Université de Montréal

Évaluation de programmes de prétraitement de signal d'activité électrodermale (EDA)

Par

Claudéric DeRoy

Département de psychologie, Faculté des arts et des sciences

Mémoire présenté en vue de l’obtention du grade M. Sc. en psychologie

août 2023

Université de Montréal

Département de psychologie, Faculté des arts et des sciences

Ce mémoire intitulé

Évaluation de programmes de prétraitement du signal d'activité électrodermale (EDA)

Présenté par

Claudéric DeRoy

A été évalué(e) par un jury composé des personnes suivantes

Floris Van Vugt

Président-rapporteur

Sébastien Hétu

Directeur de recherche

Alexandru Hanganu

Membre du jury

Résumé

L'activité électrodermale (EDA), particulièrement la *skin conductance response* (SCR), est un signal psychophysiologique fréquemment utilisé en recherche en psychologie et en neuroscience cognitive. L'utilisation de l'EDA entraîne son lot de défis particulièrement son prétraitement. En effet, encore très peu de recherches effectuent un prétraitement adéquat. Notre objectif est donc de promouvoir l'utilisation du prétraitement du signal SCR et de proposer des recommandations pour les chercheurs en fournissant des données sur l'impact du prétraitement sur la capacité à discriminer les SCR entre deux conditions expérimentales. En utilisant des travaux similaires, nous avons testé les effets de combinaisons de prétraitement utilisant différentes méthodes de filtrage, différentes méthodes de remise à l'échelle, l'inclusion d'une étape de détection automatique des artefacts de mouvement et en utilisant différentes métriques opérationnalistes (le *peak-scoring* (PS) et l'aire sous la courbe (AUC)) et d'approches par modèle. Enfin, nous avons testé si une seule combinaison de filtrage pourrait être utilisée avec différents jeux de données ou si le prétraitement devrait plutôt être ajusté individuellement à chaque jeu de données. Nos résultats suggèrent que 1) l'inclusion d'une étape de détection automatique des artefacts de mouvements n'affecte pas significativement la capacité à discriminer entre deux conditions expérimentales, 2) l'approche par modèle semble être un peu meilleure à discriminer entre deux conditions expérimentales et 3) la meilleure combinaison de prétraitement semble variée en fonction du jeu de données utilisé. Les données et outils présentés dans ce mémoire devraient permettre de promouvoir et faciliter le prétraitement du signal SCR.

Mots-clés : SCR, prétraitement de signal, approche par modèle, approche opérationnaliste, artefacts de mouvement.

Abstract

Electrodermal activity (EDA), particularly the skin conductance response (SCR) is a psychophysiological signal frequently used in research in psychology and in cognitive neuroscience. Nevertheless, using EDA comes with some challenges notably in regard to its preprocessing. Indeed, very few research teams adequately preprocess their data. Our objective is to promote the implementation of SCR preprocessing and to offer some recommendations to researchers by providing some data on the effect of preprocessing on the SCR ability to discriminate between two experimental conditions. Based on similar work, we have tested the effect of preprocessing combinations using different filtering methods, different rescaling methods, the inclusion of an automatic motion detection step while using different operationalist metrics (peak-scoring (PS) and area under the curve (AUC)) and different model-based approach metrics. Finally, we tested if only one combination could be used across different datasets or if the preprocessing should be optimized individually to each dataset. Our results show that 1) the inclusion of the automatic motion detection step did not significantly impact the ability to discriminate between two experimental conditions, 2) the model-based approach seems to be slightly better at discriminating between two experimental conditions and 3) the best combination of preprocessing seems to vary between different datasets. The data and tools presented in this master thesis should promote and facilitate SCR signal preprocessing.

Keywords : SCR, signal preprocessing, model-based approach, operationalist approach, motion artifacts

Table des matières

Liste des figures

List of figures in the scientific article

List of figures in the supplementary material

List of tables in the scientific article

List of tables in the supplementary materials

Liste des sigles et abréviations

AIC : *Akaike information criterion*

AUC : *Area under the curve*

BOLD : *Blood oxygen level dependent*

CS- : *Conditional stimulus absent*

CS+ : *Conditional stimulus present*

dB : *Decibel*

DCM : *Dynamic causal model*

EDA : E*lectrodermal activity*

EEG : *Electroencephalography*

fMRI : *Functional magnetic resonance imaging*

GLCM : *General linear convolutional model*

GLM : *General linear model*

Hz : Hertz

IAPS : *International affective picture system*

ISI : *Inter-stimulus interval*

ITI : *Inter-trial interval*

LTI : *Linear time invariant*

MEG : *Magnetoencephalography*

ML : *Machine learning*

MR : *Maximum response range*

ms : *miliseconds*

MSR : *Magnetically shielded room*

PM : *Processing method*

PS : *Peak-scoring*

PsPM : *Psychophysiological modelling*

rsm : *Rescaling method*

s : *seconds*

SCL : *Skin conductance level*

SCR : *Skin conductance response*

SNA : Système nerveux autonome

SOA : *Stimulus onset asynchrony*

SVM : *Support vector machine*

US : *Unconditional stimulus*

V : Volt

µV : *microvolt*

À l'honneur et à la mémoire de tous ceux et celles qui ont forgé la personne que je suis aujourd'hui et que je serais demain

Remerciements

Je tiens à remercier Sébastien Hétu, mon directeur de recherche, qui m'a pris sous son aile qui m'a surtout permis d'apprendre et de me développer en tant qu'étudiant gradué. Je n'avais définitivement pas les mêmes intérêts de recherche depuis qu'on s'était rencontré la première fois, mais il m'a toujours soutenu, peu importe ce que je voulais étudier.

J'aimerais remercier tous les enseignants, les enseignantes et les professeurs, les professeures que j'ai eus et qui m'ont partagé la passion et le plaisir qu'ils ont pour leur matière. Si je suis ici aujourd'hui c'est en grande partie à cause d'eux.

J'aimerais aussi dire un gros merci à ma famille et mes amis et amies qui ont bien souvent essayé de comprendre qu'est-ce que je fais dans vie, sans trop nécessaire comprendre mon parcours, mais qui m'ont toujours dit de ne pas lâcher et qui ont toujours cru en moi alors que des fois moi non.

Finalement, j'aimerais remercier ma mère et mon père qui m'ont appris une des choses les plus importantes : de faire ce qui me passionne dans la vie. Même si cette leçon-là n'a pas été facile pour eux et pour notre famille ça reste la chose qui me pousse dehors de mon lit le matin.

Introduction générale

Par sa facilité d'utilisation et son faible coût, l'activité électrodermale (EDA) est très souvent utilisée en recherche dans le domaine de la psychologie et des sciences cognitives. Cependant, son utilisation amène son lot de défis incluant le prétraitement du signal EDA. Dans une brève revue systématique de la littérature, Privratsky et al. (2020) ont montré que de grandes lacunes persistent dans le monde de la recherche concernant le prétraitement de l'EDA. L'objectif général de ce mémoire est d'étudier les impacts de l'utilisation de différentes combinaisons de prétraitement afin de fournir des arguments pour promouvoir l'importance du prétraitement du signal EDA, d'offrir des recommandations sur la façon d'effectuer ce prétraitement et enfin fournir des outils afin de le faciliter.

Le mémoire commence par une introduction générale qui propose d'abord une brève description de l'EDA et de son historique. Ensuite, les différentes techniques d'enregistrement, les différentes approches afin de mesurer l'EDA et les différentes étapes du prétraitement sont brièvement présentées. L'introduction se termine par la présentation de la problématique et des objectifs. Le corps du mémoire est constitué du manuscrit d'un article scientifique qui sera soumis pour publication. La dernière partie du mémoire présente une courte discussion générale dans laquelle les retombées et les limites du mémoire seront abordées.

