i, $i = 1, 2, 3$. The color bar above represents optimal symbolic encoding. Orange, green, and light blue colors mark three quasi-stable states, while transient states are shown in dark blue. (b) The system's phase trajectory resembles a triangle, the vertices of which are stable states, that the system visits successively. The trajectory is colored according to the optimal symbolic partitioning. (c) Recurrence plot obtained with optimal threshold $\varepsilon^* = 0.2$. (d) Symbolic sequences s for the range of $\varepsilon \in [0.0003,\ 0.0536]$. (e) The black solid line shows symbolic entropy Eq. (2.19), the blue dashed line shows the optimal threshold ε^* and the maximum of the entropy.

panel illustrates the situation when ε is too large and only one stable state is present (shown in light blue). By computing utility function (recurrence entropy as given in Eq. (2.19)) and by maximizing its value we obtain the optimal threshold distance ε^*. The top panel of Fig. 2.6 shows the optimal recurrence-based symbolic sequence, where three stable states are color-coded in orange, green and light blue, which also correspond to vertices of the triangle on Fig. 2.6b, and transitions between these states are shown in dark blue. Fig. 2.6c shows ε^*-RP used to generate the symbolic sequence.

2.4 Phase Space Reconstruction

The discussed technique of recurrence plots and recurrence-based symbolic dynamics usually deals with dynamical systems, whose phase space is multidimensional. Each dimension of the phase space is a certain property of a system, it can, for instance, be obtained from a system of ordinary differential equations, which defines system's evolution law or it can be presented with multivariate measurements of an observable (e.g., several EEG channels). The system's trajectory in the phase space correspond to the system's states at particular times.

In certain cases only discrete univariate measurements of an observable are available, in this situation a phase space should be reconstructed according to Takens's theorem [135], which states that phase space presented with a d-dimensional manifold can be mapped into $2d+1$-dimensional Euclidean space preserving dynamics of the system. Several method of phase space reconstruction exist: delayed embeddings [135], numerical derivatives [114] and others (see for instance [88]).

In this work we propose a new heuristic method of phase space reconstruction based on the time-frequency representation (TFR) of the signal, a distribution of the signal's power over time and frequency. Here, the power in each frequency band contributes to a dimension of the reconstructed phase space. This approach is well-adapted for non-stationary and, especially, for oscillatory data. The current section presents different phase space reconstruction methods, starting with delayed embeddings and continuing by describing different TFRs: spectrogram, reassigned spectrogram and scalograms (based on continuous wavelet transform). Below we argue that the proposed reconstruction with the signal's TFR is more advantageous for SRSA analysis of oscillatory signals.

2.4.1 Delayed Embeddings

Assume, we have a time series which represents scalar measurements of a system's observable in discrete time:

$$x_n = x(n\Delta t), \quad n = 1, \ldots, N \ , \tag{2.20}$$

where Δt is measurement sampling time.

Then reconstruction of the phase space is done in the following way:

$$\mathbf{s}_n = \left[x_n, x_{n+\tau}, x_{n+2\tau}, \ldots, x_{n+(m-1)\tau}\right], \quad n = 1, \ldots, N-(m-1)\tau \ , \tag{2.21}$$

where m is the embedding dimension and τ is the time delay. These parameters play an important role in correct reconstruction and should be estimated correctly.

Optimal time delay τ should ensure that reconstructed delay vectors from Eq. (2.21) are sufficiently independent. Several approaches to estimate this parameter exist. One can visually inspect obtained phase portraits; for data with periodic component, a good initial estimate of τ is a quarter of the period [79]. However, these are empirical approaches. Amongst more formal methods, the two leading estimates are based on autocorrelation function [1, 79] or average mutual information (AMI) [49, 98]. The autocorrelation approach considers linear independence between delayed coordinates as a sufficient criterion for optimal τ, while AMI also takes non-linear relations between the coordinates into account. Owing to this reason we chose to use AMI to estimate the proper value of the time delay τ.

Mutual information is the amount of information that one can learn about one system by observing another system. Let us call these two systems X and Y with possible observations x_i and y_j. If we denote the probability of observing x_i by $P(x)$, probability of observing y_j by $P(y)$ and the joint probability of x_i and y_j by $P(x, y)$, then average mutual information between the two systems is given in [55] as:

$$I(X,Y) = \sum_{x,y} P(x,y)\,\log_2 \left[\frac{P(x,y)}{P(x)P(y)} \right] \quad . \tag{2.22}$$

In the context of delayed coordinates, AMI is Dependant on τ and indicates the amount of information we obtain about the observation at the time $n + \tau$ from the observation at the time n. And Eq. (2.22) is then transformed into:

$$I_\epsilon(\tau) = \sum_{n=1}^{N-\tau} P(x_n, x_{n+\tau})\,\log_2 \left[\frac{P(x_n, x_{n+\tau})}{P(x_n)P(x_{n+\tau})} \right] . \tag{2.23}$$

In practice, Eq. (2.23) can be resolved by computing histogram of x_n with bin size ϵ, it will give us $P(x_n)$. If the data is stationary, $P(x_{n+\tau})$ is equal to $P(x_n)$. And $P(x_n, x_{n+\tau})$ is found from a joint histogram of x_n and $x_{n+\tau}$. As a rule of thumb, the optimal value of τ is found at the first local minimum of AMI [49].

Next important parameter in phase space reconstruction process is the embedding dimension m. The main attribute of appropriately chosen dimension m is that original d-dimensional manifold will be embedded into m-dimensional space without ambiguity, i.e., self-crossing or intersections. A method of false nearest neighbors (FNN) [81, 82] takes it into account, which allows to estimate minimal embedding dimension.

The idea behind this method is that points which are close (neighbors) in dimension m' might be located far from each other in dimension $m' + 1$. Such points are called false neighbors and might be seen as projections of two distant points onto smaller dimension. Here, closeness of two points is defined in a sense of distance, most commonly it is Euclidean distance. For each dimension m we compute the number of FNN, starting from $m = 1$ until a chosen $m_{\max}$, and the dimension in which the number of FNN is zero (or close to zero) is the optimal embedding dimension.

To determine if two points are false neighbors we have to check the following inequality:

$$\frac{|\mathbf{s}_n^{(m+1)} - \mathbf{s}_{k(n)}^{(m+1)}|}{R_n^{(m)}} > R_{\mathrm{tol}} \quad , \tag{2.24}$$

where $\mathbf{s}_{k(n)}$ is a kth nearest neighbor of $\mathbf{s}_n$, superscripts signifies dimension, R_{tol} is some threshold. Normally $10 < R_{\mathrm{tol}} < 20$ [81] and $R_n^{(m)}$ is a square root of Euclidean distance between $\mathbf{s}_n$ and $\mathbf{s}_{k(n)}$ in the dimension m.

In their paper Kennel *et al.* [82] propose an additional constraint for FNN, in order to eliminate influence of noise by putting the penalty on the largest distance between the nearest neighbors. By doing so we assure that points which are already far away from each other will not bias the statistics because distance between them cannot possibly grow when going to higher dimensions. This constraint is formulated in the following condition:

$$\frac{R_n^{(m+1)}}{\sigma} > R_{\mathrm{A}} \ , \tag{2.25}$$

where $R_n^{(m+1)}$ is the distance between two possible nearest neighbors in the dimension $m+1$, σ is the standard deviation of the data and R_A is a fixed threshold. Usually $R_A = 2$ is a reasonable choice.

For each dimension we count the number of points satisfying the above conditions, and the dimension in which total number of FNN drops to zero is considered to be the optimal one. In practice, due to noise and random fluctuations in the data this level cannot always be achieved. It would be reasonable to aim for a certain threshold (5% from the number of FNN at dimension $m = 1$ in this work) instead of zero. In addition we set a limit on total number of dimensions ($m_{\max} = 10$), because higher number of dimension might be too computationally expensive without adding meaningful information, especially in case of noisy data.

2.4.2 Time-Frequency Embedding

In contrast to the method of phase space reconstruction described above, we propose an alternative approach to this problem based on the use of a signal's time-frequency representation. A time-frequency representation is a distribution of the signal's power over time and frequency. It provides several time series, one for each frequency bin, which can later be used in reconstruction process, where each frequency band contributes to the multi-dimensional phase trajectory. The main advantage of this approach over existing ones is that it takes into account the oscillatory nature of the signal and is more robust than analyzing RPs constructed from individual amplitude values. Below we provide description and examples of the TFRs used in the present work: the spectrogram, the reassigned spectrogram and two variants of the scalogram (based on continuous wavelet transform).

Spectrogram

The spectrogram of a signal is the squared magnitude of its *short-time Fourier transform* (STFT). In this approach the power spectrum of the signal is extended with temporal information by the use of a sliding analysis window. The STFT is closely related to the Welch's PSD estimate. In order to obtain the STFT one has to follow the same steps as in the Welch's method, except for averaging in Eq. (2.9). The STFT $F(t, f)$ of a signal $x(t)$ on an infinite time interval is given by:

$$F(t, f) = \int_{-\infty}^{+\infty} x(\tau) w^*(t - \tau) \mathrm{e}^{-2\pi f i \tau} \, \mathrm{d}\tau \ , \tag{2.26}$$

where $w(t)$ is a smoothing window and $*$ denotes the complex conjugate. Then by taking a squared absolute value of Eq. (2.26), we obtain the *spectrogram*:

$$S(t, f) = |F(t, f)|^2 \ . \tag{2.27}$$

The spectrogram is computed in three steps: divide the signal into several possibly overlapping segments, apply the smoothing window $w(t)$ to the segments and compute the power spectrum for each of them. This approach makes the spectrogram a suitable tool for the non-stationary data analysis.

In the present work the spectrograms were computed with a Gaussian window given previously by Eq. (2.12) with width parameter $\alpha = 2.5$. The overlap between consecutive segments was empirically estimated to be 80% of the window length, as the one producing the best results. The length of the segments was chosen such as to provide a resolution of 0.3 Hz for synthetic data and 1 Hz for EEG data. Afterwards the segments were zero-padded to the closest power of 2 for computational efficiency. An example of the spectrogram applied to a test signal is shown on Fig. 2.7a.

An important issue that has to be addressed is application of the method to discrete-time finite signals. Equation (2.26) has to be mapped to discrete frequencies and times. The frequencies at which the spectrogram (and STFT) is computed are defined analogously to the Welch's modified periodogram:

$$\Delta f = \frac{f_s}{N_w} \ , \tag{2.28}$$

where f_s is the sampling rate of the signal and N_w is the window's length after zero padding (if any was used).

The discrete times at which the spectrogram is calculated are defined by the position of the analysis window and do not necessarily matches time resolution of the signal. It is defined through the length of the window N_w and the overlap between successive segments N_{ovlap}:

$$\Delta t' = \frac{N_w - N_{ovlap}}{f_s} \ . \tag{2.29}$$

We can see that the spectrogram performs "downsampling" of the original data, as the number of time points at which the TFR is computed now becomes:

$$M = \frac{N - N_{ovlap}}{N_w - N_{ovlap}} \ . \tag{2.30}$$

In addition, it is important to mention the edge effects appearing at the borders of the time interval under study due to its finiteness. Hatched areas in Fig. 2.7a marks the area of the TFR affected by this issue. For the spectrogram, the time window length is independent of the frequency, which leads to an edge effect of constant width. In order to make the boundary effect less prominent, we padded the signals with a mirrored version of itself, where length of the padding is half of the window length [136].

Reassigned Spectrogram

The method of spectrogram reassignment [12, 13] decreases the spreading of the conventional spectrogram by moving its value from original location where it has been computed to the centroids of the energy distribution (one can think of such centroids as "centers of gravity" of the distribution).

This method was first proposed in the end of 1970s by Kodera *et al.* under the name of *Modified Moving Window Method* [85]. Here, the authors used moving (or sliding) window method similar to STFT in order to obtain coefficients of the signal's TFR. Then each point on the time-frequency (TF) plane (t, f) was assigned to a new location $(\hat{t}, \hat{f})$, which better reflects the energy distribution. Such points were estimated based on the partial derivatives of the phase of the STFT, which is discarded when computing the spectrogram. For this reason and also because it is computationally difficult to obtain phase derivatives in the digital domain, the method did not receive wide attention until mid-1990s, when it was rediscovered by Auger and Flandrin.

In their paper [12] Auger and Flandrin proposed computation of TFR centroids through the *Wigner-Ville distribution* (WVD) [36]:

$$WV(t,f) = \int_{-\infty}^{+\infty} x(t+\tau/2)\, x^*(t-\tau/2)\mathrm{e}^{-2\pi f i \tau}\, \mathrm{d}\tau \ . \tag{2.31}$$

The WVD posses many "desirable" theoretical properties [67, 75], amongst which the most important one for the reassignment method is a perfect localization of a linear chirp: a signal with linear frequency modulations represented by a straight line on the TF plane. However, if more than one component is present in the signal, WVD produces non-negligible cross-terms due to interactions between the components and the quadratic nature of the WVD. To overcome this problem a low-pass smoothing of the distribution was proposed, the spectrogram can be regarded as one of the applications of this technique [77] and can be written in an alternative form:

$$S(t,f) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} WV_w(\tau-t,\nu-f)WV_x(\tau,\nu)\,\mathrm{d}\tau\,\mathrm{d}\nu \ , \tag{2.32}$$

where $WV_x(t,f)$ and $WV_w(-t,-f)$ are WVD of the signal and of the spectrogram analysis window, respectively.

The smoothing of the WVD reduces (or completely attenuates) the cross-terms, while sacrificing perfect localization of signal's components. To overcome this problem Auger and Flandrin proposed to tackle the problem by reassigning spectrogram values to more precise location, which, by analogy with mechanics, they refer to as the "center of gravity" of the distribution (as opposed to the geometrical center) [34]. The value of the spectrogram is reassigned from its original TF coordinates (t,f) to the *centroid* $(\hat{t},\hat{f})$:

$$RS(t,f) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} S(\tau,\nu)\delta(f-\hat{f}(\tau,\nu))\delta(t-\hat{t}(\tau,\nu))\,\mathrm{d}\tau\,\mathrm{d}\nu \ , \tag{2.33}$$

where δ is Dirac delta function, and the centroids are given by [13, 34]:

$$\hat{f}(t,f) = \frac{1}{S(t,f)}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \nu\, WV_w(\tau-t,\nu-f)\, WV_x(\tau,\nu)\,\mathrm{d}\tau\,\mathrm{d}\nu \ , \tag{2.34}$$

$$\hat{t}(t,f) = \frac{1}{S(t,f)}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \tau\, WV_w(\tau-t,\nu-f)\, WV_x(\tau,\nu)\,\mathrm{d}\tau\,\mathrm{d}\nu \ . \tag{2.35}$$

However, there exists a simplified solutions to Eq. (2.34), which requires computation of two additional STFTs with modified windows:

$$\hat{f}(t,f) = f - \mathrm{Im}\left[\frac{F^{dw}(t,f)}{F^{w}(t,f)}\right] \ , \tag{2.36}$$

$$\hat{t}(t,f) = t + \mathrm{Re}\left[\frac{F^{tw}(t,f)}{F^{w}(t,f)}\right] \ , \tag{2.37}$$

where F^w is the STFT of the signal $x(t)$ computed with the analysis window $w(t)$ as defined in Eq. (2.26), F^{dw} and F^{tw} are the STFTs computed with windows $\frac{\mathrm{d}w(t)}{\mathrm{d}t}$ and $t \cdot w(t)$, accordingly;

The discrete frequencies and times are computed identically to the conventional spectrogram. More information and practical details about the implementation of the method can be found in [53] or [47].

Scalogram

By analogy with the spectrogram, the scalogram is the energy distribution of a signal calculated as the squared magnitude of the *wavelet transform*. The continuous wavelet transform (CWT) is obtained by convolving the signal with a set of functions $\psi_{a,b}(t)$ obtained by translation b and dilation a of a mother wavelet function $\psi_0(t)$ [4]:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi_0\left(\frac{t-b}{a}\right) \ . \tag{2.38}$$

The mother wavelet function should be admissible [43], i.e., it should be centered around zero, have a compact support and a zero mean, and be square integrable. The definition of the CWT for an infinite time domain is given by:

$$T_\psi(b,a) = \int\limits_{-\infty}^{+\infty} x(t)\psi^*_{a,b}(t)\,\mathrm{d}t \ , \tag{2.39}$$

then the *scalogram* can be obtained by:

$$W_\psi(b,a) = |T_\psi(b,a)|^2 \ . \tag{2.40}$$

The continuous wavelet transform, as well as the spectrogram, is well adapted for the analysis of non-stationary data. However, unlike the spectrogram, where one faces a trade-off between the resolution in time and in frequency (called the Heisenberg relation [102]), the wavelet transform has an adaptive resolution. In spectrograms, the resolution is determined by the length of the segments: short segments provide good time resolution but have wide spread in frequency domain and vice versa (cf. Eqs. (2.28) and (2.29)). In contrast, wavelet functions have larger temporal and smaller frequency spreading for lower-frequencies and the opposite for higher frequencies. Due to this property of the CWT the choice of the scales at which the transform is computed plays a very important role.

Scales in wavelet transform are inversely proportional to frequencies [139], i.e., $a \sim 1/f$ and the pseudo-frequencies f can be approximated from scales by:

$$f = \frac{1}{\lambda a} \ , \tag{2.41}$$

here, the parameter λ is the equivalent to the primary Fourier period of the mother wavelet and hence is different for each wavelet type, i.e., each choice of mother wavelet function implies different values of λ [109].

In this study, we utilize the analytical (complex) Morlet wavelet [139] for the analysis, as it is a very common choice in EEG data analysis. The reasons for that are its shape: it is Gaussian-windowed sinusoid, and also the fact that it is complex function, thus the resulting TFR also contains phase information (which might be desirable in some applications) and the energy distribution is smoother than the one produced by real wavelet functions. The analytical Morlet wavelet is formally defined as follows:

$$\psi_0(t) = \pi^{-1/4} \mathrm{e}^{i\omega_0 t} \mathrm{e}^{-t^2/2} \ , \tag{2.42}$$

where ω_0 is the center frequency of the wavelet. Here, $\omega_0 = 6$ in order to satisfy admissibility condition [43, 111, 139]. The primary Fourier period in Eq. (2.41) for this mother wavelet is $\lambda = 1.03$ s.

Despite the fact that scales can be chosen arbitrarily several general practices exist. For a given range of frequencies, one has to choose the corresponding smallest and the largest scales a_0 and $a_{\max}$, respectively. The smallest scale is defined through the largest frequency which is limited by the Nyquist frequency, i.e., half of the sampling frequency. In the present work we do not need such a high frequency limit, since the major activity of interest evolves at lower frequencies. For that reason we choose the maximum frequency to be $f_{\max} = 10$ Hz for the transient oscillations, $f_{\max} = 5$ Hz for the Lorenz system and $f_{\max} = 40$ Hz for the EEG data.

Similarly to the spectrogram, we can observe edge effects occurring at the borders of the TFR (cf. hatched areas in Figs. 2.7c and 2.7d); in wavelet-based analysis, it is common to refer to the affected regions as the cone of influence (COI), due to their shape. In wavelet analysis, the sliding time window changes with scale, i.e., with frequency, and the edge effect is larger for low frequencies than high frequencies due to the corresponding longer sliding time windows for low frequencies. Such nonlinear nature of boundary effect gives us an idea for estimating the largest scale, and thus the lowest frequency. Its value is chosen such that only 50% of the CWT coefficients will be affected by the edge effect (the approach adopted from [78]). The next step is to sample the scale values within the above limits. In this work, we study the influence of populating the scale vector on the SRSA and compare two different scale sampling schemes.

In a first method variant, the set of scales was constructed in such a way that the scale vector corresponds to the frequency vector defined for the spectrogram. We sampled the frequency range under study with a fixed frequency step Δf and converted the frequencies of number N_a into scales by inverting Eq. (2.41):

$$a_{n+1} = a_n \frac{f_{\max} - n\Delta f}{f_{\max} + (n+1)\Delta f} \ , \ \ n = 0, \ldots, N_a - 1 \ , \tag{2.43}$$

with the maximum frequency $f_{\max}$. This method corresponds to the conventional wavelet analysis.

In another variant, we are aiming to cover frequency domain in the most efficient way, as the wavelet's frequency support varies for each scale. To fulfill this requirement, we populate the scale vector by selecting subsequent scales based on the values of the previous scales [78]:

$$a_{n+1} = K a_n, \ \ n = 0, \ldots, N_a - 1 \ . \tag{2.44}$$

The factor K ensures that scale values cover sufficiently the frequency domain and can be found as:

$$K = \exp\left[\frac{\ln a_{\max} - \ln a_0}{N_a - 1}\right] \ , \tag{2.45}$$

where $a_{\max}$ and a_0 are the largest and the smallest scales, respectively, and N_a is the total number of scales. In the following, N_a is chosen as the number of frequency bins used in the spectrogram for comparison reasons.