L'EDA simplement

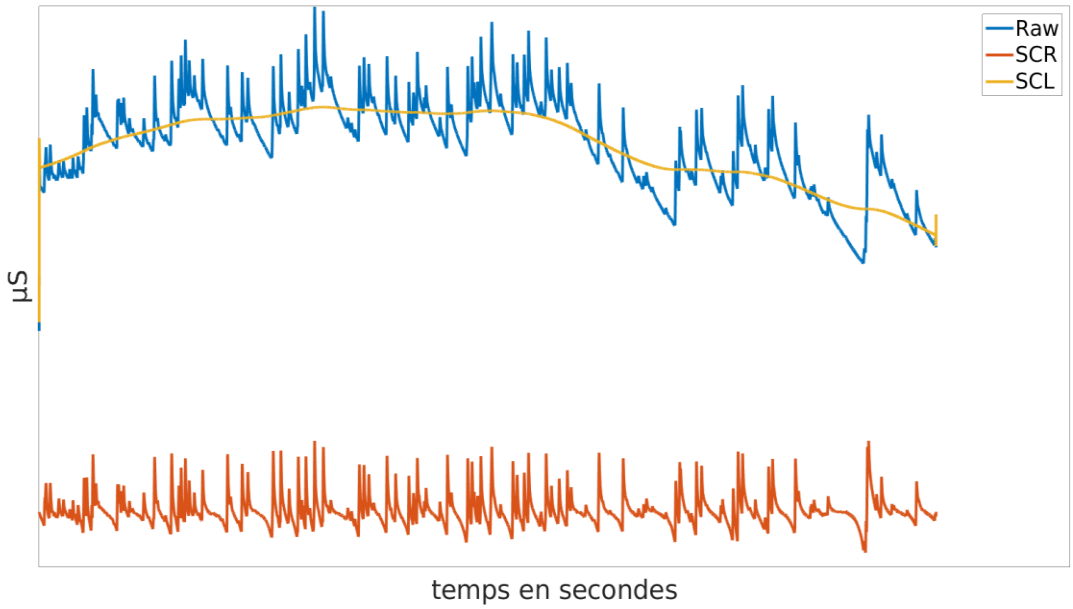
L'EDA provient de la mesure des variations de la conduction cutanée mesurée en microvolt (µV) et est utilisée en recherche comme une mesure indirecte du système nerveux autonome (SNA). Les variations des propriétés conductrices de la peau à l'origine du signal EDA sont dues entre autres au relâchement de la sueur par les glandes eccrines (Boucsein, 2012). Les glandes eccrines sont innervées par des neurones efférents du SNA qui projettent leurs axones par

la voie sympathique spinale descendante (Boucsein, 2012). À ce point, une connexion avec un neurone qui passe par la racine ventrale est faite et qui se propage à travers le tronc sympathique et le nerf spinal mixte pour finalement aller innerver les glandes eccrines de la peau (Boucsein, 2012). Pour de plus amples détails, le lecteur est dirigé vers la lecture de Boucsein (2012).

Le signal EDA est composé de deux signaux offrant de l'information différente. Le *skin conductance level* (SCL) représente les changements lents des capacités conductrices de la peau, il est associé plus à l'état général de la personne (p. ex. le niveau de stress général chez la personne durant la journée). Le *skin conductance response* (SCR) représente les changements rapides des capacités conductrices de la peau, il est associé aux évènements vécus (p. ex. exposition à une image aversive). Le mémoire portera en particulier sur le SCR (voir Fig. 1).

Figure 1

Décomposition du signal EDA en SCL et SCR



Note. Exemple des deux signaux de l'EDA. En bleu, le signal brut où l'on peut clairement voir les fluctuations lentes du niveau de base et les fluctuations rapides notamment reliées à la perception de stimuli. En jaune, le *skin conductance level* (SCL; pour lequel les fluctuations rapides ont été ôtées) et en orange le *skin conductance response* (SCR; pour lequel les fluctuations lentes ont été ôtées). Les données ayant servi à cette figure ont été tirées du jeu de données HRA-1 (Bach, Daunizeau et al., 2010).

L'histoire de l'EDA

Des conditions historiques et culturelles auront attribué sans nuances l'avènement de l'EDA à Vigouroux (Neumann & Blanton, 1970), d'autres l'attribuèrent plus à Dubois-Reymond (Boucsein, 2012). Évidemment, l'histoire est beaucoup plus nuancée et compliquée. Au 19e siècle, les relations entre la France et l'Allemagne ne sont pas les meilleures et donc la correspondance entre les scientifiques français et allemands n'était pas non plus importante (Neumann & Blanton, 1970). Ainsi, deux pôles se développèrent avec un certain parallélisme, le français et l'allemand. En Allemagne, Dubois-Reymond fût le premier à travailler sur les

phénomènes électrophysiologiques en 1849, mais ils attribuèrent incorrectement le phénomène qu'ils observèrent au potentiel d'action des muscles (Boucsein, 2012). Ensuite, ce fut Hermann qui démontra que l'activité que Dubois-Reymond avait attribuée au potentiel musculaire était en réalité attribuable à la sécrétion de la peau due à l'activité des glandes de la peau (Neumann & Blanton, 1970). En France, c'est Vigouroux qui attribua ses observations d'activité électrodermale à un facteur psychologique vers 1879 (Neumann & Blanton, 1970). Quelque temps après Vigouroux, c'est Féré qui commença à étudier l'EDA, il effectua des expériences sur la présentation de stimuli émotionnels chez des patients hystériques et observa une relation avec l'EDA. Ensuite, un Russe, Tarchanoff, réussit à joindre le travail français et allemand sous une seule et même idée grâce à sa familiarité du travail français et allemand (Neumann & Blanton, 1970). Essentiellement, l'idée française concevait le corps humain comme un conducteur, sans tenir compte de l'innervation des glandes eccrines alors que l'idée allemande concevait que c'était la sécrétion produite par les glandes eccrines qui produisait le signal EDA. Ainsi, la théorie de Tarchanoff fut la première théorie à faire cette relation à l'époque et à décrire l'EDA comme le résultat des glandes eccrines qui sont dépendantes de l'activité nerveuse (Boucsein, 2012). Pour les lecteurs souhaitant en apprendre davantage Boucsein (2012) et Neumann & Blanton (1970) sont des ouvrages inestimables à cet égard.

Les techniques d'enregistrement de l'EDA

Afin d'utiliser l'EDA comme mesure de l'activité du SNA, il faut donc mesurer les variations de la conductance de la peau. Trois types d'enregistrements existent actuellement : endosomatiques, exosomatiques à courant continu et exosomatique à courant alternatif (Boucsein, 2012). Des trois techniques, l'enregistrement exosomatique à courant continu est la plus utilisé (Posada-Quintero & Chon, 2020) et aussi le plus souvent recommandé (Boucsein,

2012; Boucsein et al., 2012). L'enregistrement exosomatique à courant continu utilise un courant continu et constant généralement de 0.5 V (Boucsein, 2012) appliqué sur deux électrodes. Une troisième électrode, une électrode de référence, est aussi employée afin de bien pouvoir mesurer la différence de courant entre les deux électrodes (Boucsein et al., 2012). Le signal résultant provient de l'enregistrement du courant dans le circuit. Par la loi de Ohm et puisque le courant est connu et constant (p. ex. de 0.5 V), le courant *I* est donc proportionnel à l'inverse de la résistance (c.-à-d. $1/R$) ce qui équivaut à *G* la conductance (c.-à-d. la conductance de la peau) (Boucsein, 2012). Enfin, le courant *I* est donc proportionnel à la conductance *G* (c.-à-d. la conductance de la peau) : plus le courant passe, plus la conductance est élevée (Boucsein et al., 2012). Ainsi, lorsque le SNA s'active, il y a une augmentation de la sueur sur la peau qui entraîne une augmentation de la conductance qui peut être mesurée par des fluctuations dans les enregistrements exosomatiques à courant continu.

Les différentes métriques pour mesurer le SCR

Plusieurs façons de quantifier le signal SCR existent (Boucsein, 2012), dont deux grandes familles d'approches. La première, l'approche opérationnaliste, englobe toutes les techniques qui opérationnalise un construit psychologique (p. ex. la peur) par une seule métrique calculée par un petit nombre de points du signal (p. ex. l'amplitude) (Bach et al., 2018). Le lien entre le concept et la réponse est proportionnel (c.-à-d. plus la personne a peur, plus un SCR augmente et par conséquent la métrique aussi). L'approche opérationnaliste est simple et facile à implémentée et donc plus souvent utilisée (Posada-Quintero & Chon, 2020; Privratsky et al., 2020). La seconde, l'approche par modèle, utilise des modèles mathématiques afin d'estimer l'amplitude de l'activité neuronale à l'origine du SCR observé. Cette approche est beaucoup plus complexe et requière habituellement plus de capacités de traitement de données (c.-à-d. des ordinateurs puissants et un

temps considérable de computation). À l'instar de l'approche opérationnaliste, plus le concept est intense (c.-à-d. la peur) plus la réponse électrodermale estimée par l'approche par modèle sera importante.

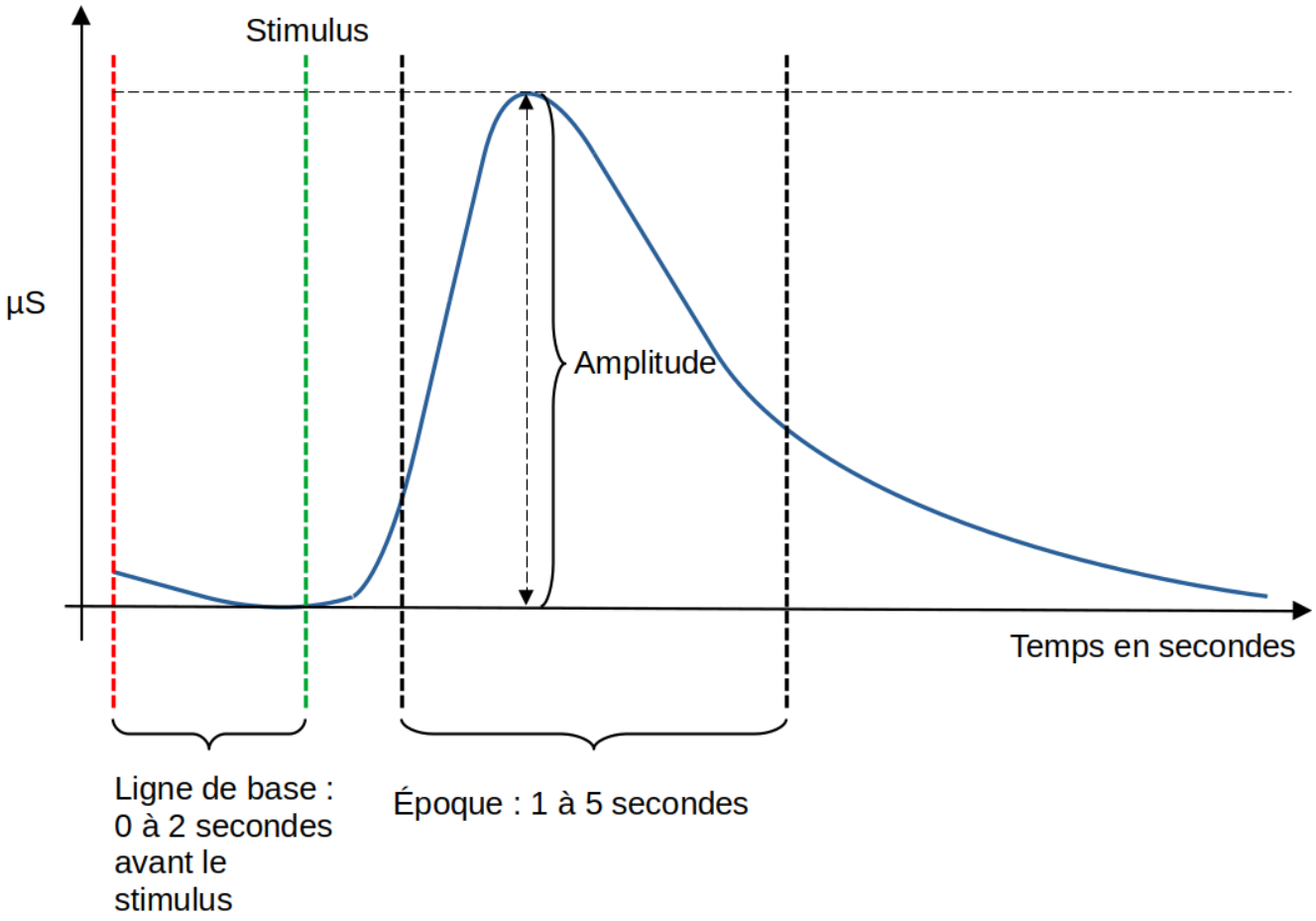
Approche opérationnaliste

Le *peak-scoring* (PS)

Le *peak-scoring* (PS) ou encore l'amplitude est une métrique de l'approche opérationnaliste qui consiste essentiellement à une mesure d'amplitude du SCR (Boucsein et al., 2012; Posada-Quintero & Chon, 2020; Privratsky et al., 2020). Généralement, il est calculé par la différence entre le maximum du signal mesuré dans une période après le stimulus (c.-à-d. une époque) et la mesure d'une ligne de base. Cependant, il existe plusieurs façons de le calculer et il existe aussi plusieurs façons de définir le début, la fin et la durée autant de l'époque que de la ligne de base (Bach et al., 2013; Boucsein, 2012; Boucsein et al., 2012; Braithwaite et al., 2013; Figner & Murphy, 2011; Posada-Quintero & Chon, 2020; Privratsky et al., 2020). Par exemple, comme ligne de base, certains utilisent la valeur minimale du signal de 2 s avant l'apparition du stimulus jusqu'à l'apparition du stimulus alors que d'autres vont plutôt calculer la moyenne du signal durant cette période pour obtenir la valeur de la ligne de base. Pour l'époque, le *maximum response range* (MR) correspond à l'étendue de l'époque après l'apparition du stimulus (p. ex. 1 s à 5 s après l'apparition du stimulus) (voir Fig. 2 pour un exemple).

Figure 2

Exemple de mesure du PS



Note. Ici, l'amplitude est calculée en effectuant la différence entre le maximum du signal obtenu dans une époque allant de 1 s à 5 s (MR=5) après l'apparition du stimulus et la valeur minimale obtenue entre 2 s et 0 s avant l'apparition du stimulus.

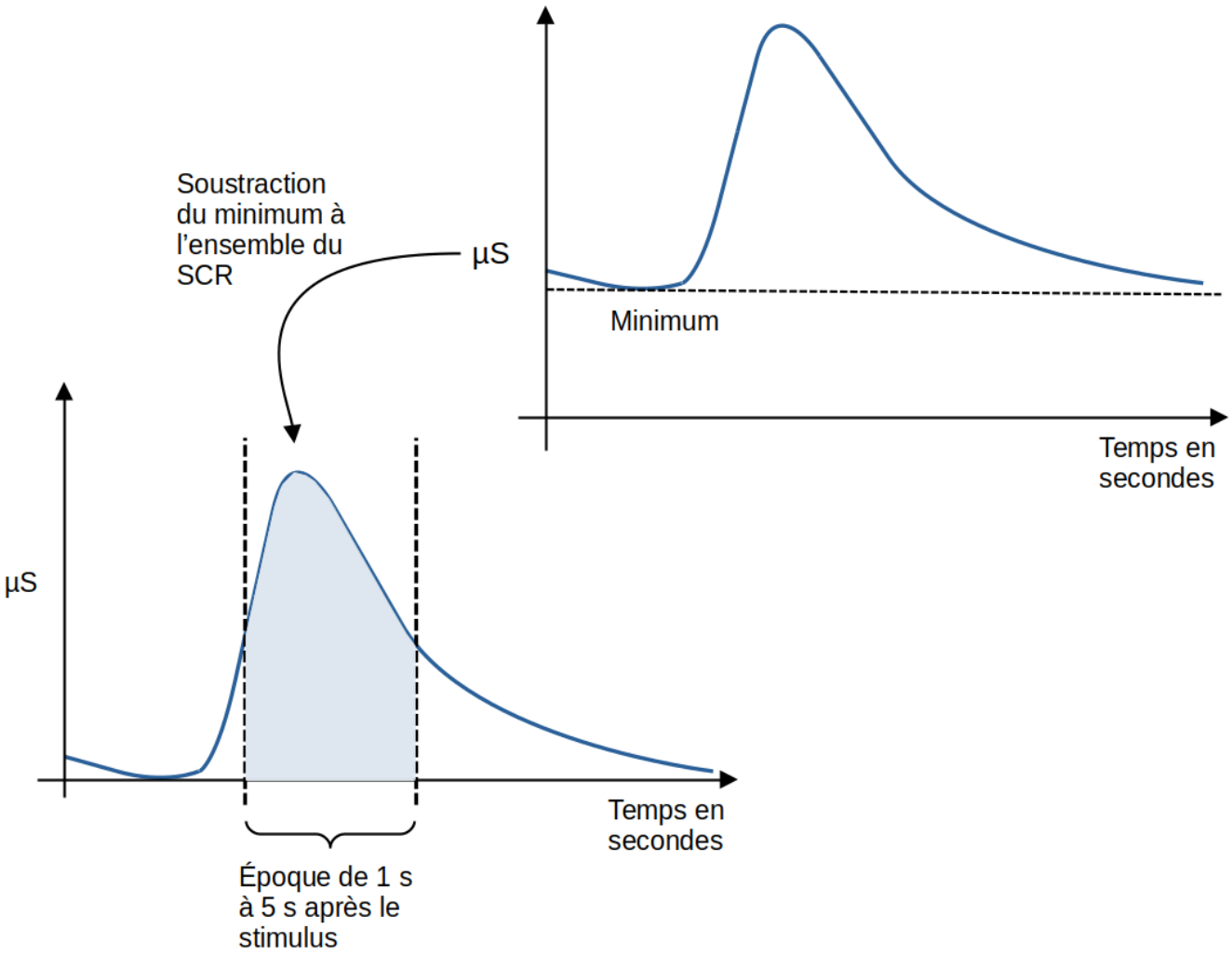
L'aire sous la courbe (*area under the curve*;AUC)