To conclude the topic of time-frequency representations, Fig. 2.7 presents the four TFR discussed above and computed for an example signal. A quadratic chirp is selected as a test signal for its simplicity and a well known TFR. A chirp signal is a general notation for frequency-modulated signals. A chirp signal is defined by:

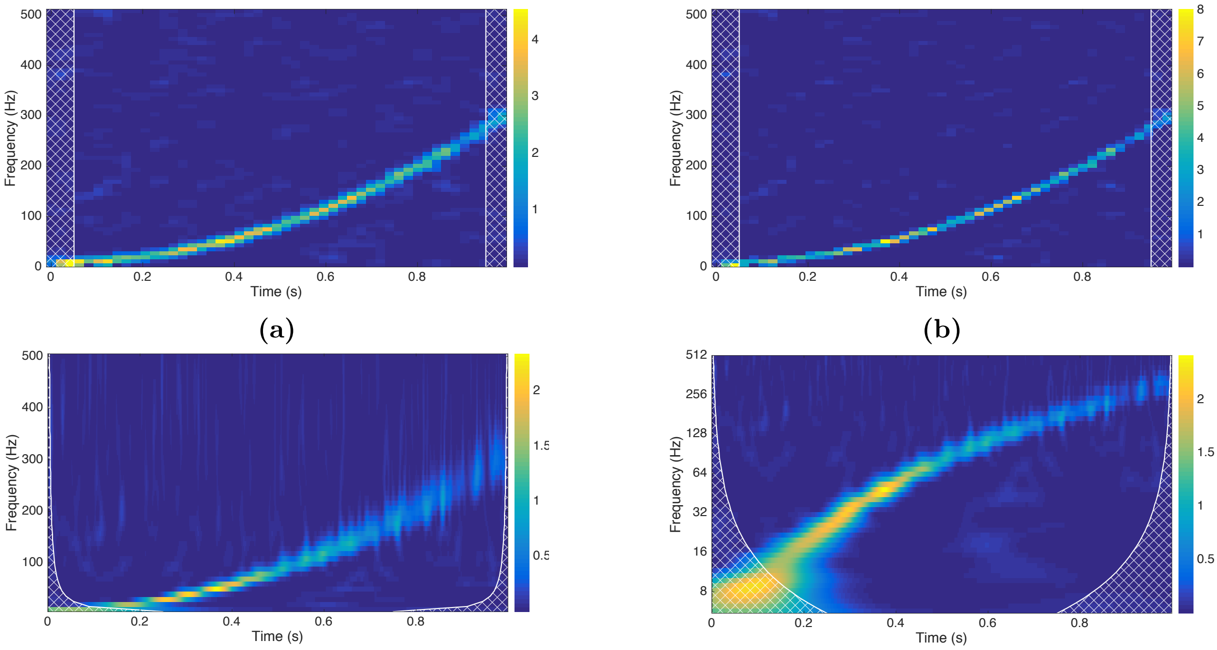
$$x_n = \sin(\phi_0 + 2\pi t_n f_n) + \xi_n \ \ , n = 0, \dots, N \ \ , \tag{2.46}$$

where ϕ_0 is the initial phase, ξ_n is AWGN, and f_n is the instantaneous frequency, which, for quadratic chirp, is given by:

$$f_n = f_0 + k t_n^2 \ \ , \tag{2.47}$$

where k is a coefficient defining the shape of the instantaneous frequency.

We can see from the above equations that the TFR of the quadratic chirp is presented with parabola. In this example, the chirp is defined with the following parameters: $\phi_0 = 0$, the observation time is 1 s sampled at 1024 Hz, $\xi_n \sim \mathcal{N}(0, 0.5)$, $f_0 = 5$ Hz, $k = 100$. So the expected TFR model resembles a parabola with $f = 5$ Hz at time $t = 0$ s and $f = 300$ Hz at $t = 1$ s.



(a) **(b)**

(c) **(d)**

Fig. 2.7: Four different time-frequency representation of the quadratic chirp embedded in white Gaussian noise defined by Eqs. (2.46) and (2.47). (a) Spectrogram. (b) Reassigned spectrogram. (c) Scalogram with equidistant frequencies. (d) Scalogram with nonlinear frequency vector. Hatched regions mark regions affected by the edge effect.

2.5 Complexity Measure

To quantitatively measure the obtained symbolic sequences we propose to measure their complexity. We present here three different complexity measures: the alphabet cardinality of the sequence (or the alphabet size), the well-known Lempel-Ziv (LZ) complexity, and the number of distinct words the sequence can be composed of. These complexity measures have in common the notion of complexity, which is the number of distinct elements required to encode a string of symbols. The more complex the sequence is the more elements are needed to represent it without redundancy. Despite the similarity of the three proposed measures, they capture different types of dynamics.

Alphabet Size

Our first choice of complexity measure is the size of the alphabet, with which the symbolic string is constructed. It is a logical representation of complexity of a signal, as the more complex the signal is, the more states will be encoded as a distinct symbol. Let us consider an example string $s = 0011110002220033000$, in this case, the alphabet is $A(s) = \{0, 1, 2, 3\}$, thus the complexity in this case is $C(s) = 4$. However, this approach is rather trivial and might not necessarily capture subtle differences in the signal's dynamics, for example, a strictly periodic signal can be encode with the same alphabet, but the order of the symbols differs. To capture such case with suggest to use another complexity measures.

Lempel-Ziv Complexity

A common complexity measure of a sequence is Lempel-Ziv complexity [94]. It was developed as a measure of "randomness" of a sequence, and it found an application in the domain of biomedical signal analysis [2, 71, 127]. This complexity measure is related to the number of distinct substrings and the rate of their occurrence along the symbolic sequence.

In order to obtain LZ complexity one parses an original string s into a set of substrings (or words), once a new substring is added the value of complexity c is incremented by one. Before we define the parsing procedure, let us first revise terms used in the process. A substring of s starting at ith symbol and ending at jth symbol is denoted by $s(i, j)$, if $i > j$ $s(i, j) = \{\}$ is the null set. Next, the vocabulary of the string $V(s)$ is a set of all substrings of the string s, for example $V(010) = \{0, 1, 01, 10, 010\}$. The parsing is done from left to right starting with the first symbol $s(1, 1)$, and during this process the substring $s(i, j)$ is compared to the vocabulary $V(s(1, j-1))$. If the substring is not present in $V(s(1, j-1))$, then we add it to the set of words, increase complexity by 1 and update the analyzed substring to $s(j+1, j+1)$ and the vocabulary to $V(s(1, j))$. If the substring is present, then update the substring to $s(i, j+1)$ and the vocabulary to $V(s(1, j))$. Repeat the previous steps until the end of the string. Sometimes the last word might be presented in the vocabulary of the preceding substring, but it is still added to the set of distinct words due to finiteness of the sequence (see example below).

Let us consider the previous example string $s = 0011110002220033000$. By following the described procedure we parse the string s into the set of words $W(s) = \{0 \cdot 01 \cdot 11 \cdot 10 \cdot 002 \cdot 22 \cdot 003 \cdot 30 \cdot 00 \cdot\}$ and the complexity measure is $C(s) = 9$. We can see, that this method of parsing sometimes extracts words by grouping symbols representing different type of activity, e.g., 002 joins transient states (which denoted by $s_i = 0$ in the current work) and a stable state (here, $s_i = k,\ k > 0$). The next method in this section is a modification of Lempel-Ziv complexity, where words are extracted based on the dynamics they represent.

Number of Words

The complexity measure discussed here is defined as number of unique words extracted from a sequence, similarly to Lempel-Ziv complexity. The only difference is that we define a word as a consecutive sequence of the same symbol. Let us illustrate it on the sequence $s = 0011110002220033000$. The extracted words are $W(s) = \{00 \cdot 1111 \cdot 000 \cdot 222 \cdot 00 \cdot 33 \cdot 000 \cdot\}$. Here, the total number of words is 7 and the number of unique words is 5. Hence, the complexity value is $C(s) = 5$.

The three described complexity measure are selected as features for further classification. In addition to these three features we also add two more: optimal threshold distance ε^* and maximum value of symbolic entropy $h(\varepsilon^*)$ as defined in Eq. (2.18) and Eq. (2.19), respectively. The reasoning for that choice is that these two parameters are estimated from the data and can characterize it. In addition, these two parameters have been successfully applied to classification of brain states in epileptic patients from neural signals [147]. To bring this topic to a conclusion, we provide a list of all features obtained with symbolic recurrence structure analysis in Table 2.2 alongside notations used throughout the manuscript.

Table 2.2: Features generated during the symbolic recurrence structure analysis. First column shows the notation used to refer to the feature further in the manuscript. Second column provides the description of the feature.

Notation	Description
`alphabet`	alphabet cardinality of the symbolic string
`lz`	Lempel-Ziv complexity of the symbolic string
`words`	number of words constructed of the same symbol in the symbolic string
`entr`	maximum symbolic entropy, correspond to $h(\varepsilon^*)$ in Eq. (2.19)
`eps`	optimal threshold distance, correspond to ε^* in Eq. (2.18)

Examples of recurrence-based symbolic representation of synthetic and the EEG data, as well as distributions of the features are presented in the Results section (cf. Section 2.6.2).

2.6 Results

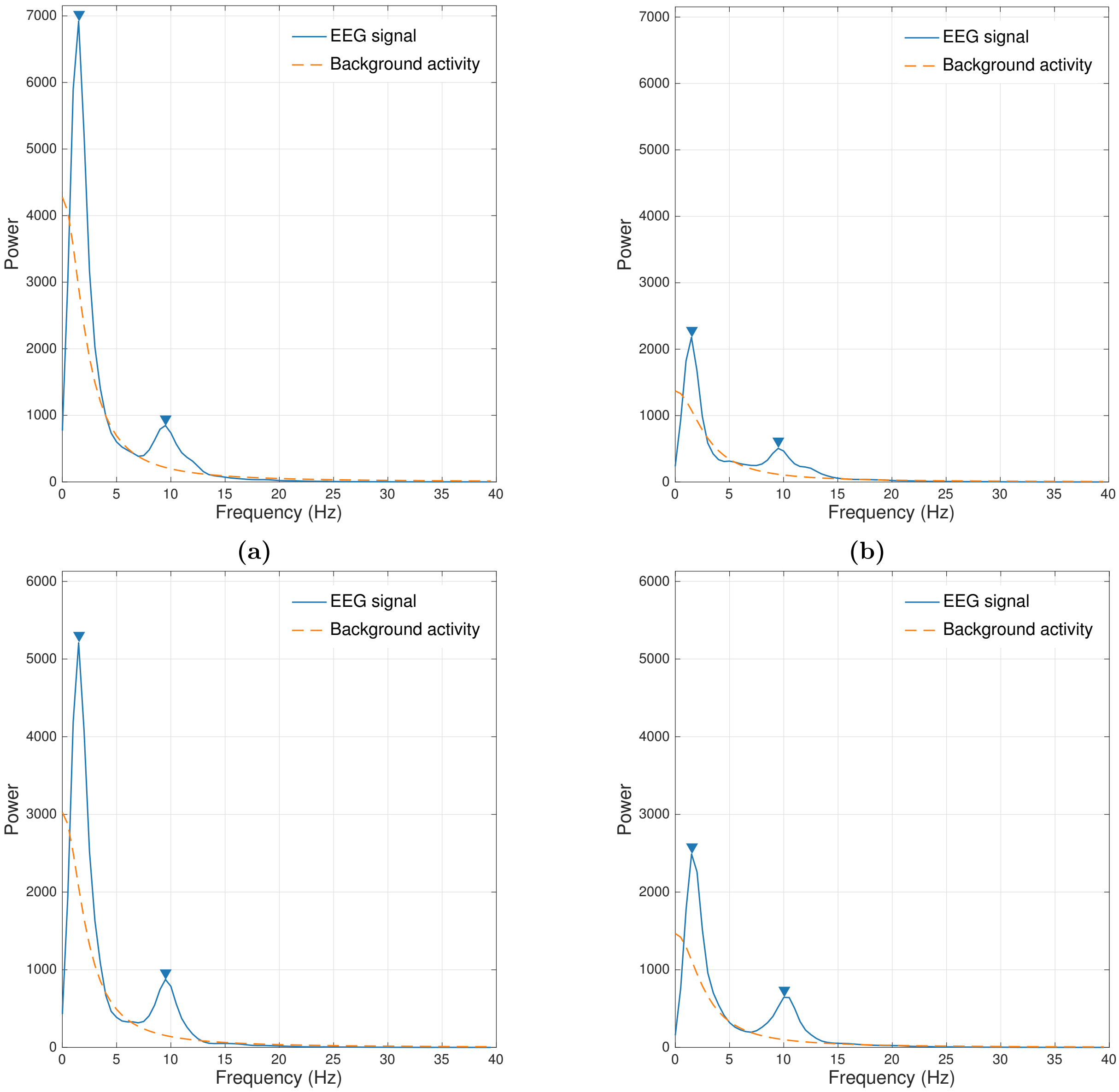
This section presents results of the feature construction part of our work. It is divided into two parts: the first part presents the features obtained with the power spectral analysis. And the second part presents the features constructed with the symbolic recurrence structure analysis.

2.6.1 Power Spectral Analysis

We start by presenting the results of the power spectral analysis. The features extracted at this stage include various measures of oscillatory power contained in the signal and the features arising from description of the EEG's background activity. More detailed explanation of the methods and the features are presented in Section 2.2.

Figure 2.8 shows examples of PSD estimates and the fitted PSD of background activity of EEG signals recorded during a surgery with propofol or desflurane. The corresponding EEG signals are shown on Fig. 2.2. We can see that the fitted curve follows well the general $1/f$ shape of the EEG's PSD. This curve models the PSD of the Ornstein-Uhlenbeck process and has two parameters: $2D$ and γ^2. The former controls intensity of the driving noise and the latter is inversely proportional to the correlation time of the process. These parameters affect the shape of the fitted curve in the following way: $2D$ controls the y-intercept of the curve and is indirectly related to the magnitude of the δ-peak, whilst γ^2 controls how fast the curve is decaying and can be indirectly related to the cut-off frequency of the EEG's power spectrum.

Figures 2.9 and 2.10 show boxplots of every feature obtained with the power spectral analysis. A boxplot is one of the ways to visualize the distribution of data. The box's edges mark the first and the third quartiles of the data, the horizontal line on the box marks the median, the whiskers



(a) (b) (c) (d)

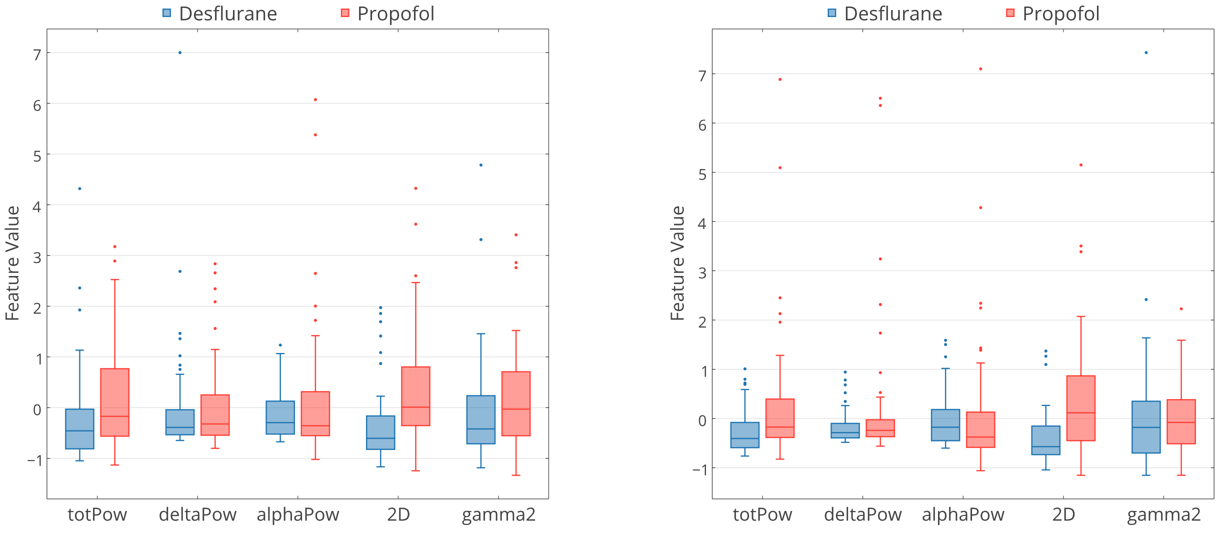
Fig. 2.8: The PSD estimates for EEG signals recorded during general anesthesia (shown in Fig. 2.2). The blue solid line shows the PSD of the signal, the dashed orange line shows the fitted model of the background activity's PSD, the blue triangles indicate location of δ- and α-peaks. (a) Desflurane pre-incision; (b) desflurane post-incision; (c) propofol pre-incision; (b) propofol post-incision.

indicate the range of the data, and the points above or below the whiskers indicate possible outliers. The Y-axis of the plots presents z-scores of the features. z-score or standardization can be computed as following:

$$z = \frac{x - \mu}{\sigma}, \tag{2.48}$$

where x is the values of a feature, μ is its mean value, and σ is the standard deviation from the mean. The standardization is carried out in order to overcome the problem of different features having different ranges of values.

By visually inspecting Figs. 2.9 and 2.10, we can see that the features describing desflurane and propofol groups expose larger differences than the features describing pre- and post-incision groups. For desflurane/propofol the larger difference is found in total power of the signal and $2D$ parameter of the OU process. These changes are consistent both in pre-incisional signals and post-incisional ones. In case of pre-/post-incision groups, we see that the distributions of the features do not differ strongly, the power contained in alpha oscillations appears to be the most discriminative features.



(a) **(b)**

Fig. 2.9: Boxplots of the distribution of the PSD features between two classes: desflurane (shown in blue) and propofol (shown in red) in (a) pre-incision and (b) post-incision signals. Y-axis shows z-scores of the features.

In order to quantitatively measure the differences between the values of the features we apply the two-sample Kolmogorov-Smirnov (KS) statistical test. The KS test is a nonparametric test, which evaluates a distance between empirical distribution functions of two samples. If this distance is larger than a critical value, we reject the null hypothesis that the two samples are drawn from the same distribution. When comparing two samples the main advantage of the KS test is that it takes into account both shape and location of the empirical distribution functions [56]. The results of the two-sample Kolmogorov-Smirnov test are presented in Table 2.3.

We can see that the results of the KS test are similar to the visual inspection of the boxplots shown in Figs. 2.9 and 2.10. Table 2.3 shows that the largest difference between desflurane/propofol groups is found in $2D$ parameter, which is indirectly related to the δ-power, the same results can be observed in Fig. 2.9: we see that the values of $2D$ are smaller and expose

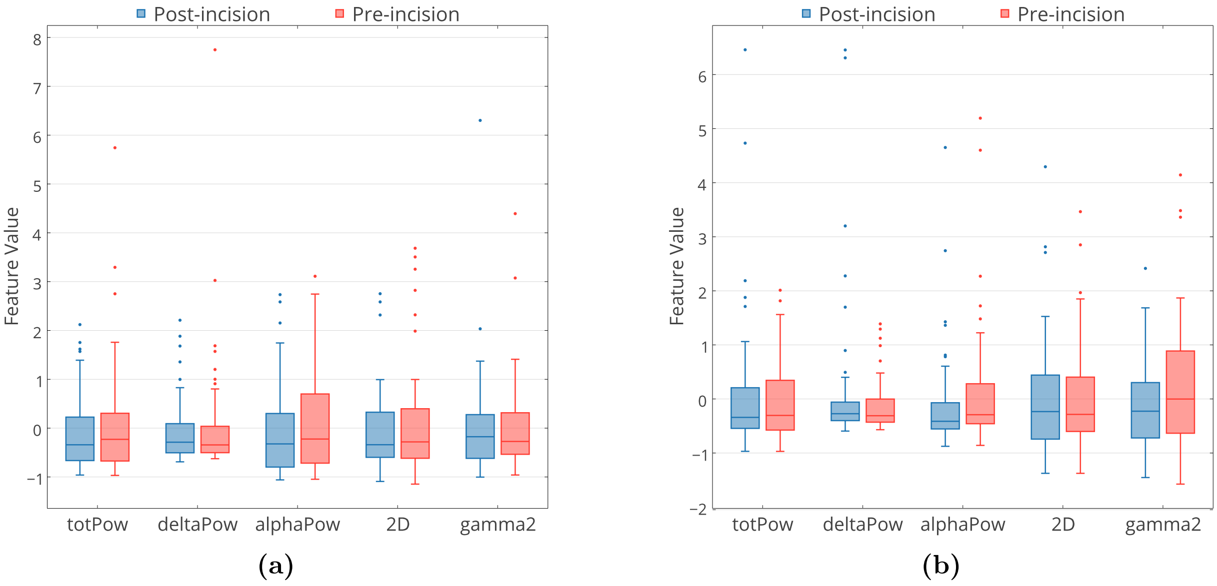


Fig. 2.10: Boxplots of the distribution of the PSD features between two classes: pre-incisional (shown in red) and post-incision (shown in blue) in (a) desflurane and (b) propofol signals. Y-axis shows z-scores of the features.

Table 2.3: Computed p-values for the Kolmogorov-Smirnov test for the features extracted during the power spectral analysis. The test is computed for each feature separately, where two samples are the values of a candidate feature belonging to each class. First column shows two classes, which are being compared, and the remaining five columns show the analyzed features. Significant differences between two samples are marked by bold values with $p < 0.05$.