L'aire sous la courbe (AUC) est une autre métrique de l'approche opérationnaliste qui quoique moins utilisée que le PS dans la littérature semblerait meilleur que ce dernier (Shukla et al., 2021). L'AUC est obtenue en calculant l'aire sous la courbe d'une époque du SCR. Comme pour le PS, il existe plusieurs façons de calculer l'AUC (Bach, Friston et al., 2010; Boucsein, 2012; Shukla et al., 2021). Évidemment, puisqu'il est essentiellement impossible de déterminer avec exactitude la fonction mathématique du SCR et donc qu'une

intégrale n’est pas possible, les techniques de calcul de l’AUC reposent entièrement sur des approximations (Boucsein, 2012). À notre connaissance, personne n’utilise la somme de Riemann (c.-à-d. la méthode du point médian) afin d’approximer la valeur de l’AUC. Plutôt, certains utilisent comme approximation l’aire d’un rectangle dont la hauteur est l’amplitude de la réponse et la largeur déterminée par les points où la montée de la réponse atteint 50% de son amplitude et où la descente de la réponse atteint 50% de son amplitude. D’autres, comme c’est le cas dans ce mémoire, utilisent une approche développée par Bach, Friston et al. (2010) qui consiste à soustraire le minimum d'un SCR à l’ensemble du SCR et ensuite calculer sa moyenne (voir Fig. 3 pour un exemple).

Figure 3

Exemple de mesure de l’AUC



Note. L'aire sous la courbe (AUC) représente l'aire sous la réponse SCR représentée en bleu dans la figure. Dans le cadre de ce mémoire, cette valeur sera estimée en soustrayant chaque point de la réponse SCR à la valeur minimum et en effectuant ensuite une moyenne de ces différences (Bach, Friston et al., 2010).

Une critique souvent faite de l'approche opérationnaliste est qu'elle vise à opérationnaliser un concept psychologique latent (p. ex. la peur) qui est généralement très complexe, cette information est en fait réduite, voire perdue, puisque ces métriques ne sont calculées que par un nombre très restreint de points du signal (Bach et al., 2018). Afin de résoudre le problème réductionniste de l'approche opérationnaliste, l'approche par modèle fait appel à des modèles mathématiques qui utilisent l'ensemble des points du signal.

Approche par modèle

Comme mentionné plus tôt, l'approche par modèle cherche à pallier aux limites de l'approche opérationnaliste. L'approche opérationnaliste se base essentiellement sur l'idée suivante : l'information capturée par un évènement observable (p. ex. le SCR) est considéré comme informative d'un évènement inobservable (p. ex. la peur) (Bach & Friston, 2013). Les chercheurs vont donc considérer que l'évènement observable est informatif de l'évènement inobservable (Bach & Friston, 2013). Toutefois, Bach & Friston (2013) révèlent deux grandes limites de l'opérationnaliste. Premièrement, l'approche opérationnaliste utilise des hypothèses qui ne sont pas falsifiables, en fait les liens causaux entre l'évènement observable et inobservable ne sont pas spécifiés et donc pas testables ou rejetables. Deuxièmement, l'approche opérationnaliste ne spécifie pas la manière dont le lien entre l'évènement observable et inobservable sont liés. À titre d'exemple, si l'on prétend que le SCR permet de mesurer la peur qu'on peut représenter par la relation g suivante : $peur \overset{g}{\Rightarrow} EDA$ il est difficile de savoir avec exactitude qu'elle est la relation inverse de g (c.-à-d. g^{-1}) : $peur \overset{g^{-1}}{\Leftarrow} EDA$ implique à cause du bruit, mais aussi de la nature compliquée de la relation g. L'approche par modèle cherche justement à rendre cette relation inverse prédictive. L'approche par modèle se base sur deux modèles mathématiques– le *linear time invariant model* (LTI) et un modèle neuronal– afin de rendre intelligible la relation, ou chemin, inverse entre le SCR et n'importe quel concept étudié conjointement avec ce dernier. Le résultat final obtenu par l'approche par modèle est une estimation de l'amplitude de l'activité neuronale qui est à l'origine de la production du SCR mesuré. Pour plus de détails sur cette approche, voir (Bach et al., 2018; Bach & Friston, 2013).

Avant de calculer ces différentes métriques et d'effectuer des analyses statistiques, comme la plupart des signaux utilisés en recherche en psychologie et en neuroscience cognitive (Esteban

et al., 2019; Gramfort et al., 2013; Lakshmi et al., 2014; Liu, 2016), le signal SCR devrait être prétraité. Le prétraitement du signal SCR fera l'objet de la prochaine section.

Prétraitement du signal SCR

Le SCR devrait être soumis à un prétraitement aussi rigoureux que possible afin de mesurer les concepts à l'étude. Dans le cadre de ce mémoire nous nous intéresserons à trois étapes du prétraitement : 1) la gestion des artefacts de mouvement, où les artefacts dus au mouvement du participant doivent être retirés afin de ne pas fausser les résultats des analyses statistiques; 2) le filtrage qui permet de réduire ou d'éliminer certains bruits qui se retrouvent dans le signal SCR, mais qui n'intéressent pas les chercheurs; et 3) la remise à l'échelle qui permet de réduire la variance intersujet.

Identification et retrait des artefacts de mouvement

Les artefacts de mouvement sont produits généralement des deux façons suivantes : 1) lorsqu'une ou des électrodes placées sur le participant bouge(nt) ; 2) lorsque la pression entre l'électrode et la peau du participant change (Taylor et al., 2015). De plus, le frottement de soulier avec des semelles en caoutchouc produit aussi des artefacts (Braithwaite et al., 2013). Les mouvements du participant doivent donc être gardés au minimum ce qui peut représenter un défi considérable dans le cas d'étude ambulatoire ou écologique où le participant doit bouger. Le problème avec les artefacts de mouvement est qu'ils ressemblent beaucoup à des SCR (c.-à-d. du signal qui intéresse les chercheurs) sans en être pour autant (Taylor et al., 2015). Encore plus problématique, le filtrage du signal n'est pas suffisant pour retirer ces artefacts du signal et par conséquent, ils doivent être retirés en inspectant souvent manuellement le signal, en se basant sur des critères prédéterminés afin de décider si une partie du signal possède un artefact

ou non (Taylor et al., 2015). Malheureusement, beaucoup de techniques de gestion des artefacts de mouvement requièrent beaucoup de temps et d'effort pour les chercheurs puisqu'elles ne sont pas automatisées (Taylor et al., 2015; Tronstad et al., 2015). De plus, les approches reposant sur l'inspection visuelle du signal sont plus sensibles à la subjectivité. Aussi, différentes équipes ont mis au point des techniques automatisées afin de faciliter et de rendre plus objective cette étape du prétraitement (Taylor et al., 2015; Tronstad et al., 2015). Dans le cadre de ce mémoire, nous utiliserons un algorithme d'apprentissage machine (ML) développé par Taylor et al. (2015) qui permet de détecter automatiquement des artefacts de mouvement. Nous avons choisi d'utiliser l'algorithme binaire afin d'être plus conservateurs dans la détection des artefacts.