	Feature				
Classification scenario	`totPow`	`deltaPow`	`alphaPow`	2D	`gamma2`
desflurane/propofol in pre-incision	**0.026**	0.271	0.178	$<$ **0.001**	0.154
desflurane/propofol in post-incision	**0.007**	0.429	**0.048**	$<$ **0.001**	0.818
pre-/post-incision in desflurane	0.970	0.998	0.735	0.735	0.998
pre-/post-incision in propofol	0.999	0.477	0.064	0.629	0.160

more compact distribution in desflurane than in propofol. The total power of the PSD estimates also expose significant difference and the results are similar to $2D$ parameter: values of power in propofol are larger and more spread. In addition, alpha power exposes statistically significant difference but only in post-incisional signals, revealing smaller values in propofol. In case of pre-/post-incision groups, as expected from inspecting Fig. 2.10, the Kolmogorov-Smirnov test do not reveal any significant difference between the features.

2.6.2 Symbolic Recurrence Structure Analysis

This section presents the features constructed with the symbolic recurrence structure analysis. As the SRSA is a relatively recent method and, to our knowledge, has not been applied to the incisional EEG data. Thus, before we apply the SRSA to the EEG data, whose dynamics are not known a priori, we we will first demonstrate the effectiveness of the method on the artificial data which was designed to model dynamics of brain activity.

Artificial Data

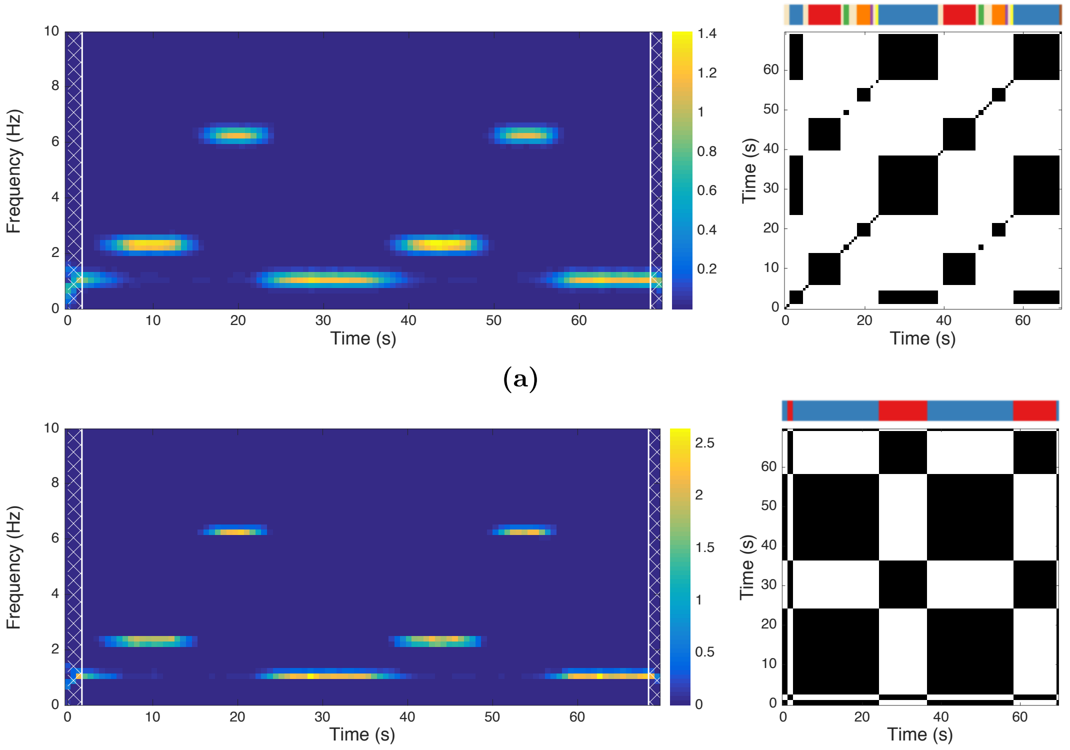
In this subsection we present application of the SRSA to the synthetic data described in Section 2.1.2. Figures 2.11 and 2.12 show the results of the SRSA for the transient oscillations dataset shown in Fig. 2.3. Figure 2.11 shows the analysis performed with conventional and reassigned spectrograms, while Fig. 2.12 shows the results obtained with two variations of the scalogram: with linearly and nonlinearly spaced frequencies.

We can see that the spectrogram-based symbolic sequence and RP are able to detect three types of activity, which are shown in blue, red, and orange colors in the symbolic sequence and as three rectangles of different size in the symbolic recurrence plot. We can also see that the 6.3 Hz oscillations are presented with the symbolic domains of a shorter duration than they actually are, and also several additional states are created (shown in green, purple, and yellow). The reassigned spectrogram fails to capture the underlying dynamics, creating only two large metastable states (red and blue) without transient states. One of the reasons for that might be that the process of reassignment of spectrogram values from one location to another, increase the sharpness of the TFR but also increases its variance.

Similarly to reassigned spectrogram, scalograms cannot precisely capture the dynamics underlying the data as it tends to neglect higher-frequency oscillatory components. On the symbolic sequences constructed with two versions of the scalograms, only two main states at lower frequencies are detected. Scalogram with equally spaced frequencies (cf. Fig. 2.12a) creates two symbols representing 1 Hz and 2.25 Hz oscillations while neglecting 6.3 Hz component. The wavelet transform with non-equidistant frequencies (cf. Fig. 2.12b) separates two low-frequency oscillations from the 6.3 Hz oscillations.

There might be several reasons for that. The primary reason of the neglect of components at larger frequencies is the well-known $1/f$ scaling of the power intrinsic to the wavelet spectrum, i.e., wavelet power decreases with higher frequencies due to the spread of the wavelet's frequency support while preserving the total power. Moreover, in wavelet analysis, high-frequency components are not resolved well in frequency (due to the Heisenberg principle), hence the power peak is spread over several frequencies. In addition, COI of the scalogram, unlike the one of the spectrogram, does not have a fixed width and varies with frequency. In literature, regions affected by edge effects are said to be "treated with care", therefore they might account for the disturbances in the final symbolic encoding.

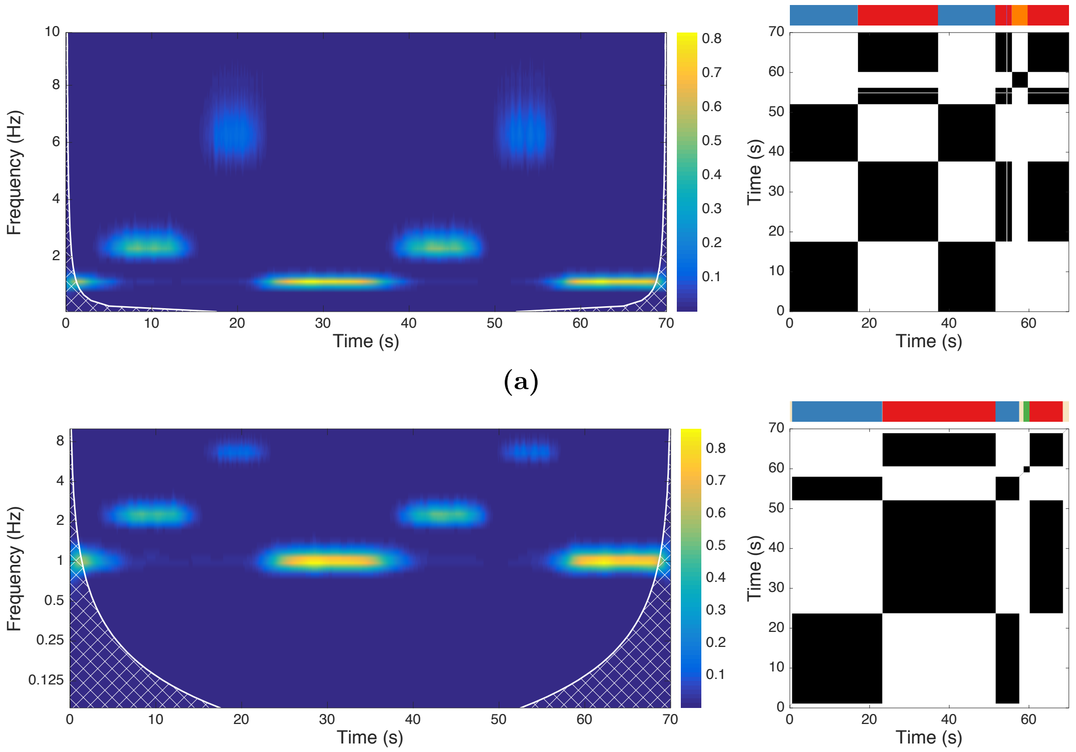
Next, we present results for the Lorenz system. Solutions of the Lorenz system reveals chaotic



(a)

(b)

Fig. 2.11: The SRSA results for the transient oscillations. (a) spectrogram; (b) reassigned spectrogram. On each panel, left: time-frequency representation; right: RPs with corresponding symbolic sequences above them. The colors in the symbolic sequences reflect a metastable state and transitions between the states, i.e., $s_i = 0$, are shown in beige color. In the time-frequency representations hatched regions mark the cone of influence reflecting edge effects.



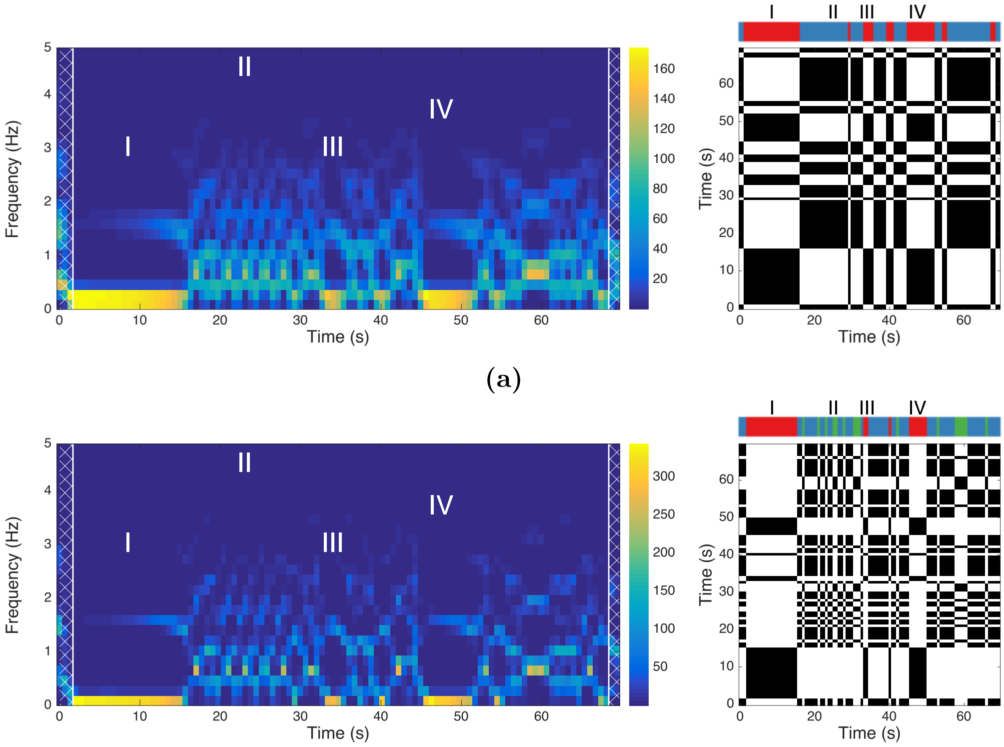
(a)

(b)

Fig. 2.12: The SRSA results for the transient oscillations. (a) scalogram with equidistant frequencies; (b) scalogram with non-equidistant frequencies. On each panel, left: time-frequency representation; right: RPs with corresponding symbolic sequences above them. The colors in the symbolic sequences reflect a metastable state and transitions between the states, i.e., $s_i = 0$, are shown in beige color. In the time-frequency representations hatched regions mark the cone of influence reflecting edge effects.

behavior, that exhibits metastable oscillatory states. Figures 2.13 and 2.14 show four TFRs of the Lorenz signal. On every time-frequency representation one can identify visually the four signal states I to IV marked in Fig. 2.4.

The symbolic sequence extracted from the spectrogram (cf. Fig. 2.13a) correctly identifies the time windows of the signal states I to IV. The states I, II, and IV are well captured, whereas the short state III is not well identified. The results for the reassigned spectrogram are similar (cf. Fig. 2.13b). Moreover, it creates additional states (shown in green) and the time windows of the states II, III are not defined very precisely.

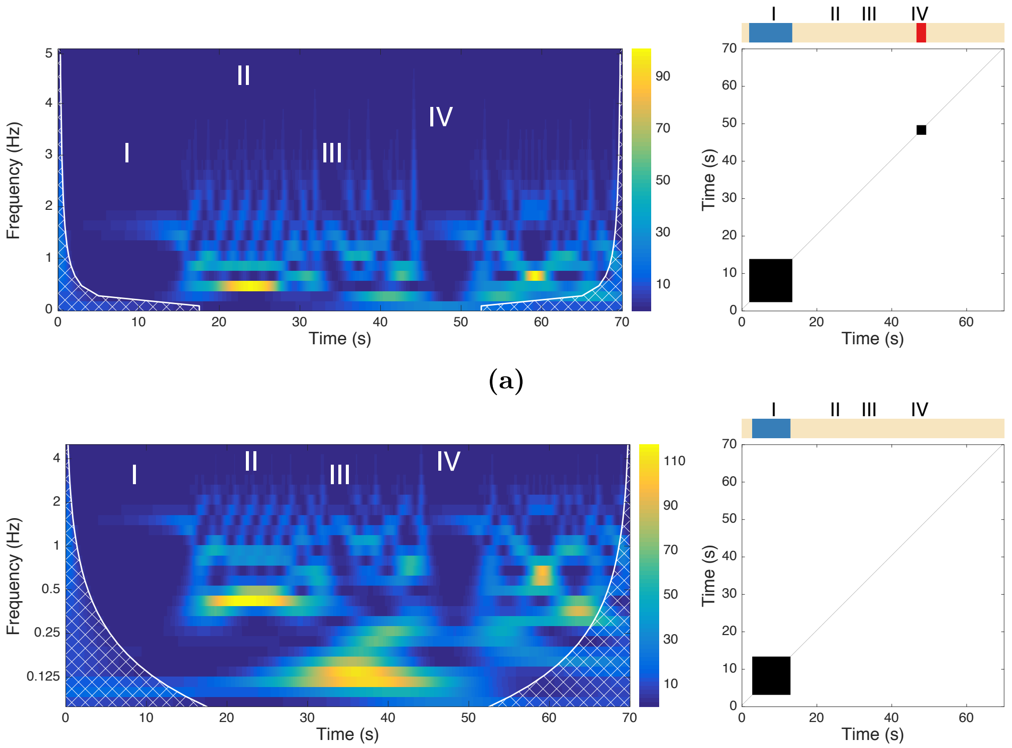


(a)

(b)

Fig. 2.13: The SRSA results for the Lorenz system. (a) spectrogram; (b) reassigned spectrogram. On each panel, left: time-frequency representation; right: RPs with corresponding symbolic sequences above them. The colors in the symbolic sequences reflect a metastable state and transitions between the states, i.e., $s_i = 0$, are shown in beige color. In the time-frequency representations hatched regions mark the cone of influence reflecting edge effects.

The symbolic sequences obtained with two variants of scalogram are shown on Fig. 2.14 fail to identify states II and III completely. And the symbolic domains extracted with wavelet-based methods are worse than the once extracted with Fourier-based methods and express no recurrence. A reason for such a poor detection, in addition to earlier mentioned reasons, can be found by examining the low-frequency parts of the corresponding time-frequency representations. It is known that the scalogram can not estimate power around $f = 0$ Hz, since this would correspond to an infinite scale due to $f \sim 1/a$. Taking a closer look at the Lorenz time series in Fig. 2.4, one observes that the “wings” of Lorenz system are not only defined by the frequency of the oscillations but also by the bias of these oscillations, despite the fact that the whole signal is zero-mean. Since wavelet-based techniques cannot resolve this constant shift, they miss an important feature and hence cannot properly handle the signal's dynamics.



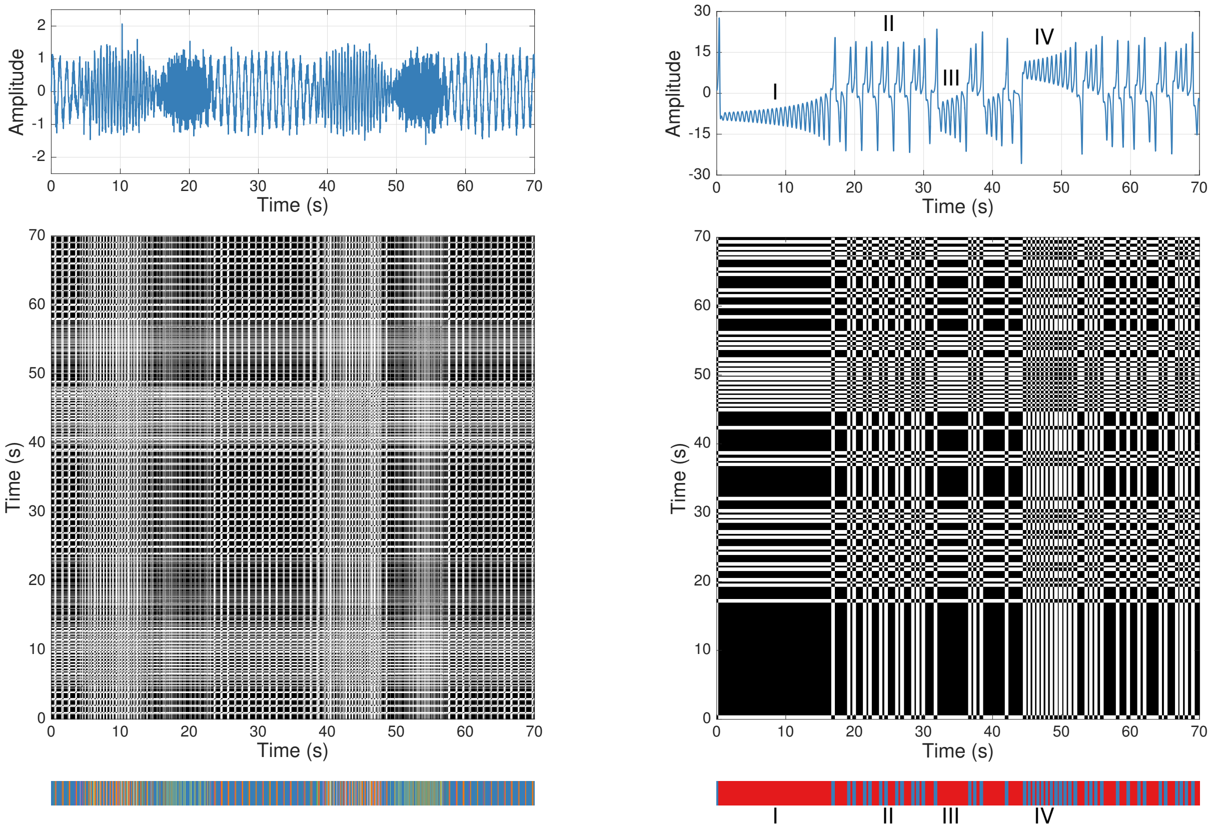
(a)

(b)

Fig. 2.14: The SRSA results for the Lorenz system. (a) scalogram with equidistant frequencies; (b) scalogram with non-equidistant frequencies. On each panel, left: time-frequency representation; right: RPs with corresponding symbolic sequences above them. The colors in the symbolic sequences reflect a metastable state and transitions between the states, i.e., $s_i = 0$, are shown in beige color. In the time-frequency representations hatched regions mark the cone of influence reflecting edge effects.

Finally, we present symbolic encodings of the two synthetic signals, when using delayed embeddings as the reconstructions step in the SRSA. The resulting RPs and symbolic sequences are shown on Fig. 2.15. We can observe, that for the transient oscillations the SRSA with delayed embeddings fails to identify underlying recurrence domains (cf. Fig. 2.15a). In case of Lorenz system, the SRSA extracts two interchanging metastable states, however, the dynamics of these states is not reflected properly. We notice that the SRSA with the delay embedding tends to extract stable states based on the amplitude values of the signal. This is especially noticeable for the transient oscillations, where symbolic sequence consist of almost 30 unique symbols instead of expected 3 for each type of activity.