Filtrage

Le filtrage du signal SCR a plusieurs utilités. Il sert notamment à réduire au maximum le bruit et à augmenter la qualité du signal afin qu'il soit le plus proche de la réalité. Malgré son rôle important et la publication de travaux démontrant son utilité et proposant des recommandations sur son utilisation (Boucsein et al., 2012; Braithwaite et al., 2013; Figner & Murphy, 2011; Privratsky et al., 2020), plusieurs chercheurs n'appliquent toujours pas de filtrage avant d'analyser leurs données SCR (Privratsky et al., 2020). De plus, bien que plusieurs méthodes de filtrage soient disponibles, ceux qui filtrent leur signal EDA utilisent très souvent un filtrage passe-bas (Privratsky et al., 2020). Le filtrage passe-bas n'affecte pas les composantes de basse fréquence du signal et il est utilisé pour retirer plusieurs sources de bruit du signal comme le bruit provenant des lignes électriques et même dans certains cas le bruit provenant de champ magnétique (p. ex. en enregistrement simultané en IRMf) (Boucsein, 2012). Le filtre passe-haut est l'opposé du filtre passe-bas puisqu'il n'affecte pas les composantes du signal de haute fréquence, mais bloque les composantes de basse fréquence. Le filtre passe-haut peut être utilisé pour retirer certains artefacts (Boucsein, 2012; Braithwaite et al., 2013), mais aussi pour retirer la

composante tonique du signal (c.-à-d. le SCL) (Figner & Murphy, 2011). Enfin, le filtrage passe-bande consiste à l'implémentation d'un filtrage passe-bas et d'un filtrage passe-haut, il peut également être utilisé. Notez que dans certains cas, d'autres types de filtrage peuvent être également utilisés, par exemple le filtrage médian lorsque l'EDA est enregistré en même temps que le signal IRMf (Staib et al., 2015). Dans le cadre de ce mémoire, nous nous intéresserons seulement aux filtres passe-bas, passe-haut et passe-bande.

Remise à l'échelle (*rescaling*)

Braithwaite et al. (2013) et Privratsky et al. (2020) qui citent Lykken & Venables (1971) précisent l'importance de la remise à l'échelle lors du prétraitement afin de retirer la variance intersujet avant d'effectuer n'importe quelle analyse statistique. Privratsky et al. (2020) proposent sept méthodes de remise à l'échelle portant soit sur le signal brut ou sur les réponses obtenues du signal analysé (c.-à-d. les métriques). Braithwaite et al. (2013) proposent d'autres méthodes, mais toujours avec le même objectif d'éliminer la variance intersujet. Dans le cadre de ce mémoire, nous utiliserons les différentes méthodes proposées par Privratsky et al. (2020).

Pour chacune de ces étapes du prétraitement, plusieurs paramètres peuvent être ajustés (p. ex., la fréquence de coupure du filtre passe-haut) et les étapes peuvent être combinées de multiples manières afin de déterminer le pipeline de prétraitement qui sera utilisé.

Problématique

Malgré les connaissances sur les différentes étapes du prétraitement en EDA (p. ex. Boucsein et al. (2012), Braithwaite et al. (2013), Figner & Murphy (2011) et Privratsky et al. (2020)) ainsi que le développement d'outils pour réaliser (p. ex. Bach, Flandin et al. (2010), Chaspari et al. (2016), Greco et al. (2016), Posada-Quintero et al. (2016), Taylor et al. (2015) et Tronstad et al. (2015)) le prétraitement du signal EDA, il existe encore d'importantes lacunes

concernant le prétraitement du signal EDA dans le milieu de la recherche (Privratsky et al., 2020). En effet, dans une étude contenant une brève revue de la littérature sur les habitudes de prétraitement de signal EDA de 65 articles entre 2019 et 2020, Privratsky et al. (2020) concluent qu'il n'y avait toujours pas de consensus dans la littérature scientifique quant à la bonne méthode de prétraitement de signal EDA. Plus précisément, la revue de littérature de Privratsky et al. (2020) montre que 36 articles parmi 44 qui n'avaient pas utilisé de filtrage ont utilisé des techniques d'inspection visuelle, de retrait manuel d'artefact et d'interpolation comme seul prétraitement. Privratsky et al. (2020) soulignent que toutes ces méthodes sont très subjectives et en proie aux biais. De plus, des 61 articles qui ont rapporté avoir eu des problèmes concernant la qualité des données ou qui ont dû rejeter des participants dus à dû mauvais signal, seulement 54% ont rapporté avoir utilisé du filtrage et que 23% d'entre eux ont utilisé un filtre passe-haut (Privratsky et al., 2020). Des 56 articles qui ont rapporté avoir utilisé du filtrage, de basses fréquences de coupure de trois ordres de grandeurs différentes ont été utilisées (p. ex. 0.1 Hz, 1 Hz et 100 Hz) (Privratsky et al., 2020). Privratsky et al. (2020) suggère que ce manque de consensus est probablement expliqué par une certaine incompréhension du prétraitement et de ses effets potentiels sur le signal – certains chercheurs croyant que le prétraitement altère (négativement) le signal ou qu'il n'est tout simplement pas nécessaire.

Afin de remédier à ces inquiétudes et incompréhensions, dans ce même article, Privratsky et al. (2020) ont tenté de démontrer l'effet sur les métriques du SCR de différentes étapes de prétraitement touchant les méthodes de filtrage et de remise à l'échelle. Ils ont ainsi testé les effets de trois techniques de filtrage, de sept types de remise à l'échelle ainsi que deux méthodes de quantification du signal (c.-à-d. le PS et une approche par modèle). De plus, pour le PS, ils ont testé différents MR. Pour quantifier ces effets, ils ont testé l'ensemble des combinaisons possibles

sur la taille d'effet des contrastes entre les réponses SCR à évènements aversifs vs neutres (c.-à-d. la capacité à discriminer ces deux conditions expérimentales) d'un jeu de données où le signal EDA était enregistré de manière concomitante avec le signal IRMf. En se basant sur leurs résultats, Privratsky et al. (2020) proposent plusieurs recommandations concernant le prétraitement du signal EDA en général, mais aussi plus particulièrement dans le contexte d'enregistrements concomitants EDA-IRMf. Ainsi, les auteurs proposent d'utiliser pour des SCR mesurés par l'approche par modèle un filtrage passe-bande avec une basse fréquence de coupure de 5 Hz et une haute fréquence de coupure entre 0.01 Hz et 0.03 Hz ainsi que l'utilisation d'une remise à l'échelle correspondant à un score z effectué sur le signal brut (p. ex., pas sur les réponses transformées). Pour ce qui est des SCR mesurés par le PS, ils proposent encore une fois un filtrage passe-bande, mais avec une basse fréquence de coupure de 5 Hz, mais une haute fréquence de coupure de 0.02 Hz ainsi qu'une remise à l'échelle utilisant un score z effectué sur les réponses obtenues du signal (p. ex. le PS obtenu du signal).

Il est important de noter que malgré les recommandations récentes de Privratsky et al. (2020), il y a encore, aujourd'hui, de grandes différences concernant le prétraitement du signal EDA dans la littérature (voir la section *Introduction* de l'article qui suit) suggère que la manière d'effectuer ce prétraitement, mais également que l'importance de cette étape demeure très méconnue des chercheurs et chercheuses qui l'utilisent ou qui voudraient le faire. Puisque le prétraitement est central dans l'obtention de données de qualité, une meilleure connaissance sur le prétraitement du SCR pourrait avoir des répercussions positives importantes dans les domaines de recherche utilisant cette approche.

Objectifs

Nos objectifs généraux sont donc de promouvoir l'utilisation du prétraitement du signal SCR et de proposer des recommandations pour les chercheurs et chercheuses en fournissant des données sur l'impact du prétraitement sur la capacité à discriminer les SCR entre deux conditions expérimentales. Ainsi, nous visons à poursuivre et complémenter le travail de Privratsky et al. (2020) en utilisant la même méthodologie que ces derniers, mais tout en ajoutant une étape supplémentaire de prétraitement, une nouvelle métrique et de tester toutes ces différentes combinaisons avec des jeux de données différents. Nous avons donc au total trois objectifs spécifiques.

Premièrement, malgré le travail important de Privratsky et al. (2020), ils n'ont pas testé de méthode de gestion des artefacts de mouvement. Comme mentionné plus tôt, les artefacts de mouvement peuvent survenir de plusieurs façons et être interprétés comme du signal d'intérêt (c.-à-d. un SCR) sans en être. Nous avons donc comme objectif d'ajouter une étape de gestion des artefacts de mouvement par l'utilisation de l'algorithme de Taylor et al. (2015) et d'en tester l'impact sur un pipeline de prétraitement du signal EDA.

Deuxièmement, comme mentionné dans la section sur les métriques, il existe deux grandes familles d'approches : l'approche opérationnaliste et l'approche par modèle. Dans leur article, Privratsky et al. (2020) ont seulement utilisé le PS comme métrique de l'approche opérationnaliste. Nous allons donc ajouter l'AUC comme métrique supplémentaire. Shukla et al. (2021) ont montré la supériorité de l'AUC par rapport au PS comme métrique utilisée par un algorithme ML dans la reconnaissance d'émotion. De plus, Bach, Friston et al. (2010) ont théorisé que l'AUC pourrait être utilisée comme métrique, mais sans toutefois le tester. En plus, à notre connaissance aucune recherche n'a été faite sur la comparaison entre l'AUC et l'approche

par modèle. Ainsi, en rajoutant l'AUC, nous avons comme objectif de tester les différentes combinaisons de filtrage sur l'AUC ainsi que de comparer cette métrique opérationnaliste au PS et à l'approche par modèle.

Troisièmement, Privratsky et al. (2020) ont testé leurs différentes combinaisons de prétraitement sur un seul jeu de données limitant ainsi la généralisabilité des recommandations qu'ils ont émises à d'autres jeux de données. De plus, le jeu de données que Privratsky et al. (2020) ont utilisé provient d'enregistrement d'EDA avec un enregistrement simultané d'IRMf, ce qui est peu représentatif des conditions générales d'enregistrement d'EDA. De plus, Privratsky et al. (2020) ont aussi émis une limite par rapport à leurs recommandations, expliquant les résultats qu'ils ont obtenus ne sont pas nécessairement applicables à tous les designs expérimentaux. Cette limite selon eux s'applique particulièrement aux designs avec de longs *inter-trial interval* (ITI) (Privratsky et al., 2020). En fait, ils recommandent aux chercheurs avec de long ITI de trouver les paramètres de filtrage les plus adaptés à leurs données. Ainsi, nous allons tester toutes les différentes combinaisons de prétraitement sur trois jeux de données collectées sans IRMf (c.-à-d. dans un contexte plus typique) et publiquement accessibles en ligne. Donc, notre objectif est de tester si la meilleure combinaison pour prétraiter le signal SCR est généralisable ou devrait plutôt être ajustée en fonction du jeu de données.

Essentiellement, nous voulons pouvoir offrir de grandes lignes directrices concernant le prétraitement du signal EDA afin de promouvoir l'importance du prétraitement du signal EDA. De plus, nous allons fournir un outil gratuit et publiquement accessible en ligne qui pourra être utilisé afin de faciliter le prétraitement de l'EDA et l'étude de ses effets sur les données (https://github.com/neurok8050/eda-optimisation-processing-tool).

Apport à l’article

L’article au coeur de ce mémoire a été fait durant la maîtrise de Claudéric DeRoy. Le projet d’étudier le prétraitement du signal d’EDA provient de Sébastien Hétu qui avait déjà auparavant proposé un projet similaire, mais de moins grandes envergures à Claudéric durant son baccalauréat. Claudéric est le premier auteur de cet article, il a écrit l’entièreté du code qui teste les différentes combinaisons de prétraitement ainsi que le code qui effectue les analyses statistiques. De plus, il a aussi écrit l’article avec l’aide, le support et les conseils de Sébastien Hétu. L’article devrait être prochainement publié.

Scientific article

1.Introduction

Electrodermal activity (EDA) is a psychophysiological signal that has been used in research for more than 100 years including in the fields of psychology and cognitive science. In particular, skin-conductance response (SCR) is widely used as a measure of sympathetic arousal. SCR signal is quite easy to acquire, inexpensive and extremely noninvasive (Boucsein, 2012) which explains its widespread use. However, even with the publication of guidelines (Boucsein et al., 2012; Braithwaite et al., 2013; Figner & Murphy, 2011) including more recent work (Privratsky et al., 2020), issues related to preprocessing still hinder the quality of the SCR data being analyzed in several studies. Hence, we propose to add to previous efforts by providing researchers with new data-based suggestions and tools about how to preprocess SCR signals.

Similar to fMRI, EEG or MEG signals (Esteban et al., 2019; Gramfort et al., 2013), SCR signal needs to be preprocessed in line with particularities related to its acquisition (Boucsein, 2012). SCR preprocessing involves different steps such as resampling, filtering and artifact removal (Boucsein et al., 2012; Braithwaite et al., 2013; Figner & Murphy, 2011). Rescaling focuses on inter-subject standardization by removing inter-subject variance in baseline SCR. Filtering is used to attenuate the noise that is not part of SCR (e.g. electric noise from the power outlet). Artifact removal is done to remove a portion of the signal that is not related to physiological responses and often produced by movement during the acquisition. Even though many standard practices have been proposed (Boucsein, 2012; Boucsein et al., 2012; Braithwaite et al., 2013; Figner & Murphy, 2011), a short systematic literature review by Privratsky et al. (2020) looking at the articles published between 2019 and 2020, showed that there was still no consensus in the scientific community concerning the importance of doing any preprocessing at all nor the proper way to preprocess SCR signals. Privratsky et al. (2020) suggested that this lack

of consensus was due to misconceptions about filtering and its impacts–that filtering is not necessary or even that filtering can alter the signal.

To address these concerns, Privratsky et al. (2020) demonstrated the effect of different preprocessing approaches on SCR signals and analyses. To do so, they used a previously collected SCR dataset recorded during a functional magnetic resonance imaging (fMRI) session when participants completed a task in which they had to go through a fear-conditioning and extinction paradigm: participants had to look at 3 sec presentations of geometric stimulus, one conditional stimulus that could predict an electric shock (CS+; conditioned aversive stimulus) and the other a conditional stimulus that did not predict an electrical shock (CS-; neutral stimulus) (Privratsky et al., 2020). To test the effects of different preprocessing combinations, they used three nested filter methods, seven different rescaling methods (rsm) and two quantification metrics often used in SCR studies: the peak-scoring (PS) and a general linear convolutional model (GLCM). They also added different maximum-response ranges (MR) for the PS. Maximum-response range is the timing from which the metric will be computed from the signal (e.g. from 1 s after the stimulus to 3.5 s after the stimulus). They tested the impact of all the different combinations on the effect size of the contrasts of interest (i.e. aversive vs. neutral stimuli). Based on their empirical results, Privratsky et al. (2020) proposed some general recommendations regarding preprocessing of SCR signal for PS: an MR of 5 s (if the CS-US interval is of 2.5 s), a bidirectional Butterworth band-pass filter with a high cutoff frequency of 0.02 Hz and a low cutoff frequency of 5 Hz and a z-score rescaling done on all the responses (Privratsky et al., 2020). For the GLCM metrics, they again proposed using a band-pass filter but with a high cutoff frequency between 0.01 Hz and 0.03 Hz and a low-pass with a low cutoff frequency of 5 Hz and using a z-score rescaling done on all the responses. Importantly, Privratsky et al. (2020) stated that one of the limits of their work was that their findings might not be

applicable to all experimental designs especially those with longer inter-trial intervals (ITI). Thus, they recommended researchers to try different filter parameters to find the best one for their own experimental design (Privratsky et al., 2020).

Since these recommendations have been published, an overview of some recent articles found on APA PsycNet (research done on 2023-02-14) suggests that there be still a wide variety of preprocessing practices being used in SCR studies. While some of the articles between 2020 and 2023 mention using the recent recommendation stemming from Privratsky et al. (2020) (Bilodeau-Houle et al., 2023; Wehrli et al., 2022) or used a software called Ledalab which uses low-pass filters, downsamples the signal and makes artifact correction (Cuve et al., 2023; Moretta et al., 2023), there are still several articles that only mention downsampling (Koppold et al., 2022; Starita et al., 2022), only using a smoothing filter (Liberman & Dubovi, 2022), only using a low-pass filter (Sklivanioti Greenfield et al., 2022; Stefaniak et al., 2021), or even mention not preprocessing their data (Harnett et al., 2023). Thus, there is still a clear need for the development and dissemination of better evidence-based practices regarding the preprocessing of SCR signals.

To address this issue, we propose to add to Privratsky's important work by using the same methodology but adding extra preprocessing steps, new metrics and testing the impacts of preprocessing approaches on different datasets. We therefore have three main objectives.