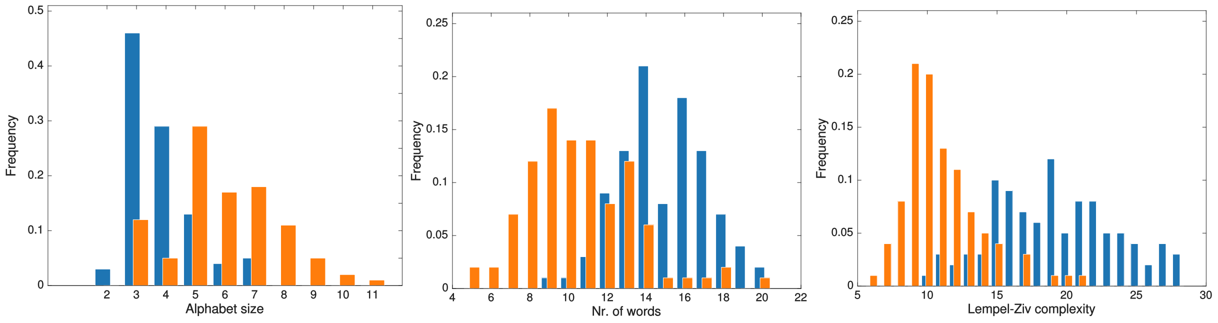
To conclude the analysis of the synthetic data we demonstrate that the proposed complexity measures might serve as quantitative feature of different types of activity. To do so we generated a set of 100 trials of each type of synthetic signals with random initial conditions and computed distributions of the complexity measure with the two-sample Kolmogorov-Smirnov. The initial conditions are drawn from normal distribution, i.e., $\mathbf{x}_0 \sim \mathcal{N}(\mu, \sigma^2)$, the parameters of the distribution are: $\mu = [1, 0, 0]$ and $\sigma = 0.2$ for transient oscillations and $\mu = [20, 5, -5]$ and $\sigma = 15$ for the Lorenz system.



(a) **(b)**

Fig. 2.15: The SRSA results for the synthetic data obtained with delay embeddings. (a) Transient oscillations, reconstruction parameters: $m = 5$, $\tau = 0.1$ s; (b) Solutions of the Lorenz system, reconstruction parameters: $m = 3$, $\tau = 0.18$ s.

The recurrence symbolic complexities obtained with all four time-frequency representations and delayed embeddings show significant difference according to Kolmogorov-Smirnov test ($p < 0.05$). Figure 2.16 shows an example of distributions of three proposed complexity measures obtained with the SRSA and the spectrogram. The distributions for other phase space reconstruction methods are similar (not shown here for the sake of brevity).



(a) **(b)** **(c)**

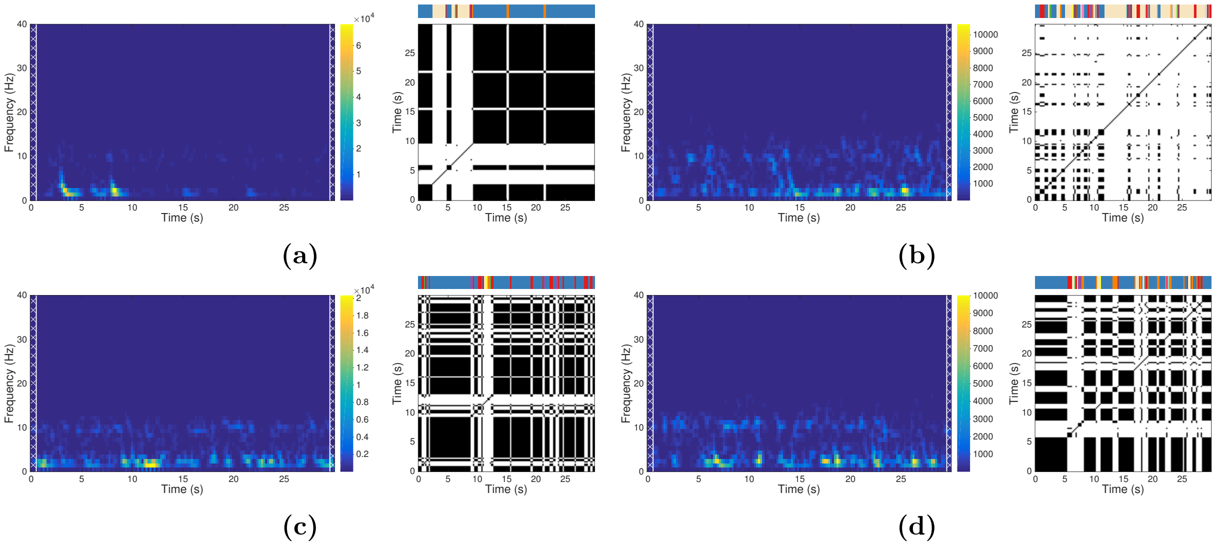
Fig. 2.16: Three complexity measure distributions for Lorenz system (blue) and transient oscillations (orange). The complexities are obtained using the spectrogram as phase space reconstruction in the SRSA. (a) Alphabet size; (b) number of words; (c) Lempel-Ziv complexity.

EEG Data

In this subsection we present the SRSA results for experimental data. In the previous subsection we have demonstrated the method on the synthetic data using four different TFRs and delayed embeddings in the phase space reconstruction step. The results reveal that time-frequency embeddings is more suitable for the SRSA of oscillatory signals. For that reasons we perform analysis of the EEG data using only time-frequency embeddings. In addition, the four TFRs consist of variants of only two essentially different TFRs: spectrogram and scalogram. In order to avoid redundancy, we present EEG analysis with only these two phase space reconstructions methods. We select conventional spectrogram and scalogram (with equidistant frequencies), as they demonstrated better performance than their variants. This could be seen by visually inspecting the obtained symbolic sequences of the synthetic data presented in the previous subsection. In the case of EEG, as its precise dynamics are not known a priori, we used Kolmogorov-Smirnov test, which has shown that the conventional spectrogram outperforms its reassigned version and the conventional scalogram demonstrate better results than the scalogram with non-equidistant frequencies (see Table 2.4).

We start by demonstrating the TFRs, symbolic RPs and corresponding symbolic dynamics of the EEG data shown in Fig. 2.2. Figure 2.17 shows the results of the SRSA with the conventional spectrogram, while Fig. 2.18, with scalogram with equidistant frequencies. Observing the TFRs we can see that the main EEG activity is concentrated mainly in δ- and α-bands (1–4 Hz and 8–12 Hz, respectively). This oscillatory activity is a typical feature of deep anesthesia found in frontal EEG electrodes and was also found in number of previous studies (e.g., [62, 95, 133]).

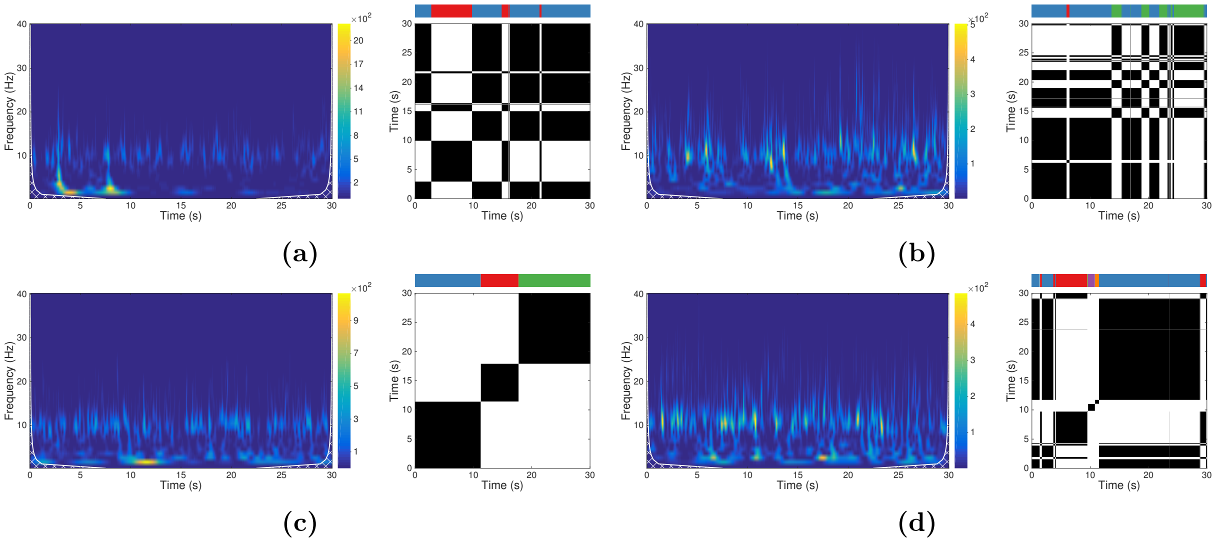
Figure 2.17 shows well-localized spectral power in both bands of interest, while wavelet-based transform in Figure 2.18 revels spread of α-activity and good localization of activity in δ-band. Hence, we can expect the wavelet-based recurrence analysis to capture primarily the temporal structure at lower frequencies, i.e., in the $\delta-$band. This leads to a coarser recurrence structure



(a) **(b)**

(c) **(d)**

Fig. 2.17: The SRSA results (conventional spectrogram) for the EEG data. (a) Desflurane pre-incision; (b) desflurane post-incision; (c) propofol pre-incision; (d) propofol post-incision.



(a) **(b)**

(c) **(d)**

Fig. 2.18: The SRSA results (conventional scalogram) for the EEG data. (a) Desflurane pre-incision; (b) desflurane post-incision; (c) propofol pre-incision; (d) propofol post-incision.

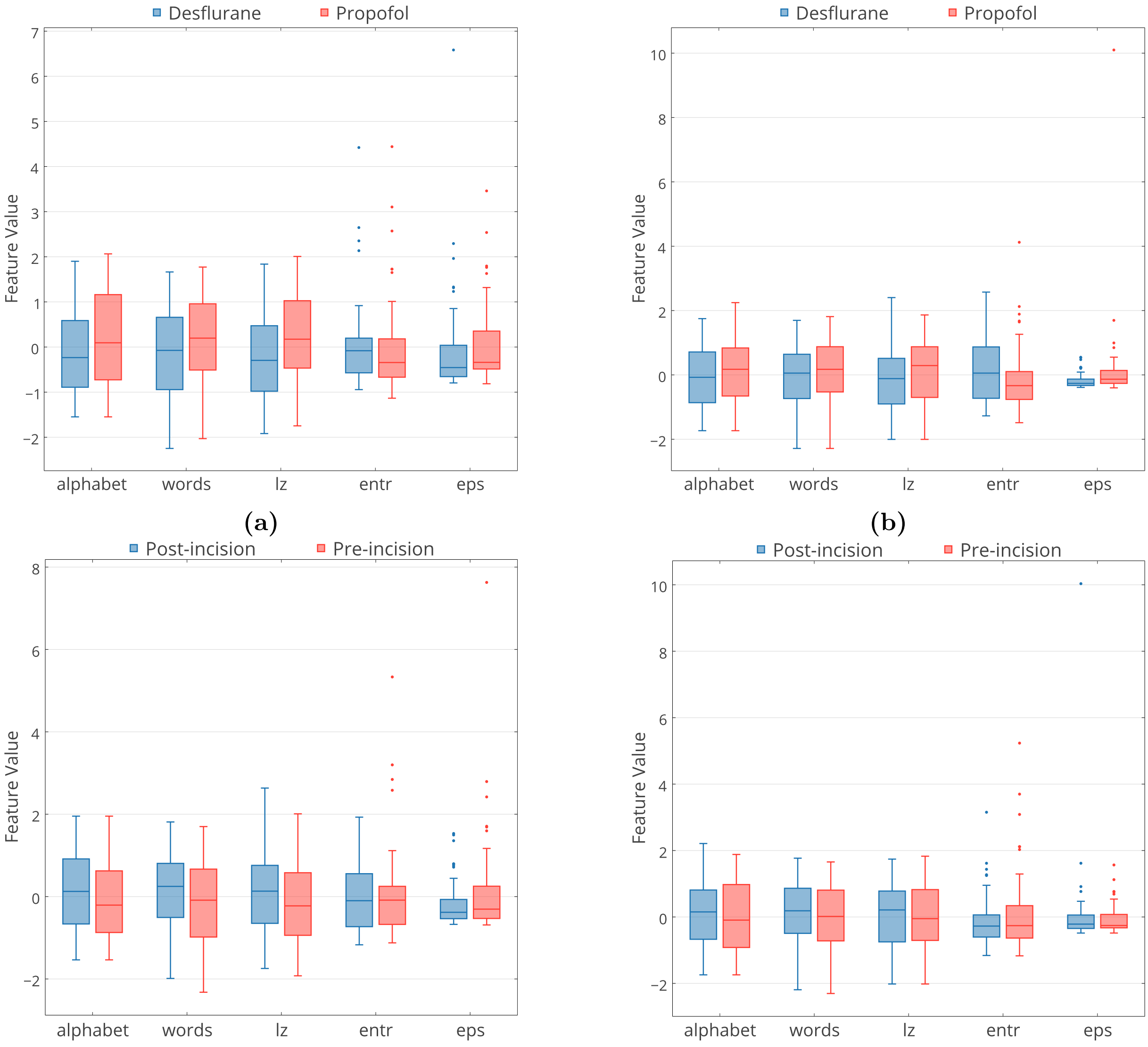
than for the spectrogram and hence sequences with less symbols, this can be seen in Figs. 2.17 and 2.18.

When comparing recurrence structures between desflurane and propofol or between pre- and post-incision, it is difficult to identify fundamental differences by visual inspection. It appears that the recurrence structure computed from scalogram exhibits longer recurrent periods in propofol data than in desflurane data. However, such a visual impression has to be validated statistically. For that we present below boxplots of the distribution of the recurrence features Figs. 2.19 and 2.20 and the results of the two-sample Kolmogorov-Smirnov test Table 2.4.

Comparing the distributions obtained with the spectrogram to the ones obtained with the scalogram, we can see that the largest difference between distributions is achieved with the scalogram. The same results are seen from Table 2.4. The SRSA with the scalogram reveals significant differences between desflurane/propofol groups. While in case of spectrogram-based features, only optimal threshold distance `eps` shows significant difference between desflurane and propofol in post-incisional EEG. The features do not show any difference in case of pre- and post-incision groups.

Table 2.4: Computed p-values for the Kolmogorov-Smirnov test for the features extracted during the symbolic recurrence structure analysis. The test is computed for each feature separately, where two samples are the values of a candidate feature belonging to each class. First column shows the phase space reconstruction method which was used in the SRSA, and the remaining five columns show the analyzed features. Significant differences between two samples are marked by bold values with $p < 0.05$.

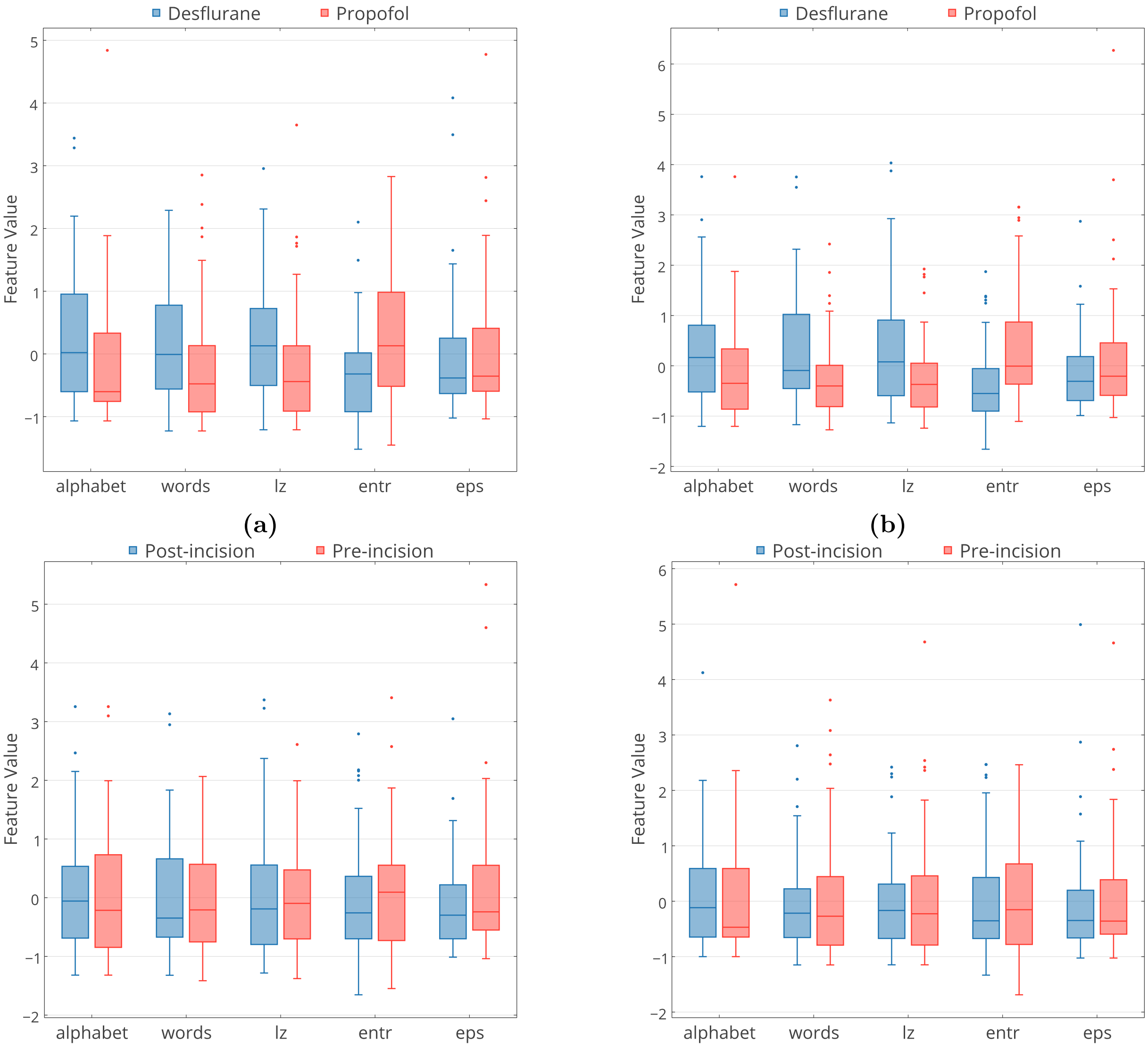
Embedding	`alphabet`	`words`	`lz`	`entr`	`eps`
			desflurane/propofol in pre-incision		
Delayed embeddings	0.19	0.22	0.15	**0.02**	0.70
Spectrogram	0.34	0.37	0.06	0.43	0.06
Reassigned spectrogram	0.06	0.10	0.19	0.07	0.13
Scalogram (linear frequencies)	0.05	**0.01**	**0.03**	**0.003**	0.67
Scalogram (nonlinear frequencies)	0.42	**0.02**	0.05	0.34	0.93
			desflurane/propofol in post-incision		
Delayed embeddings	**0.01**	0.07	0.39	$<$ **0.01**	**0.03**
Spectrogram	0.44	0.68	0.19	0.05	**0.02**
Reassigned spectrogram	0.57	0.75	0.91	**0.01**	0.06
Scalogram (linear frequencies)	**0.04**	**0.04**	**0.01**	$<$ **0.01**	0.34
Scalogram (nonlinear frequencies)	0.08	**0.01**	**0.01**	0.86	0.40
			pre-/post-incision in desflurane		
Delayed embeddings	0.13	0.13	0.99	0.13	0.20
Spectrogram	0.29	0.29	0.42	0.57	0.57
Reassigned spectrogram	0.88	0.57	0.29	0.57	0.74
Scalogram (linear frequencies)	0.97	0.99	0.88	0.19	0.73
Scalogram (nonlinear frequencies)	0.99	0.88	0.88	0.74	0.20
			pre-/post-incision in propofol		
Delayed embeddings	0.78	0.78	0.98	0.91	0.48
Spectrogram	0.63	0.63	0.91	0.91	0.91
Reassigned spectrogram	0.91	0.98	0.98	0.91	0.78
Scalogram (linear frequencies)	0.24	0.48	0.48	0.34	0.78
Scalogram (nonlinear frequencies)	1.00	0.91	0.98	0.78	0.98



(a) **(b)**

(c) **(d)**

Fig. 2.19: Boxplots of the distribution of the SRSA features obtained with the conventional spectrogram between classes: (a) desflurane/propofol in pre-incision; (b) desflurane/propofol in post-incision; (c) pre-/post-incision in desflurane; (d) pre-/post-incision in propofol. Y-axis shows z-scores of the features.



(a) **(b)**

(c) **(d)**

Fig. 2.20: Boxplots of the distribution of the SRSA features obtained with the scalogram (equidistant frequency) between classes: (a) desflurane/propofol in pre-incision; (b) desflurane/propofol in post-incision; (c) pre-/post-incision in desflurane; (d) pre-/post-incision in propofol. Y-axis shows z-scores of the features.

2.7 Discussion

In this chapter we perform feature construction. This process is carried out in two steps: first we analyze stationary properties of the EEG signals by power spectral analysis, next we generate features which might reflect nonlinear dynamical properties of the signal with the use of symbolic recurrence analysis.

Power spectral features, which reveal the most difference between desflurane and propofol groups are total power of a signal, power of α-activity of the EEG and the parameter of the background activity's model, proportional to δ-power, which is similar to the results found in [95]. In case of pre- and post-incision features we cannot achieve significant difference between the groups, but the largest difference was found in α-power of the signal relatively to the background activity and γ^2 parameter of the model of the EEG's background activity (it is indirectly related to the highest frequency present in the EEG).The former might be related to paradoxical arousal [84], while the latter might be related to the SEF as discussed in [62].