First, while Privratsky et al. (2020) did not directly look at artifact identification and removal, we will add a preprocessing step for detecting and removing motion artifacts with a machine learning (ML) algorithm developed by Taylor et al. (2015) and test its impacts. Artifacts may arise from many different reasons even if the participant is restrained or told to not move. These reasons include simple things like rubber-sole shoes being moved (Braithwaite et al., 2013), muscle activity (Boucsein et al., 2012), an uncomfortable setup for the participant that might make him/her wiggle out of discomfort (Figner & Murphy, 2011) or even just the

movements required to do the experimental task (Boucsein, 2012). All of these situations make SCR easily affected by artifacts produced by voluntary or involuntary movements from the participant. Motion artifacts can be mistaken as genuine psychophysiological reactions to a stimulus (SCR) which can lead to incorrect results and to misinterpretations (Taylor et al., 2015). Often, motion artifact detection is done manually by visually inspecting the signal which takes a significant amount of time and is easily biased by subjective factors (Taylor et al., 2015). In order to automatically and efficiently clean SCR signal from motion artifacts, Taylor et al. (2015) developed a ML algorithm aimed at detecting motion artifacts and identifying the epochs where those artifacts occur in order to remove them. Hence, our first main objective is to test the impacts of adding an artifact detection step during the preprocessing of SCR signals.

Secondly, there are different methods for extracting information from SCR signal that are generally regrouped under two main approaches: the operationalist approach and the model-based approach. Operationalist approach work by extracting a value from a small number of points in the time series to operationalize a psychological construct (Bach et al., 2018; Posada-Quintero & Chon, 2020). For example, the PS is an amplitude measure that is computed using the difference between the maximum and minimum values of the SCR response. Another operationalist approach is the area under the curve (AUC) – essentially the area under the SCR curve. Both are used to measure SCR responses to various stimuli and for example to test if certain stimuli presented to participants produce larger PS or AUC values which can be interpreted as being more stressful than others (related to increased sympathetic arousal). The model-based approach uses a mathematical model to map a psychological construct to the observable signal (Bach et al., 2018). More precisely, it uses two models, the first which is called the neural model, is used to essentially map a psychological construct (e.g. fear) to a sudomotor nerve activity (Bach et al., 2018). The second model called the peripheral linear time invariant model (LTI) is used to map

the sudomotor nerve activity to psychophysiological activity (e.g. SCR) (Bach et al., 2018). In the end, when using the model-based approach (i.e. the GLM or the DCM), one gets a parameter estimate of the neural amplitude that created the observed SCR. So just like PS and the AUC, a larger parameter estimate is related to a larger sympathetic arousal. In their paper, Privratsky et al. (2020) looked at the two different metrics to quantify SCR signals; one operationalist (PS) and one model-based (GLCM). To complement their work, we will add the area under the curve (AUC) as a third metric in our analyses. Indeed, Shukla et al. (2021), using an emotion recognition task, have shown the superiority of the AUC comparatively to PS. Also, it has been theorized in Bach, Friston et al. (2010) that the AUC could be used as an event-related metric, but to our knowledge, it has not been tested with different preprocessing combinations such as those used in Privratsky et al. (2020). Furthermore, to our knowledge, while Privratsky et al. (2020) showed that a model-based metric was better at discriminating between experimental conditions than the PS using different preprocessing combinations, no such work has looked at comparing model-based metrics to the AUC nor comparing the AUC to the PS. Hence, our second main objective is to test the impacts of various preprocessing combinations on the PS, AUC and model-based metrics and to compare each metric's ability to maximize the discrimination between the SCR from two experimental conditions.

Thirdly, Privratsky et al. (2020) have tested different preprocessing combinations using a single dataset. This raises the question: can their results and recommendations be generalized to other SCR datasets? This issue is indeed mentioned by Privratsky et al. (2020). Furthermore, their dataset was collected in a very specific experimental context where SCR was recorded simultaneously with fMRI recordings. This is not perfectly representative of most SCR signal recording environments. Hence, our third main objective is to study if optimal preprocessing approaches are dataset specific. To do this we will test various preprocessing combinations with

different openly accessible datasets that were collected in more usual laboratory settings (i.e. not in a scanner) (Bach et al., 2009; Bach, Daunizeau et al., 2010).

Overall, we hope to stress the importance of preprocessing in SCR research. To do this we will propose data-based suggestions on how to preprocess SCR signals. Moreover, we will provide freely accessible code for SCR signal preprocessing that can be used to build a pipeline for SCR preprocessing and identify the optimal preprocessing combination.

2. Methods

2.1. Datasets

2.1.1. HRA-1

The HRA-1 dataset is the second experiment from Bach, Daunizeau et al. (2010). In this experiment, 20 participants from the general population were recruited (10 females; mean age = 22.2 years, SD = 4 years). The task from HRA-1 is a classic Pavlovian learning task between conditional and unconditional stimuli (i.e. CS and US respectively). CS were orange and blue filled circles that appeared on each trial on either the right or left side of the screen. One of the colours of the CS (CS+) was paired 50% of the time with an aversive US: an electrical shock of 500 Hz for 0.5 ms. The other colour (CS-) was never followed by US. Participants had to indicate with the cursor or the arrow which was the colour of the CS. The stimulus onset asynchrony (SOA) was 3.5 s and the inter-trial interval (ITI) was randomly selected between 7, 9, and 11 s, there were 180 trials, 90 for each CS. In the original article, the signal was preprocessed with a bidirectional first order Butterworth band-pass filter with a cutoff frequency of 5 Hz and 0.0159 Hz and finally down sampled to 10 Hz sampling rate, also a z-score of the continuous SCR time series rescaling was applied (Bach, Daunizeau et al., 2010). Original data analysis showed that both dynamic causal model (DCM), a model-based approach, and PS managed to significantly

discriminate between aversive CS+ and neutral CS-, although DCM produced better discrimination than PS (Bach, Daunizeau et al., 2010).

2.1.2. SCRV-1

The SCRV-1 dataset is the third experiment in Bach et al. (2009). In this experiment, 24 healthy participants were recruited from the general population (12 females; mean age = 27 years, SD = 4.6 years). Participants were shown aversive and neutral pictures drawn from the International Affective Picture System (IAPS) (Lang et al., 2005): aversive stimuli were the 90 most arousing and negative pictures from the IAPS; neutral stimuli were the 90 least arousing neutral pictures from the IAPS. The 90 aversive pictures were separated into three sets with different inter-stimulus intervals (ISI) (i.e. 3, 9 or 19 sec), the same thing was done with the neutral pictures. Each participant saw stimuli randomly presented from these aversive and neutral sets with a random stimulus onset around the mean ISI with a jitters of 0.08, 0.68 and 1.68 sec, respectively. Finally, participants were asked to signal if they liked the stimulus or not by pressing the up or down key. In the original article, the signal was preprocessed with a first order Butterworth band-pass filter with cutoff frequency of 5 Hz and 0.0159 Hz and then downsampled to 10 Hz, z-score to the continuous time series rescaling was also applied (Bach et al., 2009). Original data analysis showed that the general linear model (GLM), a model-based approach managed to discriminate between aversive and neutral stimuli, but PS did not (Bach et al., 2009).

2.1.3. SCRV-4

The SCRV-4 dataset is the first experiment from Bach, Daunizeau et al. (2010). Note that although both datasets come from the same article (Bach, Daunizeau et al., 2010), participants from HRA-1 and SCRV-4 were not the same. In this experiment, 32 participants from the general

population were recruited (16 females; mean age = 22.4 years, SD=4.6 years). The task was essentially the same as the HRA-1 task, except that the aversive US was different–instead of being an electric shock it was a loud white noise at approximately 95 dB. Also, participants had to indicate where on the screen the CS appeared on each trial with the cursor buttons. SOA randomly varied between 4, 10 and 16 sec and the ITI was also randomly selected between 14, 19 and 24 sec. In total there were 64 trials, 32 of each CS. Just like in HRA-1, CS+ was followed 50% of the time by the US (i.e. aversive condition) and CS- was never followed by US (i.e. neutral condition). In the original article, preprocessing steps were the exact same as the one in HRA-1. Original data analysis showed that the DCM could differentiate between aversive and neutral events while PS could not (Bach, Daunizeau et al., 2010).