The symbolic recurrence structure analysis carried out on the synthetic data has shown the advantage of the proposed phase space reconstruction method based on the signal's time-frequency representation over delayed embeddings. The proposed method takes into account oscillatory behavior of the signal, focusing on the power of separate rhythms instead of analyzing amplitude values of the signal, which might be more prone to noise and random fluctuations. For synthetic data, consisting of nonstationary oscillatory signals, we observe that the spectrogram provides a good time-frequency representation and allows to extract the underlying recurrence structure of a signal. Reassigned spectrogram shows worse performance in the SRSA than the spectrogram. This might be related to the fact that the process of reassignment, providing an almost perfect energy localization in the TFR, increase its variance. A possible solution to this problem might be reassignment of TFRs with initially lower variance, such as multitaper spectrogram [146]. The scalogram performs bad on the synthetic data, however, shows much better results in case of experimental data. This might be related to the intrinsic property of the wavelet transform to scale power with frequency. Similar power scaling can be found in many biological signals, including EEG. While EEG's power spectrum expose $1/f$ behavior, the synthetic datasets have a uniform distribution of power over frequency.

A number of previous studies have extracted stationary neural activity, but very few have investigated fluctuations of spectral power [93, 145]. In this thesis we propose combination of time-frequency analysis of signal with dynamical system theory, which provides a measure of nonstationary processes.

The features constructed with the SRSA shows significant differences between desflurane/propofol groups in almost all the features obtained with scalograms with linearly spaced frequencies. We observe the opposite situation in features generated with spectrogram (and with other phase space reconstruction methods). A possible explanation is the fact that activity at low frequencies in the $\delta-$band exhibit unique recurrence structures. These unique recurrence patterns exist in spectrogram data as well, but the additional $\alpha-$activity may disturb the detection of the recurrence structure in the $\delta-$band and hence the recurrence structures involving both bands are blurred. In addition, three complexity measure obtained with the SRSA show higher values in desflurane, while symbolic entropy is lower. The complexity measures are directly related to number of recurrence states and different types of activity present in the signal, whilst symbolic entropy reaches its maximum when the distribution of these domains is uniform. Thus, it might indicate that propofol produce small number of metastable states, in which system spends almost equal amount of time. Whilst propofol produces larger number of states, but some of these states are more dominant than the others.

3
Classification

Contents

After generating features from the data, as described in the previous chapter, we proceed with classification. In the current work we consider only two-class problem, namely, our objective is to distinguish between pre- and post-incision, as well as between propofol and desflurane. We believe that it is a good practice to separate different experimental conditions, that is why we are going to sacrifice the number of data points in each class, in order to treat different type of EEG signals independently. Therefore, we are going to handle four cases:

1. classify desflurane/propofol in pre-incision signals;
2. classify desflurane/propofol in post-incision signals;
3. classify pre-/post-incision in patients receiving desflurane;
4. classify pre-/post-incision in patients receiving propofol.

The current chapter presents classification carried out using two classifiers: linear discriminant analysis and support vector machines. Linear discriminant analysis is a very basic and well-established method allowing to build linear class boundaries. Support vector machine is a more complex classifier, and, as a result, is known to be one of the most efficient supervised learning techniques. Even though support vector machine originally is a linear classifier, it allows the use of kernels, which produces complex nonlinear class boundaries. This classifiers can be used with many kernel functions, however, in this manuscript, we choose the radial basis function (also known as Gaussian) kernel, as the one yielding the best results.

We perform the classification in several steps. First, we analyze the features obtained with the power spectral analysis, as they reveal linear dynamics inside the system under consideration. Next, we use only the features constructed with the symbolic recurrence structure analysis, for they capture nonlinear behavior of the system. And, lastly, we combine the two feature subsets to analyze in which way it affects the classification accuracy.

In addition, to gain more insight on each feature's discriminative power, we perform feature ranking based on a test statistic. This method allows to order features by their importance in a classification procedure. Amongst many possibilities, we select Wilcoxon rank sum test as the most suitable one, more details on the method as well as on other feature ranking techniques are given below. The final results are obtained by performing classification using only the first three features with support vector machines.

3.1 Linear Discriminant Analysis

Linear discriminant analysis (LDA) is a generalization of Fisher's linear discriminant [46], which projects the original dataset onto a lower-dimensional space that yields better class separability. Fisher's discriminant operates by finding such a projection that minimizes the within-classes variability while maximizing the variability between classes, this idea is illustrated on Fig. 3.1. This technique can be used as a preprocessing step (dimensionality reduction) in other machine learning applications or as a self-contained classifier.

Classification with LDA utilizes linear discriminant functions, which map an input feature vector $\mathbf{x}$ to one of the existing K classes. In order to obtain such functions, let us recall Bayes rule:

$$p(k|\mathbf{x}) = \frac{p(\mathbf{x}|k)p(k)}{\sum_{l=1}^{K} p(\mathbf{x}|l)p(l)} \quad , \tag{3.1}$$

where $p(k|\mathbf{x})$ is the posterior probability that an observation $X = \mathbf{x}$ belongs to class k, $p(k)$ is the prior probability of class k with $\sum_{k=1}^{K} p(k) = 1$, and $p(\mathbf{x}|k)$ is the class-conditional probability of X in class k.

We can observe that estimating the posterior probability requires computation of the class prior probabilities and the value of $p(\mathbf{x}|k)$. The value of $p(k)$ can be estimated as a ratio of training samples belonging to class k to the total number of training samples:

$$\hat{p}(k) = \frac{N_k}{N} \quad . \tag{3.2}$$

However, estimation of $p(\mathbf{x}|k)$ requires an assumption regarding the underlying data distribution. LDA models each class conditional density as a multivariate Gaussian distribution in p-dimensions:

$$p(\mathbf{x}|k) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mu_k)^T \Sigma_k^{-1} (\mathbf{x} - \mu_k)^T\right] \quad , \tag{3.3}$$

where p is the number of features (or predictors), Σ_k is a covariance matrix of the kth class, μ_k is a mean vector of the kth class, and $|\Sigma_k|$ denotes the determinant of the Σ_k. The values of Σ_k and μ_k are estimated from the training data.

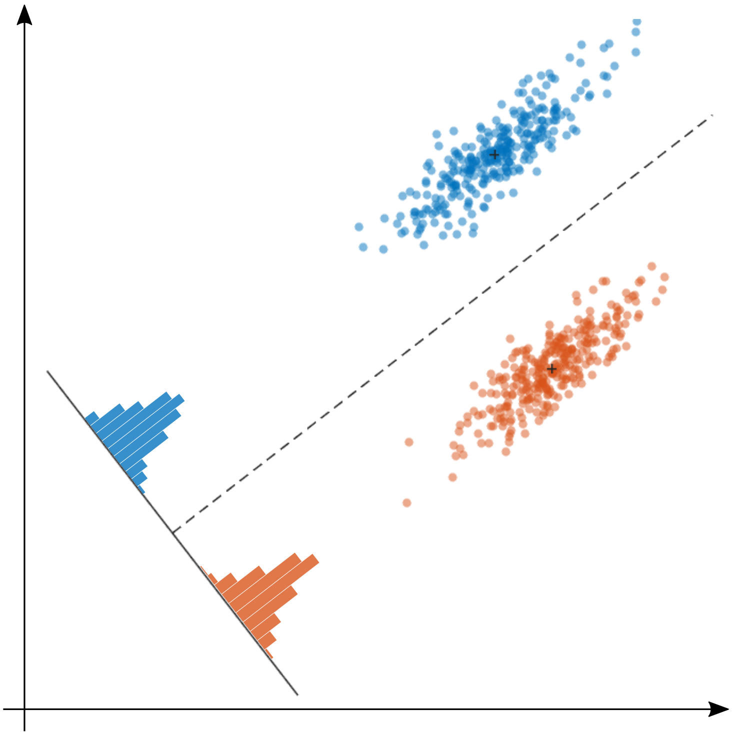


Fig. 3.1: Samples from two classes (shown in blue and orange) and the histograms resulting from projection obtained with Fisher's linear discriminant analysis, this projection ensures good class separability.

Another assumption made when applying LDA is the homogeneity assumption [73] that classes share common covariance matrix, i.e., $\Sigma_k = \Sigma, \ \forall k$. Without this assumption the decision boundaries are no longer linear and we are dealing with quadratic discriminant analysis (QDA).

By taking log-ratio of class the posteriors and rearranging the terms we obtain the discriminant functions [66]:

$$\delta_k(\mathbf{x}) = \mathbf{x}^T \hat{\Sigma}^{-1} \hat{\mu}_k - \frac{1}{2} \hat{\mu}_k^T \hat{\Sigma}^{-1} \hat{\mu}_k + \log \hat{p}(k) \ . \tag{3.4}$$

Now, a new observation can be assigned to one of the class based on the value of Eq. (3.4):

$$\hat{y}(\mathbf{x}) = \arg\max_k \delta_k(\mathbf{x}) \tag{3.5}$$

3.2 Support Vector Machines

Support Vector Machine (SVM) is an approach that was developed in the 1990s by Vapnik and co-authors [27, 37] and that can be used for regression or classification problems. Presently, it remains amongst the best "off-the-shelf" supervised classifiers. The success of the SVM is due to the idea of margin maximization behind it. The SVM not only separates two classes by a linear decision boundary, but also ensures that the boundary has the largest distance to the closest training observations. This distance is known as a *margin* and the observations which are the closest to the separating hyperplane are called *support vectors*. Support vectors play an important role since the model obtained during the training phase can be expressed solely in terms of support vectors. This makes the SVM less sensitive to the outliers located far from the decision boundary, thus margin maximization reduces the generalization error. Additionally, wide success of the SVM can be attributed to, so-called, “kernel trick”, which allows to combine the classifier with various nonlinear kernels producing more complex decision boundaries.

3.2.1 Margin Maximization

Let us consider a binary classification problem on a training set consisting of N observations:

$$S = \{(\mathbf{x}_i, y_i);\ i = 1, \ldots, N\} \ , \tag{3.6}$$

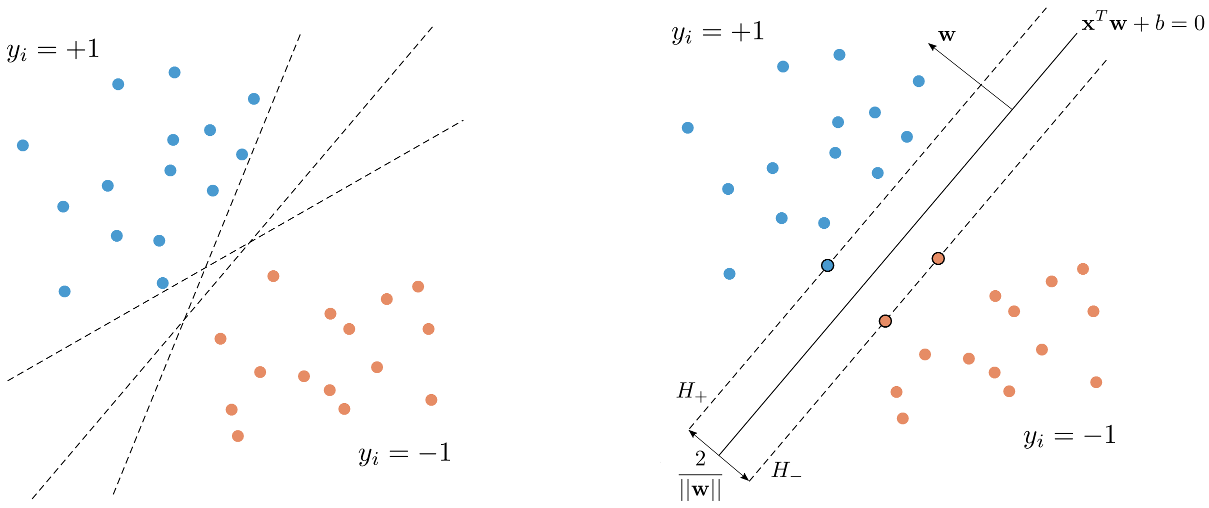
where each observation is $\mathbf{x}_i \in \mathbb{R}^p$ (p being number of features); and the labels are defined as $y_i \in \{-1, +1\}$.

Then, a hyperplane separating the two classes is defined by:

$$\{\mathbf{x}\colon\ f(\mathbf{x}) = \mathbf{x}^T\mathbf{w} + b = 0\} \ , \tag{3.7}$$

where $\mathbf{w}$ is the vector of weights and b is the bias.

New data can be classified in one of the two classes depending on which side of the hyperplane it lies. From many possible separating hyperplanes SVM finds the optimal one. Here, optimal is understood in a sense that a distance from the separating hyperplane to the closest observations from either of the classes is maximized. This distance is called *margin*, and the idea of margin maximization is illustrated in Fig. 3.2.



(a) **(b)**

Fig. 3.2: Support Vector Machines: margin maximization. The data in blue belongs to the positive class and the data in orange belongs to the negative class. (a) Black dashed lines show three out of many possible separating hyperplanes. (b) The black solid line indicates the optimal separating hyperplane defined by $\mathbf{x}^T\mathbf{w} + b = 0$, the dashed lines bound the maximal margin and shows H_- and H_+ hyperplanes.

If the data is linearly separable, then the following holds:

$$y_i(\mathbf{x}_i^T\mathbf{w} + b) \geqslant 1, \ \ i = 1, \ldots, N \ , \tag{3.8}$$

and we can define two hyperplanes H_{-1}: $\mathbf{x}^T\mathbf{w} + (b - 1) = 0$ and H_{+1}: $\mathbf{x}^T\mathbf{w} + (b + 1) = 0$. Points in the training set which lie on either H_{-1} or H_{+1} are referred to as *support vectors*. The distances from the support vectors to the separating hyperplane is given by:

$$\frac{y_i(\mathbf{x}_i^T\mathbf{w}+b)}{||\mathbf{w}||}=\frac{1}{||\mathbf{w}||} \ , \tag{3.9}$$

then we can define the margin as $d = \frac{2}{||\mathbf{w}||}$. In order to obtain optimal classifier, we need to maximize the margin subject to correct class labels. Then the constraints are given by Eq. (3.8), and the optimization problem becomes:

$$\begin{aligned} \min_{\mathbf{w},b} \quad & ||\mathbf{w}|| \\ \text{s.t.} \quad & y_i(\mathbf{x}_i^T\mathbf{w}+b) \geqslant 1, \quad i=1,\dots,N \ . \end{aligned} \tag{3.10}$$

In most of the cases data cannot be linearly separated, i.e., there is always an overlap between two classes. To deal with this problem we can relax the constraints in Eq. (3.10) by introducing positive *slack variables* ξ_i. If $\xi_i > 1$, then an observation $\mathbf{x}_i$ is lying on the wrong side of the margin. Thus by bounding the sum $\sum \xi_i$ by a constant, we define the upper bound of a training misclassification rate. By taking the slack variables into account and re-expressing Eq. (3.10) in a slightly different form, we obtain optimization problem for nonseparable case:

$$\begin{aligned} \min_{\mathbf{w},b} \quad & \frac{1}{2}||\mathbf{w}||^2 \quad + \quad C\sum_{i=1}^{N}\xi_i \\ \text{s.t.} \quad & \xi_i \geqslant 0, \quad y_i(\mathbf{x}_i^T\mathbf{w}+b) \geqslant 1-\xi_i, \quad i=1,\dots,N \ . \end{aligned} \tag{3.11}$$

This is a convex optimization problem: quadratic problem with linear constraints. This means that any local minimum is a global minimum and can be computed using Lagrange multipliers. One of the ways to solve this problem is to use the Lagrange primal function, however it is more common to use dual formulation [48]:

$$L_D(\alpha)=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j \left(\mathbf{x}_i^T\mathbf{x}_j\right) \ , \tag{3.12}$$

which we have to maximize subject to $0 \leqslant \alpha_i \leqslant C$ and $\sum_{i=1}^{N}\alpha_i y_i = 0$, where α_i is the Lagrange multipliers.

For brevity, we have to skip the details of solution of Eq. (3.12), and we present here only key ideas, however an interested reader might refer, for instance, to [29] for a more in-depth explanation and examples.

Lagrangian coefficients α_i are non-zero only for support vectors [25]. Since the points, for which $\alpha_i = 0$ do not contribute to the above sum, the solution of Eq. (3.12) is sparse and is defined by support vectors solely (see Fig. 3.2b). The constant C in constraints of Eq. (3.12) is referenced in the literature as *box constraint.* It plays an important role because it controls the amount of misclassified data while maximizing the margin. High values of C imply high training accuracy, however it makes the classifier more susceptible to the outliers. Figure 3.3 demonstrates how different values of box constraint affect classification results. This parameter is crucial to the predictive power of the model, that is why it has to be fine-tuned, in our work we use 5-fold cross-validation (see Section 3.3) in order to obtain optimal C value.

The optimal weights can be found as follows:

$$\hat{\mathbf{w}}=\sum_{i\in\mathcal{S}}\hat{\alpha}_i y_i \mathbf{x}_i \ , \tag{3.13}$$

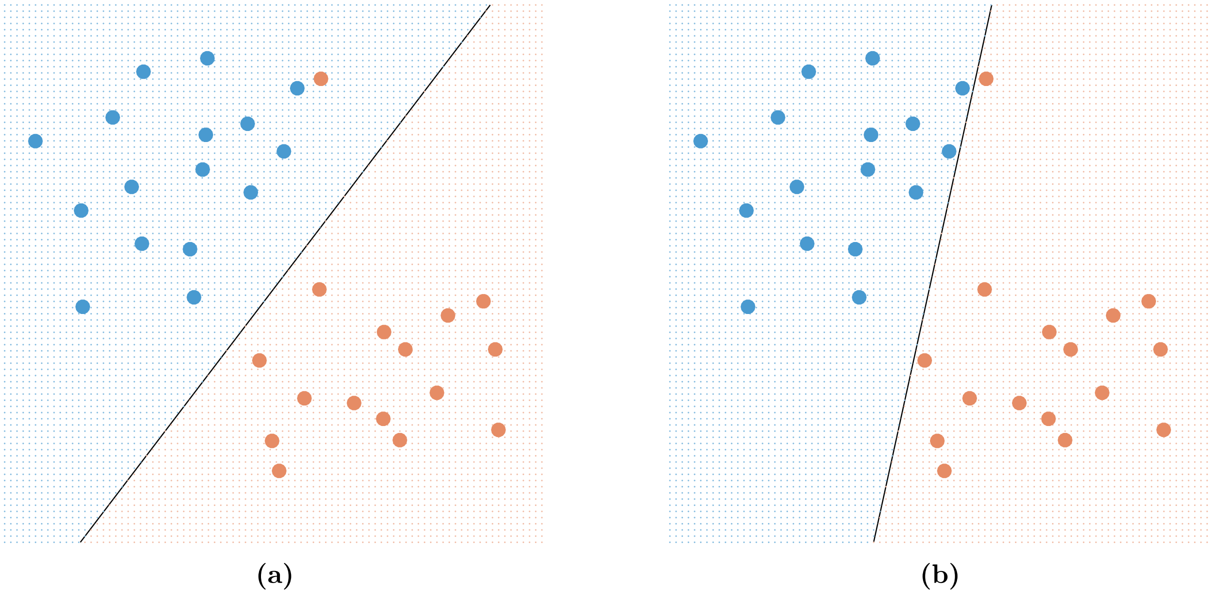


Fig. 3.3: Influence of box constraint values on SVM. The same data as shown in Fig. 3.2 but containing an outlier. The data in blue belongs to the positive class and the data in orange belongs to the negative class. The black solid line indicates the optimal separating hyperplane. The blue and orange grid indicates the decision rule made by the classifier. (a) $C = 1$; small box constraint maximizes margin while sacrificing training accuracy. (b) $C = 100$; large value of box constraint leads to high accuracy, however the model becomes less robust to outliers.

where $\hat{\alpha}_i$ solves the optimization problem, $\mathcal{S}$ is the set of indices of the training data which are the support vectors. While the weights $\hat{\mathbf{w}}$ are estimated explicitly during the training procedure, the bias term $\hat{b}$ is estimated from the Karush-Kuhn-Tucker conditions averaging over all support vectors [73].

Finally, we can write the decision rule for new data as:

$$\hat{y}(\mathbf{x}) = \operatorname{sign}\Big[f(\mathbf{x})\Big] = \operatorname{sign}\Big[\mathbf{x}^T\hat{\mathbf{w}} + \hat{b}\Big] \ . \tag{3.14}$$

3.2.2 Use of Kernels in SVM

When the decision boundary between classes is nonlinear the SVM can be successfully applied by transforming original feature space with a technique known as "kernel trick" [37]. The main idea behind this method is to compute a nonlinear transformation of original features without actually having to compute the transformation itself (which can have infinite dimensions), but instead computing only their inner product.

To demonstrate it let us first re-write $f(\mathbf{x})$ in Eq. (3.14) in a slightly different way:

$$f(\mathbf{x}) = b + \sum_{i \in \mathcal{S}} \alpha_i y_i \langle \mathbf{x}, \mathbf{x}_i \rangle \ , \tag{3.15}$$

where $\langle \cdot, \cdot \rangle$ denotes inner product.

We can replace the inner product in the above formula with a *kernel* $K = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$, where $\phi(\mathbf{x})$ is a nonlinear transformation of the input feature vector. We do not need to know the transformation itself, only the values of the kernel. A function, in order to be a valid kernel, should be a symmetric positive (semi-) definite (for more details see [29, Section 4.1]). There are several common types of kernel functions, here are some of them:

$$\begin{aligned}
&\text{Linear:} && K(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle \ , \\
&\text{Polynomial of degree } d\text{:} && K(\mathbf{x}_i, \mathbf{x}_j) = (1 + \langle \mathbf{x}_i, \mathbf{x}_j \rangle)^d \ , \\
&\text{Radial basis function:} && K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left[-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma} \right] \ , \\
&\text{Neural Network:} && K(\mathbf{x}_i, \mathbf{x}_j) = \tanh\left(p_1 \langle \mathbf{x}_i \mathbf{x}_j \rangle + p_2 \right) \ .
\end{aligned} \tag{3.16}$$

In this manuscript, we present results of SVM classification with the radial basis function (RBF) kernel. We do not use linear SVM as we perform linear classification with LDA. We chose the RBF kernel due to the fact that it is one of the mostly used nonlinear kernels. The idea behind RBF is that it measures similarity between $\mathbf{x}_i$ and $\mathbf{x}_j$: if the two points are far from each other the value of the kernel function will be close to 0 and almost one, otherwise. The RBF exposes local behavior, i.e., only a close training observation will influence a predicted label of a test observation. Due to the nonlinear nature of the produced class boundary, the RBF kernel can classify training data almost perfectly, even in presence of the outliers. Because of that, the box constraint C has to be chosen with care in order to avoid overfitting. The radial basis kernel have another tuning parameter, namely denominator in Eq. (3.16). These two parameters, C and σ should be carefully selected using a cross-validation technique which is discussed in the next section.

3.3 Cross-Validation

When building a classification model, it is important to ensure not only high classification accuracy but also good predictive power. Usually, classification model has a range of parameters, which have to be correctly estimated, the process of finding the optimal parameters is called model selection. In order to select the model with the highest predictive power, one can perform training and testing of the model with different sets of parameters, and then select the one which gives the best accuracy. However, this method is prone to overfitting, since the parameters are tuned for the specific dataset. To overcome this complication, one can use a *validation set* in addition to *training* and *testing sets*. In this case, parameters of the model are chosen such as to give the highest accuracy on the validation set, and the final classification accuracy is reported on the test set. This allows to estimate how well a particular model performs on unseen data. However, often the number of available observations is limited and, if the validation set is small, it cannot fully reflect on the predictive performance of the model and it also depends on how the data was partitioned into these three subsets. In this case, it may be more appropriate to use *K-fold cross-validation* (CV), illustrated in Fig. 3.4.

In this approach, the available data is randomly partitioned into K non-overlapping groups, known as *folds*, then one fold is set aside for validation and the classifier is trained on the remaining data, this procedure is repeated for all choices of the validation sets. The final classification error is obtained by averaging the errors from all the runs. Normally, the folds are of equal size (as much as possible) and a reasonable choice for K is 5 or 10 folds, however in the case of very scarce data one might use $K = N$, this is known as *Leave-one-out cross-validation* (LOOCV).

Cross-validation allows us to use $(K-1)/K$ amount of data for training of the model and validate it on the whole dataset. It also ensures that the model does not depend on the particular data partitioning, thus increasing generalization power of the model.

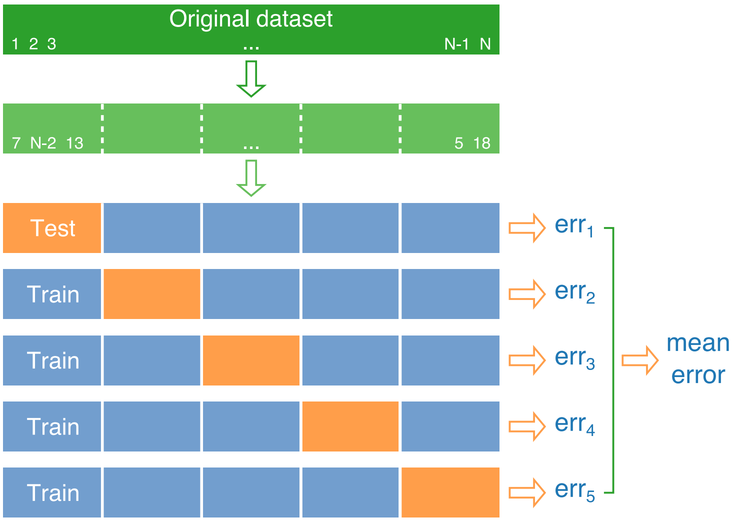


Fig. 3.4: A schematic illustration of K-fold cross-validation for $K = 5$. Original dataset (shown in dark green) is randomly partitioned into K disjoint sets (shown in light green). Then $K - 1$ parts are used for training a model (shown in blue) and remaining part is used for evaluation (shown in orange). This process is repeated K times for all possible choices of the test set, producing test errors. The final performance is reported by averaging the errors from each iteration.

Assume that a measure of the test error is given by misclassification rate:

$$\text{err} = \frac{1}{n}\sum_{i=1}^{n} I(\hat{y}_i \neq y_i) \quad , \tag{3.17}$$

where I is the indicator function and n is the number of observations in a test dataset (i.e., in a fold). Then we can write the CV error as follows:

$$\text{err}_{\text{CV}} = \frac{1}{K}\sum_{k=1}^{K} \text{err}_k \tag{3.18}$$

In this manuscript, we use 5-fold cross-validation. The choice of the CV technique (5-fold over LOOCV) is defined by the fact that the number of training runs is proportional to the number of folds. Therefore, this particular choice offers reasonable combination of computational efficiency and good performance estimate. In LDA there are no parameters to estimate, that is why we use CV to report final mean accuracy and its standard deviation. However, SVM with an RBF kernel requires tunning of two parameters: the box constraint C and the width of the kernel controlled through the σ parameter. They are estimated using pattern search optimization [70] over the values of both parameters ranging from 0.1 to 100 with CV error as a cost function. After the best parameters were found we report mean error computed over the same CV partition.

3.4 Feature Ranking

After performing classification of the available data using LDA and SVM, we would like to gain an insight into how particular features (predictors) contribute to the results of classification and if some features are more important than the others. For that we rank features based on how good they allow to discriminate two classes. Feature ranking is a tool used in feature space dimensionality reduction, therefore, before we proceed with definition of ranking technique used in this work, let us consider general notion of dimensionality reduction. In majority of situations, one wants to diminish number of available features for two main reasons: (i) it lowers a risk of overfitting; (ii) it speeds up computations.

Dimensionality reduction methods can be divided into two broad groups: *feature extraction* and *feature selection*. The former group involves a transformation of original feature space into a new, possibly smaller, space with increased predictive power. This group includes such techniques as Principal Component Analysis (PCA) [38], Independent Component Analysis (ICA) [57], LDA [52] and others. However, this approach produces completely new set of features, which makes their interpretation more difficult.

Feature selection techniques, as the name suggests, select a subset of the most informative and relevant features. Ideally, one wants to test all possible combinations of features and select the one which yields the highest value of an objective function (which can be, for instance, classification accuracy). However, exhaustive search can quickly become implausible as the number of features grows. If an original feature set consist of p features and one wants to identify optimal subset consisting of p' features, then the objective function should be evaluated $\binom{p}{p'}$ times. If the size of new subset is unknown, then the number of evaluations reaches 2^p. For example, if one wants to select 3 features, ensuring the lowest classification error, from original 10 features, one will have to perform training and validation of the model 120 times.

To overcome computational limitations, suboptimal search strategies are usually used instead of the exhaustive search. Feature selection algorithms can be classified in three main groups: *wrappers*, *embedded*, and *filters* [61, 126]. *Embedded* methods are usually specific to a classifier and a feature selection process is carried out during the training phase, for example, decision trees [121], LASSO technique (Least Absolute Shrinkage and Selection Operator) [137]. Next group of methods, *wrappers*, selects a feature subset based on the increase in the predictive power and are tailored to each specific classifier [86]. These methods add or remove features to a subset of features based on the classifier's accuracy, usually, involving cross-validation in the evaluation phase. Finally, *filter* methods, unlike wrappers and embedded methods, perform feature selection without optimizing performance of a particular classifier [99], normally it relies on intrinsic properties of the data, such as correlation between features and class labels, information gain or test statistics. Each of the discussed methods have their own advantages and downsides [60, 61], however, we focus our attention on filter methods as they are independent of the classifier and they provide us with the possibility of feature ranking.

One of the most widely used filter methods is mutual information between features and class labels [9, 115, 138, 142], which measures the information gain from each feature, however, it requires estimation of joint probabilities between features and class labels. Some of the solutions use histograms [17, 26] or Parzen window [35, 90], nonetheless, accuracy of the estimate remains limited. For that reason we utilize test statistics. As we are specifically interested in non-parametric tests, we choose Wilcoxon rank sum test as a feature ranking technique [61, 64].

Wilcoxon test is a nonparametric test, where the null hypothesis is that two distributions are the same, while no assumption about the distributions is made, and the alternative hypothesis is that there is a location shift between the two distributions. It is similar to t-test when the normality assumption cannot be ensured. The test procedure for two independent samples A and B with sample sizes N_A and N_B, respectively, is the following. First, combine two samples into one set of the size $N = N_A + N_B$; then sort the elements in ascending order and assign a rank r_i to each element, starting from 1, in case of ties, these values receive an average of their ranks.

W is a sum of ranks of the A-values, i.e.

$$W = \sum_{i=1}^{N_A} r_i \ . \tag{3.19}$$

The null hypothesis is rejected at the α level of significance if $W \geqslant w_{\alpha/2}$, where $w_{\alpha/2}$ is the critical value of the test corresponding to α significance level [69].

For large samples ($N \geqslant 10$) the distribution of W converges to a normal distribution and can be approximated using z-score:

$$z = \frac{W - \mathbb{E}_0[W]}{\sigma_0[W]} \ , \tag{3.20}$$

where $\mathbb{E}_0[W]$ and $\sigma_0[W]$ are the expected value and the standard deviation under the null hypothesis and are given by:

$$\mathbb{E}_0[W] = \frac{N_A(N_A + N_B + 1)}{2} \tag{3.21}$$

$$\sigma_0[W] = \sqrt{\frac{N_A N_B (N_A + N_B + 1)}{12}} \ . \tag{3.22}$$

Now, the null hypothesis should be rejected if z-score is larger than the critical value, i.e., $|z| \geqslant z_{\alpha/2}$.

In this work, we compute approximation of Wilcoxon statistics for each feature, where two samples are values of a candidate feature belonging to each class. We, then, order the features according to their z-scores, assuming that the larger the score is, the more discriminative power is carried by the feature, ranking is normalized over all features.

3.5 Results

In this section we present results of the classification. In the present work, we perform only binary classification. We separate experimental conditions, for example pre-/post-incision classification for patients receiving desflurane from pre-/post-incision in propofol.

We use two classifiers: LDA and SVM with nonlinear kernel (more specifically, with the RBF kernel). The tune-in parameters of the SVM classifier were estimated using CV technique and the final accuracy is reported using 5-fold CV, for more details see Section 3.3.

We also perform the classification using three feature sets: the features obtained with the power spectral analysis (see Section 2.2), the features obtained with the symbolic recurrence structure analysis (see Section 2.3), and, finally, the full feature set combining the aforementioned subsets. This allows us to see how different features affect the accuracy. As the PSA studies linear dynamics in the system and the SRSA, nonlinear, we can make conclusions on the system's behavior.

3.5.1 Power Spectral Analysis Features

We start by classifying the features obtained with the power spectral analysis. Those features include: total power of the signal (`totPow`), power in δ- and α-bands (`deltaPow` and `alphaPow`), and parameters of the model of EEG background activity (`2D` and `gamma2`). For more details see Section 2.2.

Table 3.1 shows results of classification using LDA and SVM (with the RBF kernel) classifiers. We use classification error (misclassification rate) as a measure of a classifier's performance.

The performance was evaluated using 5-fold cross-validation and results are reported as mean classification error over all folds and standard deviation in percentage.

Table 3.1: Mean classification error and standard deviation using the features obtained with the power spectral analysis. First column reports the performance of the LDA classifier and the second column shows the performance of the SVM classifier with the RBF kernel.

	Classification error	
Classification scenario	LDA	SVM
desflurane/propofol in pre-incision	31 ± 5%	24 ± 5%
desflurane/propofol in post-incision	28 ± 7%	25 ± 2%
pre-/post-incision in desflurane	52 ± 4%	45 ± 5%
pre-/post-incision in propofol	53 ± 3%	46 ± 11%

We can see from the table that the classification carried out with nonlinear SVM outperforms the classification with LDA. In case of pre-/post-incision classification error barely exceeds 50%, which is a random chance. This clearly shows that the class distributions have nonlinear boundaries and requires more complex decision. For the desflurane/propofol classification with SVM we achieve an average accuracy of 75%, which is a good indicator taking into account that the classification used only the PSA features, which describe linear dynamics in the signals.

3.5.2 Symbolic Recurrence Structure Analysis Features

In this subsection we present results of the classification with the features obtained from the symbolic recurrence analysis. These features include various complexity measures: cardinality of the symbolic alphabet (`alphabet`), number of distinct words (`words`), Lempel-Ziv complexity (`lz`), symbolic entropy (`entr`), and threshold distance in optimal recurrence (`eps`), see Section 2.5 for details. The features obtained with the SRSA depend on the time-frequency representation used, thus providing us with different feature sets for each TFR. Here, we focus on only two time-frequency methods: the spectrogram and the scalogram (with equidistant frequencies). As it was shown in Section 2.6.2, these two methods outperforms other considered TFRs.

Table 3.2 shows the results of classification with the features obtained with the spectrogram, performed with LDA and SVM. As before, we use misclassification rate as a measure of classification performance, which was evaluated using 5-fold cross-validation, and results are reported as mean classification error over all folds and standard deviation in percentage.

The classification with only the PSA features obtained with the spectrogram decreases the error for pre-/post-incision classification, but it decreases the accuracy for desflurane/propofol. SVM still shows better performance, however, it is less noticeable in desflurane/propofol classification as it is in Table 3.2. Also we can observe an increase in error variation for the desflurane/propofol case.

Table 3.3 shows the results of the classification with the features obtained with the scalogram (equidistant frequencies). We can observe that the features obtained with the scalogram yield higher classification accuracy than the ones obtained with the spectrogram (cf. Table 3.2) in every classification scenario. We can see that the classification accuracy in case of pre-/post-incision is improved by approximately 10% compared to the PSA features (while standard deviation remains the same). However, this method is less efficient in case of desflurane/propofol signals, the accuracy is lower and standard deviation of the error is higher than in case of the PSA

Table 3.2: Mean classification error and standard deviation using the features obtained with the symbolic recurrence structure analysis. The features were extracted using spectrogram in the SRSA. First column reports the performance of the LDA classifier and the second column shows the performance of the SVM classifier with the RBF kernel.

	Classification error	
Classification scenario	LDA	SVM
desflurane/propofol in pre-incision	41 ± 8%	36 ± 5%
desflurane/propofol in post-incision	35 ± 8%	31 ± 9%
pre-/post-incision in desflurane	45 ± 9%	42 ± 5%
pre-/post-incision in propofol	53 ± 12%	41 ± 11%

features.

Table 3.3: Mean classification error and standard deviation using the features obtained with the symbolic recurrence structure analysis. The features were extracted using scalogram with equidistant frequencies in the SRSA. First column reports the performance of the LDA classifier and the second column shows the performance of the SVM classifier with the RBF kernel.

	Classification error	
Classification scenario	LDA	SVM
desflurane/propofol in pre-incision	46 ± 8%	27 ± 4%
desflurane/propofol in post-incision	37 ± 8%	30 ± 12%
pre-/post-incision in desflurane	55 ± 7%	34 ± 4%
pre-/post-incision in propofol	48 ± 7%	38 ± 10%

3.5.3 Full Feature Set

In this subsection we present the results of classification with features obtained from the power spectral and symbolic recurrence analyses (full feature space). Similarly to the previous subsection, results are separated based on the TFR which was used to obtain the features. Table 3.4 shows the results of the classification performed with LDA and SVM and the features obtained with the spectrogram and Table 3.5 shows the results for the features obtained with the scalogram.

We can see that using all the features (and the spectrogram as reconstruction step in the SRSA) improves the classification of pre-/post-incision signals. The desflurane/propofol classification accuracy is also improved compared to the classification with only the SRSA features, however, it is still not as good as the classification with only the PSA features.

The classification carried out with SVM and the full feature set (using the scalogram in the feature generation stage) demonstrates the best results among all the classification procedures presented. But we can see that while the mean error is decreasing, its standard deviation is higher than standard deviation of the classification error with only the PSA features. Which might indicate lower generalization power and tendency to overfitting.

We see that combining the two feature subsets increases classification accuracy, however larger number of features increase the variance of the results, making the classifiers prone to overfitting.

Table 3.4: Mean classification error and standard deviation using full feature set. The features were extracted using spectrogram in the SRSA. First column reports the performance of the LDA classifier and the second column shows the performance of the SVM classifier with the RBF kernel.

	Classification error	
Classification scenario	LDA	SVM
desflurane/propofol in pre-incision	31 ± 5%	27 ± 10%
desflurane/propofol in post-incision	28 ± 7%	22 ± 7%
pre-/post-incision in desflurane	47 ± 8%	40 ± 9%
pre-/post-incision in propofol	47 ± 7%	36 ± 11%

Table 3.5: Mean classification error and standard deviation using full feature set. The features were extracted using scalogram with equidistant frequencies in the SRSA. First column reports the performance of the LDA classifier and the second column shows the performance of the SVM classifier with the RBF kernel.

	Classification error	
Classification scenario	LDA	SVM
desflurane/propofol in pre-incision	25 ± 8%	22 ± 10%
desflurane/propofol in post-incision	32 ± 7%	23 ± 8%
pre-/post-incision in desflurane	55 ± 9%	39 ± 7%
pre-/post-incision in propofol	52 ± 11%	38 ± 9%

Keeping that in mind, our next step is to identify features with the most discriminative power and to perform classification only with this features. Our previous results demonstrate clear superiority of the SVM classifier with the RBF kernel over the LDA classifiers, thus, further classification is going to be presented with SVM.

3.5.4 Feature Ranking

Finally, we present results of feature ranking described in Section 3.4. Two tables below show the results of feature ranking for two full feature sets obtained with the spectrogram (Table 3.6) and the scalogram with equidistant frequencies (Table 3.7). Each table contains four columns, one for each classification case.

Table 3.6: Feature ranks of all features obtained with spectrogram. The top features are the ones carrying the most discriminative power according to Wilcoxon rank sum test.

Rank	desflurane/propofol in pre-incision	desflurane/propofol in post-incision	pre-/post-incision in desflurane	pre-/post-incision in propofol
1	`2D`	`2D`	`words`	`alphaPow`
2	`entr`	`alphaPow`	`lz`	`gamma2`
3	`alphaPow`	`entr`	`eps`	`deltaPow`
4	`deltaPow`	`totPow`	`alphaPow`	`words`
5	`totPow`	`eps`	`alphabet`	`alphabet`
6	`lz`	`words`	`totPow`	`lz`
7	`gamma2`	`gamma2`	`2D`	`entr`
8	`eps`	`alphabet`	`gamma2`	`totPow`
9	`alphabet`	`deltaPow`	`deltaPow`	`eps`
10	`words`	`lz`	`entr`	`2D`

In case of the features obtained with the spectrogram, we see consistency in desflurane/propofol groups: the first three features include the same features, namely, power in α-band, the parameter of EEG background activity model, indirectly related to the height of the δ-peak, and symbolic entropy. We can see that these features result from a combination of two feature construction methods, which means that it is necessary to use both approaches in the analysis of these experimental conditions.

We do not observe the same consistency in case of pre-/post-incision classification. We see that for incision EEG recorded from patients receiving desflurane, three most discriminative features are generated with symbolic recurrence structure analysis. However, for signals recorded from patients who received propofol, the three first features are obtained with the power spectral analysis.

In case of the features obtained with the scalogram (with equidistant frequencies), the most discriminative feature in the desflurane/propofol classification is the same as in Table 3.6: `2D` which is proportional to the power in δ-band. The next two important features are complexity measures. We observe similar pattern as in the previous table, namely, a combination of the SRSA and PSA features plays important role in the classification of desflurane/propofol signals. We do not observe the same consistency in the pre-/post-incision classification, which might might be related to the low classification accuracy.

Table 3.7: Feature ranks of all features obtained with scalogram with equidistant frequencies. The top features are the ones carrying the most discriminative power according to Wilcoxon rank sum test.

Rank	desflurane/propofol in pre-incision	desflurane/propofol in post-incision	pre-/post-incision in desflurane	pre-/post-incision in propofol
1	`2D`	`2D`	`eps`	`alphaPow`
2	`lz`	`entr`	`alphaPow`	`deltaPow`
3	`alphabet`	`lz`	`alphabet`	`gamma2`
4	`words`	`words`	`deltaPow`	`alphabet`
5	`entr`	`alphabet`	`words`	`eps`
6	`totPow`	`totPow`	`entr`	`lz`
7	`gamma2`	`alphaPow`	`gamma2`	`words`
8	`alphaPow`	`deltaPow`	`lz`	`totPow`
9	`eps`	`eps`	`totPow`	`2D`
10	`deltaPow`	`gamma2`	`2D`	`entr`

In conclusion, we present classification results obtained with only three first features from Tables 3.6 and 3.7. We do so in order to see if it is possible to achieve the same (or higher) level of accuracy with reduced feature space as with complete set. This time, however, we use only the SVM classifier, as it clearly outperforms LDA in the current work.

Table 3.8: Mean classification error obtained with SVM classifier and with the three most discriminative features. First column shows results for features obtained with spectrogram and the second column, with scalogram (equidistant frequency).

	Classification error	
Classification scenario	Spectrogram	Scalogram (linear frequency)
desflurane/propofol in pre-incision	25 ± 4%	29 ± 8%
desflurane/propofol in post-incision	20 ± 4%	24 ± 14%
pre-/post-incision in desflurane	41 ± 11%	40 ± 11%
pre-/post-incision in propofol	42 ± 12%	42 ± 12%

Classifying desflurane from propofol with only three features in case of the spectrogram improves the accuracy and variation of the results, outperforming previous classification models and classification carried out with only three features obtained with the scalogram. However, the error in pre-/post incision classification increases with the reduced number of features, as well as standard deviation of the error.

3.6 Discussion

We demonstrate that the SVM classifier with the RBF kernel outperforms LDA in each classification case, which indicates that the distribution of the classes is complex with clearly nonlinear boundaries between them.

Classifying the data with different feature sets allows us to identify certain patterns. We can conclude that only the power spectral features are sufficient for classifying desflurane and propofol. That indicates that different anesthetic agents produce different oscillatory rhythms in the brain, which is in accordance with the previous studies (cf. [45, 95]). In fact, the same level of accuracy can be achieved performing classification with the power spectral and recurrence-based features. Nonetheless, the larger feature space causes a higher level of variance, which might suggest redundancy in the features. We overcome this problem by achieving the highest accuracy and lowest standard deviation of the error by using only three features, which were obtained during the feature ranking procedure. We observe that the three best features in the desflurane/propofol classification consist of the power spectral and symbolic complexity features. The best results of the desflurane/propofol classification are obtained with only values of the power in α-band, the parameter of the model of EEG background activity (which is proportional to the height of the δ-peak), and the symbolic entropy (obtained with recurrence-based symbolic encoding using the spectrogram in reconstruction step). This indicates differences between two anesthetics not only in their power spectral densities (which characterize stationary processes), but also in the dynamics they produce. These findings are of interest as it might be an indication of different mechanisms of action between two drugs.

In case of the pre-/post-incision classification we do not observe as good results as in the case of desflurane/propofol. The best results, in terms of classification accuracy and variance, are achieved by using the SVM classifier with the full feature space, when the symbolic complexity measures were obtained with scalogram. One of the explanations for that might be the intrinsic properties of the wavelet transform to expose $1/f$ scaling of power spectra. This scaling, as it was discussed in the previous chapter (cf. Section 2.2), is found in many biological processes including human EEG. Also we observe variation of the feature ranking results as well as increased error rate in the classification performed only with the most significant features. It might indicate that the features proposed in this work are not able to properly capture the underlying dynamics in pre- and post-incisional signals.

There is a little number of studies performing quantitative analysis of EEG changes caused by intraoperative noxious stimuli. In one of such studies Hagihira [62] discusses EEG features, characteristic to incision under isoflurane, sevoflurane (both are volatile anesthetics) as well as propofol anesthesia. The author reports on changes in alpha and delta activity as a response to an incision. In the feature ranking process performed in our study, we report power in α- and δ-bands being within three most discriminative features, however, this is reported only for propofol anesthesia, while for desflurane the complexity measures are more important than spectral power values. The authors also indicate changes in BIS values, but this changes as well as changes in EEG power spectra are very patient-dependent. A similar idea was expressed in [113], where authors were analyzing noxious stimulation in sheep surgeries performed under isoflurane anesthesia. They also report on a high subject-based variability in the results. EEG related features that they proposed include a shift in EEG power from the lower to the higher frequency range or vise versa, this was also reported in [83, 84]. Aside from the EEG analysis discussed in the papers some authors propose to take into account other measures usually recorded during the surgery, for instance, hemodynamic responses [84, 113, 124, 149] or BIS values [62, 133]. However, most of the authors agree in their studies on high intrapatient variability in their response to incision. Thus, the most reasonable way to improve the analysis is to consider EEG signals recorded throughout the whole duration of the surgery. This would allow to treat incision as a stimulus in event-related potential (ERP) analysis, in which SRSA has previously shown good results [21, 22, 140]. Another improvement can be done by considering other biological variables in conjunction with the EEG.

General Conclusions

General anesthesia is an indispensable tool in a modern clinical practice, well adjusted and controlled anesthesia can guarantee a smooth surgery, reduced recovery time and lower risk of post-operative pain and post-traumatic stress disorder. The main components of anesthesia include amnesia, immobility, hypnosis and analgesia. Insights into the mechanisms underlying each of these components might ensure a good quality of surgical operations. However, understanding how anesthesia works is still an open question in the research community. Our work contributes to the field by studying analgesia effect of anesthesia.

This study was conducted on the EEG data acquired from patients undergoing surgery under general anesthesia. The objective of this thesis was to perform classification of the signals recorded during the pre-incisional and post-incisional phases of the surgery. During the surgery a group of people received desflurane as an anesthetic agent and another group received propofol. These are two different anesthetic drugs, with different properties and mechanisms. We have also performed classification of EEG signals based on the type of the administered agent.

Contributions

Our work can be divided into two major parts. In the first part we have extracted EEG signal components, i.e. features, which characterize the process under study and yield the best separability between the classes. In the second part we have performed classification of the data with the obtained features. In this work we aimed at distinguishing desflurane from propofol and pre-incisional EEG signals from post-incisional ones. Each of the two stages of our work can be subdivided into smaller steps. Feature construction has been carried out with power spectral analysis, allowing to describe stationary properties of the signals, and with recurrence-based symbolic analysis, which originates from dynamical system analysis. The classification has been performed with two classifiers, resting upon different assumptions and producing linear and nonlinear boundaries between classes. In the final step, we have performed feature ranking in order to establish features playing the most significant part in the classification process.

Power Spectral Analysis

The first stage of the feature construction has been carried out by the power spectral analysis. We have measured total power of the signal and power in α- and δ-bands, relatively to the power contained in the EEG background activity. Normally, power spectrum of such activity follows $1/f^{\alpha}$ frequency scaling. The proposed model of this behavior is the power spectrum of Ornstein-Uhlenbeck process, which makes it plausible to analyze power in low frequency components. The analyzed model has two parameters which, together with band power, have been used to characterize different types of signals. The results reveal that the largest differences between desflurane and propofol are found in total power, α-band power and a parameter of Ornstein-Uhlenbeck process, indirectly related to the power in δ-band.

Symbolic Recurrence Structure Analysis

We argued that the recurrence-based symbolic dynamics might be a good representation of a signal under study. The main idea behind the method is to identify quasi-stable states in a system's dynamics, based on its recurrence matrix. If analyzed data is presented with a single time-series, the preprocessing step is required, i.e., reconstruction of multi-dimensional phase space of the system from a single observable. The most common technique for the reconstruction is delayed embeddings. Here, we have proposed a new reconstruction technique, based on signal's time-frequency representation. The results have shown that the use of time-frequency embedding might be more advantageous than the method of delays in certain applications, such as, for instance, analysis of oscillatory signals. The analysis have demonstrated efficiency of the proposed reconstruction technique on the synthetic data, which included the transient oscillations and the Lorenz system.

The symbolic recurrence structure analysis has been successfully applied to numerous problems, including neural data processing, however, to our knowledge, it has not been applied to incisional data under general anesthesia, which is analyzed in this thesis. The results of the analysis of the synthetic data, whose dynamics is well-known, have demonstrated that domains of interest can be easily detected with a combination of recurrence-based symbolic encoding and time-frequency embedding. However, in case of the EEG analysis we don't have *a priori* knowledge of the system's dynamics and are not able to estimate correctness of the resulting symbolic representation with the naked eye. For that reason, we have proposed to use various complexity measures to quantitatively compare obtained symbolic sequences.

In addition, we have studied influence of different time-frequency representations on the resulting symbolic sequence. The TFRs studied in this manuscript include reassigned spectrogram, scalograms with equally spaced frequency vector and nonlinear frequencies. Despite a better detection of recurrent domains in the synthetic data, we observed that the best efficiency in case of the EEG data analysis was achieved with the scalogram with linearly sampled frequencies. One of the reasons for this is intrinsic property of the wavelet-based transforms to scale power with frequency, i.e., the previously discussed $1/f^{\alpha}$ behavior.

Classification

We have performed binary classification for desflurane/propofol and pre-/post-incision cases. On average, the percentage of correctly classified signals is 60% in case of pre-/post-incision and around 75% for desflurane/propofol case. We have performed classification using linear discriminant analysis and support-vector machines with Gaussian kernel. It turns out that SVM with the nonlinear Gaussian kernel clearly outperforms LDA. The classification was done using different sets of features: features extracted with the power spectral analysis, with the symbolic recurrence structure analysis, and by combining the two feature subsets. This analysis reveals that classification with the power spectral analysis features slightly outperforms classification with only recurrence symbolic features. Next, the classification results demonstrate that the performance with the full feature set is higher than for each feature subset separately. However, increasing number of features when using nonlinear classifier might cause overfitting, to overcome this risk we have conducted feature ranking in order to identify the most discriminative features. The results reveal that in the majority of the cases the three most discriminative features include features from both subsets, and that using only these three features might actually decrease the mean classification error and its variation.

Future Work

The current work has demonstrated effectiveness of the proposed combination of the SRSA with a time-frequency embedding on the synthetic data. Also the use of this analysis in the desflurane/propofol classification has shown that it is capable of capturing differences in the underlying dynamics of the two types of signals, thus indicating that the method can be applicable to EEG data analysis. However, we have not been able to achieve high classification accuracy in pre-/post-incisional signals classification, which might be an sign of no obvious response to incision in EEG signals.

One of the possible improvements to this problem is to change the data acquisition process. We considered signal's segments approximately 4 minutes apart, and post-incisional EEG activity was recorded 2 minutes after the start of incision. The analysis might be improved if the data is recorded continuously, i.e., capturing the time of incision. With such approach the incision can be treated as a stimulus in the event-related potential analysis. The recurrence-based symbolic encoding has been successfully applied to the analysis of such experimental data. Also, increased sampling rate might be a possible improvement, as events in the brain can be of a very brief duration (an order of milliseconds), and with undersampled data information of such events can be lost. Another improvement, inspired by previous studies on the same topic, can be made by analyzing other biological signals together with EEG.

Bibliography

[1] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, « The analysis of observed chaotic data in physical systems », *Reviews of Modern Physics,* vol. 65, no. 4, pp. 1331–1392, 1993.

[2] M. Aboy, R. Hornero, D. Abasolo, and D. Alvarez, « Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis », *IEEE Transactions on Biomedical Engineering,* vol. 53, no. 11, pp. 2282–2288, 2006.

[3] R. Acharya U., O. Faust, N. Kannathal, T. Chua, and S. Laxminarayan, « Non-linear analysis of EEG signals at various sleep stages », *Computer Methods and Programs in Biomedicine,* vol. 80, no. 1, pp. 37–45, 2005.

[4] P. S. Addison, *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance.* Bristol: Inst. of Physics Publ, 2002, 353 pp.

[5] P. M. Addo, M. Billio, and D. Guégan, « Nonlinear dynamics and recurrence plots for detecting financial crisis », *The North American Journal of Economics and Finance,* vol. 26, pp. 416–435, 2013.

[6] O. Akeju, K. J. Pavone, M. B. Westover, R. Vazquez, M. J. Prerau, P. G. Harrell, K. E. Hartnack, J. Rhee, A. L. Sampson, K. Habeeb, *et al.*, « A comparison of propofol-and dexmedetomidine-induced electroencephalogram dynamics using spectral and coherence analysis », *The Journal of the American Society of Anesthesiologists,* vol. 121, no. 5, pp. 978–989, 2014.

[7] W. P. Akrawi, J. C. Drummond, C. J. Kalkman, and P. M. Patel, « A comparison of the electrophysiologic characteristics of EEG burst-suppression as produced by isoflurane, thiopental, etomidate, and propofol. », *Journal of neurosurgical anesthesiology,* vol. 8, no. 1, pp. 40–46, 1996.

[8] M. T. Alkire, A. G. Hudetz, and G. Tononi, « Consciousness and anesthesia », *Science,* vol. 322, no. 5903, pp. 876–880, 2008.

[9] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, « Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b », *Frontiers in Neuroscience,* vol. 6, p. 39, 2012.

[10] A. Arcentales, B. F. Giraldo, P. Caminal, S. Benito, and A. Voss, « Recurrence quantification analysis of heart rate variability and respiratory flow series in patients on weaning trials », in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* IEEE, 2011, pp. 2724–2727.

[11] P. Arhem, G. Klement, and J. Nilsson, « Mechanisms of anesthesia: Towards integrating network, cellular, and molecular level modeling. », *Neuropsychopharmacology,* vol. 28, S40–S47, 2003.

[12] F. Auger and P. Flandrin, « Improving the readability of time-frequency and time-scale representations by the reassignment method », *Signal Processing, IEEE Transactions on,* vol. 43, no. 5, pp. 1068–1089, 1995.

[13] F. Auger, P. Flandrin, Y.-T. Lin, S. McLaughlin, S. Meignen, T. Oberlin, and H.-T. Wu, « Time-frequency reassignment and synchrosqueezing: An overview », *IEEE Signal Processing Magazine,* vol. 30, no. 6, pp. 32–41, 2013.

[14] P. Bak, C. Tang, and K. Wiesenfeld, « Self-organized criticality: An explanation of the $1/f$ noise », *Physical Review Letters,* vol. 59, no. 4, pp. 381–384, 1987.

[15] R. Baker, T. C. Gent, Q. Yang, S. Parker, A. L. Vyssotski, W. Wisden, S. G. Brickley, and N. P. Franks, « Altered activity in the central medial thalamus precedes changes in the neocortex during transitions into both sleep and propofol anesthesia », *Journal of Neuroscience,* vol. 34, no. 40, pp. 13 326–13 335, 2014.

[16] E. Başar, *Brain Function and Oscillations. Volume I: Brain Oscillations. Principles and Approaches,* red. by H. Haken, ser. Springer Series in Synergetics. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, 363 pp.

[17] R. Battiti, « Using mutual information for selecting features in supervised neural net learning », *IEEE Transactions on neural networks,* vol. 5, no. 4, pp. 537–550, 1994.

[18] C. Bédard and A. Destexhe, « Macroscopic models of local field potentials and the apparent $1/f$ noise in brain activity », *Biophysical Journal,* vol. 96, no. 7, pp. 2589–2603, 2009.

[19] C. Bedard, H. Kroeger, and A. Destexhe, « Does the $1/f$ frequency scaling of brain signals reflect self-organized critical states? », *Physical review letters,* vol. 97, no. 11, p. 118 102, 2006.

[20] P. beim Graben, « Recurrence complexity analyses of anesthetic EEG data », Talk presented at Neurosys Team, INRIA/LORIA CR Nancy, France, 2015.

[21] P. beim Graben and A. Hutt, « Detecting recurrence domains of dynamical systems by symbolic dynamics », *Physical Review Letters,* vol. 110, no. 15, p. 154 101, 2013.

[22] P. beim Graben and A. Hutt, « Detecting event-related recurrences by symbolic analysis: applications to human language processing », *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences,* vol. 373, no. 2034, p. 20 140 089, 2015.

[23] P. beim Graben, K. K. Sellers, F. Fröhlich, and A. Hutt, « Optimal estimation of recurrence structures from time series », *EPL (Europhysics Letters),* vol. 114, no. 3, p. 38 003, 2016.

[24] C. Bennett, L. J. Voss, J. P. M. Barnard, and J. W. Sleigh, « Practical use of the raw electroencephalogram waveform during general anesthesia: The art and science », *Anesthesia & Analgesia,* vol. 109, no. 2, pp. 539–550, 2009.

[25] C. M. Bishop, *Pattern Recognition and Machine Learning,* ser. Information science and statistics. New York: Springer, 2006, 738 pp.

[26] B. V. Bonnlander and A. S. Weigend, « Selecting input variables using mutual information and nonparametric density estimation », in *Proceedings of the 1994 International Symposium on Artificial Neural Networks,* 1994, pp. 42–50.

[27] B. E. Boser, I. M. Guyon, and V. N. Vapnik, « A training algorithm for optimal margin classifiers », in *Proceedings of the fifth annual workshop on Computational learning theory,* ACM, 1992, pp. 144–152.

[28] T. J. Brennan, P. K. Zahn, and E. M. Pogatzki-Zahn, « Mechanisms of incisional pain », *Anesthesiology clinics of North America,* vol. 23, no. 1, pp. 1–20, 2005.

[29] C. J. Burges, « A tutorial on support vector machines for pattern recognition », *Data mining and knowledge discovery,* vol. 2, no. 2, pp. 121–167, 1998.

[30] G. Buzsaki, « Neuronal oscillations in cortical networks », *Science,* vol. 304, no. 5679, pp. 1926–1929, 2004.

[31] A. M. Calvão and T. J. P. Penna, « The double pendulum: a numerical study », *European Journal of Physics,* vol. 36, no. 4, p. 045 018, 2015.

[32] J. Capon, « High-resolution frequency-wavenumber spectrum analysis », *Proceedings of the IEEE,* vol. 57, no. 8, pp. 1408–1418, 1969.

[33] M. J. Caterina, M. S. Gold, and R. A. Meyer, « Molecular biology of nociceptors », in *The Neurobiology of Pain: (Molecular and Cellular Neurobiology),* S. P. Hunt and M. Koltzenburg, Eds., ser. Molecular and cellular neurobiology series. Oxford: Oxford University Press, 2005.

[34] É. Chassande-Mottin, F. Auger, and P. Flandrin, « Reassignment », in *Time-Frequency Analysis,* F. Hlawatsch and F. Auger, Eds., London, UK: ISTE, 2008, pp. 249–277.

[35] T. W. S. Chow and D. Huang, « Estimating optimal feature subsets using efficient estimation of high-dimensional mutual information », *IEEE Transactions on Neural Networks,* vol. 16, no. 1, pp. 213–224, 2005.

[36] L. Cohen, « Time-frequency distributions-a review », *Proceedings of the IEEE,* vol. 77, no. 7, pp. 941–981, 1989.

[37] C. Cortes and V. Vapnik, « Support-vector networks », *Machine learning,* vol. 20, no. 3, pp. 273–297, 1995.

[38] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach.* London: Prentice Hall, 1982, 472 pp.

[39] R. Donner, U. Hinrichs, and B. Scholz-Reiter, « Symbolic recurrence plots: A new quantitative framework for performance analysis of manufacturing networks », *The European Physical Journal Special Topics,* vol. 164, no. 1, pp. 85–104, 2008.

[40] G. Dumermuth and L. Molinari, « Spectral analysis of the EEG », *Neuropsychobiology,* vol. 17, no. 1-2, pp. 85–99, 1987.

[41] J.-P. Eckmann, S. Oliffson Kamphorst, and D. Ruelle, « Recurrence plots of dynamical systems », *Europhysics Letters (EPL),* vol. 4, no. 9, pp. 973–977, 1987.

[42] S. El Boustani and A. Destexhe, « Brain dynamics at multiple scales: can one reconcile the apparent low-dimensional chaos of macroscopic variables with the seemingly stochastic behavior of single neurons? », *International Journal of Bifurcation and Chaos,* vol. 20, no. 6, pp. 1687–1702, 2010.

[43] M. Farge, « Wavelet transforms and their applications to turbulence », *Annual Review of Fluid Mechanics,* vol. 24, no. 1, pp. 395–458, 1992.

[44] P. Faure and A. Lesne, « Recurrence plots for symbolic sequences », *International Journal of Bifurcation and Chaos,* vol. 20, no. 6, pp. 1731–1749, 2010.

[45] R. Ferenets, T. Lipping, P. Suominen, J. Turunen, P. Puumala, V. Jantti, S. Himanen, and A. Huotari, « Comparison of the properties of EEG spindles in sleep and propofol anesthesia », in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE,* IEEE, 2006, pp. 6356–6359.

[46] R. A. Fisher, « The use of multiple measurements in taxonomic problems », *Annals of eugenics,* vol. 7, no. 2, pp. 179–188, 1936.

[47] P. Flandrin, F. Auger, and É. Chassande-Mottin, « Time-frequency reassignment from principles to algorithms », in *Applications in Time-Frequency Signal Processing,* A. Papandreou-Suppappola, Ed., Boca Raton: CRC Press, 2002, pp. 179–203.

[48] R. Fletcher, *Practical Methods of Optimization,* 2nd ed., ser. A Wiley-Interscience publication. Chichester: Wiley, 2008, 436 pp.

[49] A. M. Fraser and H. L. Swinney, « Independent coordinates for strange attractors from mutual information », *Physical Review A,* vol. 33, no. 2, pp. 1134–1140, 1986.

[50] W. J. Freeman and T. Y. Cao, « Proposed Renormalization Group Analysis of Nonlinear Brain Dynamics at Criticality », in *Advances in Cognitive Neurodynamics ICCN 2007,* R. Wang, E. Shen, and F. Gu, Eds., Dordrecht: Springer Netherlands, 2008, pp. 145–156.

[51] R. Friedrich and C. Uhl, « Spatio-temporal analysis of human electroencephalograms: Petit-mal epilepsy », *Physica D: Nonlinear Phenomena,* vol. 98, no. 1, pp. 171–182, 1996.

[52] K. Fukunaga, *Introduction to Statistical Pattern Recognition,* 2nd ed. Boston: Academic Press, 1990, 592 pp.

[53] S. A. Fulop and K. Fitz, « Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications », *The Journal of the Acoustical Society of America,* vol. 119, no. 1, p. 360, 2006.

[54] F. Gabbiani and S. Cox, *Mathematics for Neuroscientists,* 1st ed. Amsterdam: Elsevier, Acad. Press, 2010, 486 pp.

[55] R. Gallager, *Information Theory and Reliable Communication.* Vienna: Springer Vienna, 1972, 115 pp.

[56] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference,* 5th ed., ser. Statistics, textbooks & monographs. Boca Raton: Taylor & Francis, 2011, 630 pp.

[57] M. Girolami, A. Cichocki, and S. I. Amari, « A common neural-network model for unsupervised exploratory data analysis and independent component analysis », *IEEE Transactions on Neural Networks,* vol. 9, no. 6, pp. 1495–1501, 1998.

[58] D. Griffiths and J. G. Jones, « Awareness and memory in anaesthetized patients », *British journal of anaesthesia,* vol. 65, no. 5, pp. 603–606, 1990.

[59] A. Groth, « Visualization of coupling in time series by order recurrence plots », *Physical Review E,* vol. 72, no. 4, p. 046 220, 2005.

[60] I. Guyon and A. Elisseeff, « An introduction to variable and feature selection », *The Journal of Machine Learning Research,* vol. 3, pp. 1157–1182, 2003.

[61] I. Guyon, M. Nikravesh, S. Gunn, and L. A. Zadeh, Eds., *Feature Extraction: Foundations and Applications,* red. by J. Kacprzyk, vol. 207, ser. Studies in Fuzziness and Soft Computing, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, 778 pp.

[62] S. Hagihira, « Changes in the electroencephalogram during anaesthesia and their physiological basis », *British Journal of Anaesthesia,* vol. 115, H. C. Hemmings, Ed., pp. i27–i31, suppl 1 2015.

[63] S. Hagihira, M. Takashina, T. Mori, H. Ueyama, and T. Mashimo, « Electroencephalographic bicoherence is sensitive to noxious stimuli during isoflurane or sevoflurane anesthesia », *The Journal of the American Society of Anesthesiologists,* vol. 100, no. 4, pp. 818–825, 2004.

[64] J. A. Hanley and B. J. McNeil, « The meaning and use of the area under a receiver operating characteristic (ROC) curve. », *Radiology,* vol. 143, no. 1, pp. 29–36, 1982.

[65] F. J. Harris, « On the use of windows for harmonic analysis with the discrete Fourier transform », *Proceedings of the IEEE,* vol. 66, no. 1, pp. 51–83, 1978.

[66] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction,* 2nd ed., ser. Springer Series in Statistics. New York: Springer, 2009, 745 pp.

[67] F. Hlawatsch and G. F. Boudreaux-Bartels, « Linear and quadratic time-frequency signal representations », *IEEE Signal Processing Magazine,* vol. 9, no. 2, pp. 21–67, 1992.

[68] W. E. Hoffman and G. Edelman, « Comparison of isoflurane and desflurane anesthetic depth using burst suppression of the electroencephalogram in neurosurgical patients », *Anesthesia & Analgesia,* vol. 81, no. 4, pp. 811–816, 1995.

[69] M. Hollander, D. A. Wolfe, and E. Chicken, « The Two-Sample Location Problem », in *Nonparametric Statistical Methods,* 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc., 2015, pp. 115–150.

[70] R. Hooke and T. A. Jeeves, « "Direct search" solution of numerical and statistical problems », *Journal of the ACM,* vol. 8, no. 2, pp. 212–229, 1961.

[71] J. Hu, J. Gao, and J. C. Principe, « Analysis of biomedical signals by the Lempel-Ziv complexity: The effect of finite data size », *IEEE Transactions on Biomedical Engineering,* vol. 53, no. 12, pp. 2606–2609, 2006.

[72] A. Hutt and T. D. Frank, « Critical fluctuations and $1/f^{\alpha}$-activity of neural fields involving transmission delays », *Acta Physica Polonica A,* vol. 108, no. 6, pp. 1021–1040, 2005.

[73] A. J. Izenman, *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning,* ser. Springer Texts in Statistics. New York: Springer, 2008, 733 pp.

[74] L. C. Jameson and T. B. Sloan, « Using EEG to monitor anesthesia drug effects during surgery », *Journal of Clinical Monitoring and Computing,* vol. 20, no. 6, pp. 445–472, 2006.

[75] A. J. E. M. Janssen, « Bilinear time-frequency distributions », in *Wavelets and Their Applications,* J. S. Byrnes, J. L. Byrnes, K. A. Hargreaves, and K. Berry, Eds., Dordrecht: Springer Netherlands, 1994, pp. 297–311.

[76] M. Jedynak, A. J. Pons, and J. Garcia-Ojalvo, « Cross-frequency transfer in a stochastically driven mesoscopic neuronal model », *Frontiers in Computational Neuroscience,* vol. 9, no. 14, 2015.

[77] J. Jeong and W. J. Williams, « Kernel design for reduced interference distributions », *IEEE Transactions on Signal Processing,* vol. 40, no. 2, pp. 402–412, 1992.

[78] D. Jordan, R. W. Miksad, and E. J. Powers, « Implementation of the continuous wavelet transform for digital time series analysis », *Review of Scientific Instruments,* vol. 68, no. 3, p. 1484, 1997.

[79] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis,* 2nd ed. Cambridge, UK: Cambridge University Press, 2004, 369 pp.

[80] J. Kennedy and R. Eberhart, « Particle swarm optimization », in *Proceedings of IEEE International Conference on Neural Networks,* vol. 4, IEEE, 1995, pp. 1942–1948.

[81] M. B. Kennel and H. D. I. Abarbanel, « False neighbors and false strands: A reliable minimum embedding dimension algorithm », *Physical Review E,* vol. 66, no. 2, p. 026 209, 2002.

[82] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, « Determining embedding dimension for phase-space reconstruction using a geometrical construction », *Physical Review A,* vol. 45, no. 6, pp. 3403–3411, 1992.

[83] S. Kiyama and J. Takeda, « Effect of extradural analgesia on the paradoxical arousal response of the electroencephalogram », *British Journal of Anaesthesia,* vol. 79, no. 6, pp. 750–753, 1997.

[84] E. Kochs, P. Bischoff, U. Pichlmeier, and J. Schulte am Esch, « Surgical stimulation induces changes in brain electrical activity during isoflurane/nitrous oxide anesthesia. A topographic electroencephalographic analysis », *Anesthesiology,* vol. 80, no. 5, pp. 1026–1034, 1994.

[85] K. Kodera, R. Gendrin, and C. de Villedary, « Analysis of time-varying signals with small BT values », *Acoustics, Speech and Signal Processing, IEEE Transactions on,* vol. 26, no. 1, pp. 64–76, 1978.

[86] R. Kohavi and G. H. John, « Wrappers for feature subset selection », *Artificial Intelligence,* vol. 97, no. 1-2, pp. 273–324, 1997.

[87] B. Krese and E. Govekar, « Recurrence quantification analysis of intermittent spontaneous to forced dripping transition in laser droplet generation », *Chaos, Solitons & Fractals,* vol. 44, no. 4–5, pp. 298–305, 2011.

[88] D. Kugiumtzis and N. D. Christophersen, « State space reconstruction: method of delays vs singular spectrum approach », University of Oslo, Department of informatics, Research Report No 236, 1997.

[89] A. Kumar, A. Bhattacharya, and N. Makhija, « Evoked potential monitoring in anaesthesia and analgesia », *Anaesthesia,* vol. 55, no. 3, pp. 225–241, 2000.

[90] N. Kwak and C.-H. Choi, « Input feature selection by mutual information based on Parzen window », *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 24, no. 12, pp. 1667–1671, 2002.

[91] C. Laing and G. J. Lord, Eds., *Stochastic Methods in Neuroscience,* New York: Oxford University Press, 2010, 370 pp.

[92] P. Lánský, L. Sacerdote, and F. Tomassetti, « On the comparison of Feller and Ornstein-Uhlenbeck models for neural activity », *Biological Cybernetics,* vol. 73, no. 5, pp. 457–465, 1995.

[93] J. Lefebvre, A. Hutt, J.-F. Knebel, K. Whittingstall, and M. M. Murray, « Stimulus statistics shape oscillations in nonlinear recurrent neural networks », *Journal of Neuroscience,* vol. 35, no. 7, pp. 2895–2903, 2015.

[94] A. Lempel and J. Ziv, « On the complexity of finite sequences », *IEEE Transactions on Information Theory,* vol. 22, no. 1, pp. 75–81, 1976.

[95] K. Leslie, J. Sleigh, M. J. Paech, L. Voss, C. W. Lim, and C. Sleigh, « Dreaming and electroencephalographic changes during anesthesia maintained with propofol or desflurane », *Anesthesiology,* vol. 111, no. 3, pp. 547–555, 2009.

[96] K. Levenberg, « A method for the solution of certain non-linear problems in least squares », *Quarterly of Applied Mathematics,* vol. 2, no. 2, pp. 164–168, 1944.

[97] X. Li, G. Ouyang, X. Yao, and X. Guan, « Dynamical characteristics of pre-epileptic seizures in rats with recurrence quantification analysis », *Physics Letters A,* vol. 333, no. 1–2, pp. 164–171, 2004.

[98] W. Liebert and H. G. Schuster, « Proper choice of the time delay for the analysis of chaotic time series », *Physics Letters A,* vol. 142, no. 2, pp. 107–111, 1989.

[99] H. Liu and H. Motoda, Eds., *Computational Methods of Feature Selection,* ser. Chapman & Hall/CRC data mining and knowledge discovery series, Boca Raton: Chapman & Hall/CRC, 2008, 440 pp.

[100] D. E. Longnecker, D. L. Brown, M. F. Newman, and W. M. Zapol, Eds., *Anesthesiology,* 1st ed., New York: McGraw-Hill, Medical Pub. Division, 2008, 2278 pp.

[101] E. N. Lorenz, « Deterministic nonperiodic flow », *Journal of the Atmospheric Sciences,* vol. 20, no. 2, pp. 130–141, 1963.

[102] S. G. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way,* 3rd ed. Amsterdam; Boston: Elsevier/Academic Press, 2009, 805 pp.

[103] D. W. Marquardt, « An algorithm for least-squares estimation of nonlinear parameters », *Journal of the Society for Industrial and Applied Mathematics,* vol. 11, no. 2, pp. 431–441, 1963.

[104] R. Martin, « Noise power spectral density estimation based on optimal smoothing and minimum statistics », *IEEE Transactions on Speech and Audio Processing,* vol. 9, no. 5, pp. 504–512, 2001.

[105] V. Martinez and D. Fletcher, « Prevention of opioid-induced hyperalgesia in surgical patients: Does it really matter? », *British journal of anaesthesia,* vol. 109, no. 3, pp. 302–304, 2012.

[106] N. Marwan, M. Carmenromano, M. Thiel, and J. Kurths, « Recurrence plots for the analysis of complex systems », *Physics Reports,* vol. 438, no. 5-6, pp. 237–329, 2007.

[107] N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths, « Recurrence-plot-based measures of complexity and their application to heart-rate-variability data », *Physical Review E,* vol. 66, no. 2, p. 026 702, 2002.

[108] M. D. McKenzie, « Chaotic behavior in national stock market indices », *Global Finance Journal,* vol. 12, no. 1, pp. 35–53, 2001.

[109] S. D. Meyers, B. G. Kelly, and J. J. O'Brien, « An introduction to wavelet analysis in oceanography and meteorology: With application to the dispersion of yanai waves », *Monthly Weather Review,* vol. 121, no. 10, pp. 2858–2866, 1993.

[110] E. Niedermeyer, « The normal EEG of the waking adult », *Electroencephalography: Basic principles, clinical applications, and related fields,* vol. 167, 2005.

[111] H. Nobach, C. Tropea, L. Cordier, J.-P. Bonnet, J. Delville, J. Lewalle, M. Farge, K. Schneider, and R. Adrian, « Review of some fundamentals of data processing », in *Springer Handbook of Experimental Fluid Mechanics,* Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 1337–1398.

[112] A. V. Oppenheim and R. W. Schafer, *Discrete-time signal processing,* ser. Prentice Hall signal processing series. Englewood Cliffs, NJ: Prentice-Hall, 1989, 879 pp.

[113] K. A. Otto and P. Mally, « Noxious stimulation during orthopaedic surgery results in EEG 'arousal' or 'paradoxical arousal' reaction in isoflurane-anaesthetised sheep », *Research in Veterinary Science,* vol. 75, no. 2, pp. 103–112, 2003.

[114] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, « Geometry from a time series », *Physical Review Letters,* vol. 45, no. 9, pp. 712–716, 1980.

[115] H. Peng, F. Long, and C. Ding, « Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy », *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, no. 8, pp. 1226–1238, 2005.

[116] D. M. Philbin, C. E. Rosow, R. C. Schneider, G. Koski, and M. N. D'Ambra, « Fentanyl and sufentanil anesthesia revisited: How much is enough », *Anesthesiology,* vol. 73, no. 1, pp. 5–11, 1990.

[117] M. Plancherel and M. Leffler, « Contribution à l'étude de la représentation d'une fonction arbitraire par des intégrales définies », *Rendiconti del Circolo Matematico di Palermo (1884-1940),* vol. 30, no. 1, pp. 289–335, 1910.

[118] H. Poincaré, « Sur le problème des trois corps et les équations de la dynamique », *Acta mathematica,* vol. 13, no. 1, pp. 3–270, 1890.

[119] R. Poli, J. Kennedy, and T. Blackwell, « Particle swarm optimization: An overview », *Swarm Intelligence,* vol. 1, no. 1, pp. 33–57, 2007.

[120] P. L. Purdon, E. T. Pierce, E. a Mukamel, M. J. Prerau, J. L. Walsh, K. F. K. Wong, A. F. Salazar-Gomez, P. G. Harrell, A. L. Sampson, A. Cimenser, S. Ching, N. J. Kopell, C. Tavares-Stoeckel, K. Habeeb, R. Merhar, and E. N. Brown, « Electroencephalogram signatures of loss and recovery of consciousness from propofol », *Proceedings of the National Academy of Sciences of the United States of America,* vol. 110, no. 12, E1142–E1151, 2013.

[121] J. R. Quinlan, « Induction of Decision Trees », *Machine Learning,* vol. 1, no. 1, pp. 81–106, 1986.

[122] M. Rabinovich, R. Huerta, and G. Laurent, « Transient dynamics for neural processing », *Science,* vol. 321, no. 5885, pp. 48–50, 2008.

[123] M. I. Rabinovich, R. Huerta, P. Varona, and V. S. Afraimovich, « Transient cognitive dynamics, metastability, and decision making », *PLoS Computational Biology,* vol. 4, no. 5, e1000072, 2008.

[124] I. J. Rampil and R. S. Matteo, « Changes in EEG spectral edge frequency correlate with the hemodynamic response to laryngoscopy and intubation », *Anesthesiology,* vol. 67, no. 1, pp. 139–142, 1987.

[125] M. Rodriguez-Fernandez, J. A. Egea, and J. R. Banga, « Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems », *BMC Bioinformatics,* vol. 7, p. 483, 2006.

[126] Y. Saeys, I. Inza, and P. Larranaga, « A review of feature selection techniques in bioinformatics », *Bioinformatics,* vol. 23, no. 19, pp. 2507–2517, 2007.

[127] M. Schartner, A. Seth, Q. Noirhomme, M. Boly, M.-A. Bruno, S. Laureys, and A. Barrett, « Complexity of multi-dimensional spontaneous EEG decreases during propofol induced general anaesthesia », *PLoS ONE,* vol. 10, no. 8, pp. 1–21, 2015.

[128] S. Schinkel, N. Marwan, and J. Kurths, « Order patterns recurrence plots in the analysis of ERP data », *Cognitive Neurodynamics,* vol. 1, no. 4, pp. 317–325, 2007.

[129] A. Schuster, « On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena », *Journal of Geophysical Research,* vol. 3, no. 1, p. 13, 1898.

[130] R. S. Schwartz, E. N. Brown, R. Lydic, and N. D. Schiff, « General anesthesia, sleep, and coma », *New England Journal of Medicine,* vol. 363, no. 27, pp. 2638–2650, 2010.

[131] P. K. Sinha, T. Koshy, *et al.*, « Monitoring devices for measuring the depth of anaesthesia-an overview », *Indian Journal of Anaesthesia,* vol. 51, no. 5, p. 365, 2007.

[132] C. A. Skarda and W. J. Freeman, « How brains make chaos in order to make sense of the world », *Behavioral and Brain Sciences,* vol. 10, no. 2, p. 161, 1987.

[133] J. W. Sleigh, K. Leslie, and L. Voss, « The effect of skin incision on the electroencephalogram during general anesthesia maintained with propofol or desflurane. », *Journal of clinical monitoring and computing,* vol. 24, no. 4, pp. 307–18, 2010.

[134] P. Stoica and R. L. Moses, *Spectral analysis of signals.* Upper Saddle River, NJ: Pearson/Prentice Hall, 2005, 452 pp.

[135] F. Takens, « Detecting strange attractors in turbulence », in *Dynamical Systems and Turbulence, Warwick 1980,* ser. Lecture Notes in Mathematics, D. Rand and L.-S. Young, Eds., vol. 898, Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, pp. 366–381.

[136] G. Thakur, E. Brevdo, N. S. Fučkar, and H.-T. Wu, « The synchrosqueezing algorithm for time-varying spectral analysis: Robustness properties and new paleoclimate applications », *Signal Processing,* vol. 93, no. 5, pp. 1079–1094, 2013.

[137] R. Tibshirani, « Regression Shrinkage and Selection via the Lasso », *Journal of the Royal Statistical Society. Series B (Methodological),* vol. 58, no. 1, pp. 267–288, 1996.

[138] K. Torkkola, « Feature extraction by non parametric mutual information maximization », *The Journal of Machine Learning Research,* vol. 3, pp. 1415–1438, 2003.

[139] C. Torrence and G. P. Compo, « A practical guide to wavelet analysis », *Bulletin of the American Meteorological Society,* vol. 79, no. 1, pp. 61–78, 1998.

[140] T. Tošić, K. K. Sellers, F. Fröhlich, M. Fedotenkova, P. beim Graben, and A. Hutt, « Statistical frequency-dependent analysis of trial-to-trial variability in single time series by recurrence plots », *Frontiers in Systems Neuroscience,* vol. 9, p. 184, 2016.

[141] G. E. Uhlenbeck and L. S. Ornstein, « On the theory of the brownian motion », *Physical Review,* vol. 36, no. 5, pp. 823–841, 1930.

[142] J. R. Vergara and P. A. Estévez, « A review of feature selection methods based on mutual information », *Neural Computing and Applications,* vol. 24, no. 1, pp. 175–186, 2014.

[143] C. L. Webber and J. P. Zbilut, « Dynamical assessment of physiological systems and states using recurrence plot strategies », *Journal of Applied Physiology,* vol. 76, no. 2, pp. 965–973, 1994.

[144] P. D. Welch, « The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms », *IEEE Transactions on Audio and Electroacoustics,* vol. 15, no. 2, pp. 70–73, 1967.

[145] M. A. Whittington, R. D. Traub, and J. G. R. Jefferys, « Synchronized oscillations in interneuron networks driven by metabotropic glutamate receptor activation », *Nature,* vol. 373, no. 6515, pp. 612–615, 1995.

[146] J. Xiao and P. Flandrin, « Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement », *IEEE Transactions on Signal Processing,* vol. 55, no. 6, pp. 2851–2860, 2007.

[147] J. Yan, Y. Wang, G. Ouyang, T. Yu, and X. Li, « Using max entropy ratio of recurrence plot to measure electrocorticogram changes in epilepsy patients », *Physica A: Statistical Mechanics and its Applications,* vol. 443, pp. 109–116, 2016.

[148] J. P. Zbilut and C. L. Webber, « Embeddings and delays as derived from quantification of recurrence plots », *Physics Letters A,* vol. 171, no. 3, pp. 199–203, 1992.

[149] A. M. Zbinden, S. Petersen-Felix, and D. A. Thomson, « Anesthetic depth defined using multiple noxious stimuli during isoflurane/oxygen anesthesia. II. Hemodynamic responses », *Anesthesiology,* vol. 80, no. 2, pp. 261–267, 1994.