2.2. Ethics statement

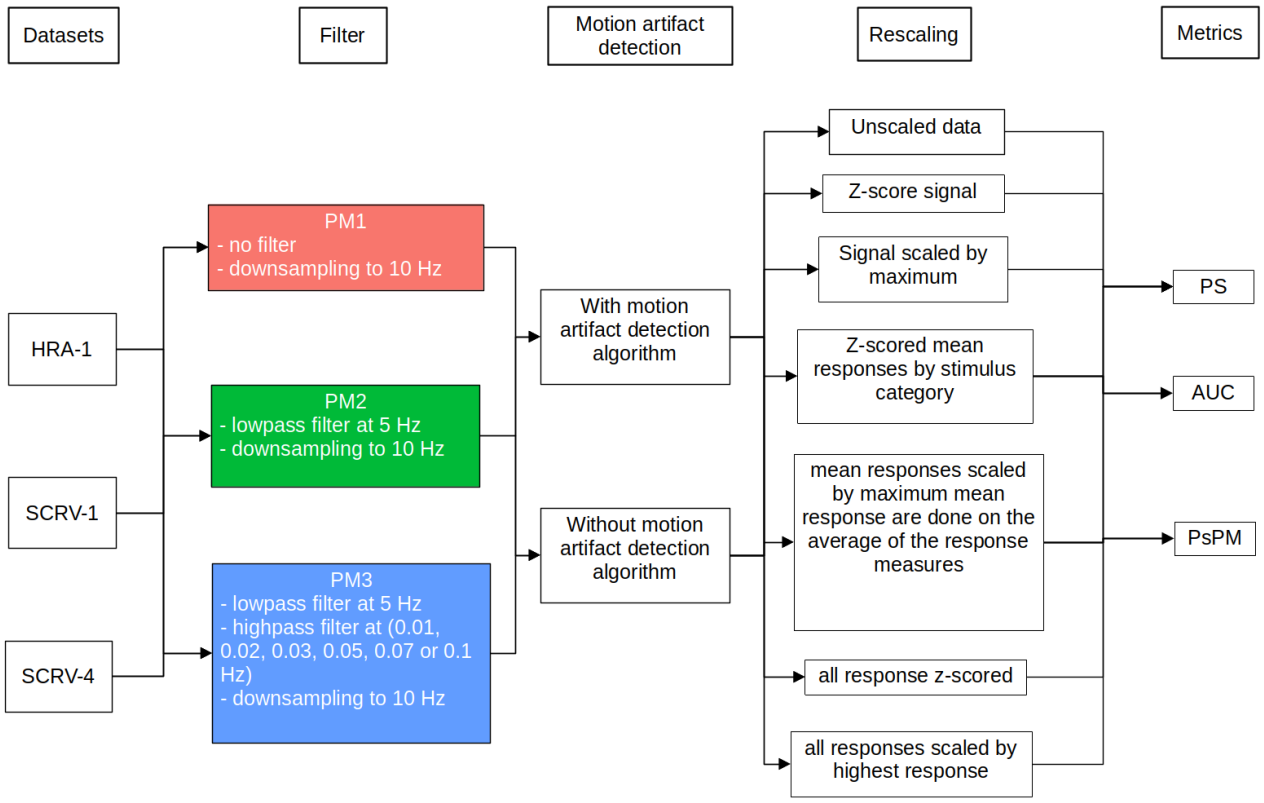
Because we used free and open access online datasets, we were not required to present this project to our IRB.

2.3. Data analysis

Based on the work by Privratsky et al. (2020), we tested different combinations of filtering methods, rescaling methods and MR. In line with our 3 main objectives, we also tested the impacts of adding a motion artifact detection algorithm, tested the combinations on the model-based metrics, the PS and also the AUC and tested these combinations in three different datasets (see Fig. 1). The code for all the data analysis is available here : https://github.com/neurok8050/eda-optimisation-processing-tool.

Figure 1

The flow chart of the SCR preprocessing pipeline used



Note. PM: Processing method. The way we processed was to filter the signal then to remove the epochs containing motion artifact or not (i.e. in the case that we did not use motion artifact detection). Afterwards, we rescaled the signal and finally computed the metrics. We did this pipeline for each of the possible combinations. So each dataset passed through each PM and then each motion artifact detection protocol, each rescaling and finally each metric.

2.3.1. Processing method 1 (PM1) : no filtering

The first filtering method proposed in Privratsky et al. (2020) is essentially not applying any sort of filtering at all. This approach represents the minimal preprocessing they found in their short systematic review of the literature (Privratsky et al., 2020). The only thing done to the signal is a downsampling to 10 Hz which should not impact the scoring of SCR (Fahrenberg et al. (1983) and Lim et al. (1997) cited in Privratsky et al. (2020)).

2.3.2. Processing method 2 (PM2) : low-pass filtering

The second filtering method proposed by Privratsky et al. (2020) is a single filter to remove high-frequency noise. First, they used a median filter with a 10 ms window to remove artifact coming from the fMRI (Staib et al. (2015) cited in Privratsky et al. (2020)). We did not implement a median filtering because all of our dataset did were not recorded during fMRI. This being said, as in Privratsky et al. (2020) we implemented a bidirectional first order Butterworth low-pass filter with a cutoff frequency of 5 Hz after which the signal was downsampled to 10 Hz. For more detail about the implementation in MATLAB consult the code freely and openly accessible on our GitHub page.

2.3.3. Processing method 3 (PM3) : low-pass and high-pass filtering

The third filtering method proposed by Privratsky et al. (2020) is the exact same as PM2 except that before resampling, a high-pass filter was applied with seven different high cutoff frequencies. The addition of the high-pass filter to the low-pass filter makes it a band-pass filter. We implemented a bidirectional first order low-pass Butterworth filter with a low cutoff frequency of 5 Hz, and then a bidirectional first order high-pass Butterworth filter with a high cutoff frequency that varied between 0.01, 0.0159, 0.02, 0.03, 0.05, 0.07 and 0.1 Hz. Finally, the signal was downsampled to 10 Hz.

2.3.4. Subject data rescaling

In order to remove the inter-subject variance, Privratsky et al. (2020) used seven different rescaling methods. Three methods are done on the entire SCR time series: (1) unscaled data: (2) z-scores continuous SCR time series and (3) continuous time series scaled by maximum recorded value. Two methods are done on the means of response measures: (4) z-scored mean responses by stimulus categories (5) mean responses scaled by maximum mean response. Two methods are applied on all responses measures: (6) all responses z-scored and (7) all responses scaled to the highest response. It is worth mentioning that the fourth rescaling method will not be used for

SCRV-1 for PS, AUC and PsPM since SCRV-1 has only two stimulus categories and computing the average, standard deviation or the z-score of two data points is not informative. Ultimately, for the model-based approach for SCRV-1 only the first, second and third rescaling method will be used. For the model-based approach for HRA-1 and SCRV-4 only the first, second, third, fourth and fifth rescaling methods will be used.

2.3.5. Motion artifact detection algorithm

In line with our first objective, we added an extra step consisting of using vs. not using Taylor et al. (2015) algorithms to detect and remove motion artifacts in the signal. To be precise, we do not try to see if the algorithm should be included in the pipeline or not, we think that artifact removal should probably be always done. Instead, we want to assess how including this step can influence the choice of optimal preprocessing approaches. The goal of this algorithm was to encode the expertise on motion artifact detection of human experts in a ML algorithm (Taylor et al., 2015). They achieved this by asking two SCR signal experts to independently identify which of 1560 epochs contained motion artifacts based on prior defined criteria (Taylor et al., 2015). These two experts agreed on the criteria and labelled each of the 1560 epochs independently. After the labelling was done, the two experts compared their labels, if they disagreed on the labelling of one epoch, the epoch was labelled as questionable. After this, the labels were fed to different ML algorithms and a support vector machine (SVM) algorithm was identified as the best algorithm. The algorithm was released online (https://SCR-explorer.media.mit.edu/) and the source code is available on their GitHub page (https://github.com/MITMediaLabAffectiveComputing/SCR-explorer). In our case, the filtered signal was given to the algorithm and we obtained an output file with time markers of 5 s epochs containing artifacts. With the file, we then proceeded to remove any event markers in the SCR files that were in an epoch that contained an artifact. Essentially, by taking out the event markers

in the SCR file, these events were not analyzed. Finally, to be as conservative as we could during the motion artifact detection, we used the binary classifier of the algorithm.

2.3.6. Response measure method 1 : operationalist approach

2.3.6.1. Peak-scoring (PS)

As in Privratsky et al. (2020), the PS was calculated as the difference between the maximum value of the signal in a window ranging from 1 s after the stimulus onset to an MR of either 3.5, 4, 4.5 or 5 s and the minimum value from a baseline (2 s period before the stimulus onset).

2.3.6.2. Area under the curve (AUC)

To calculate the AUC, we took the code from the pspm_sf_auc.m function for PsPM which was elaborated to calculate the AUC based on a paper by Bach, Friston et al. (2010). Essentially, we measured the minimum of each epoch defined by the signal from 1 s after the stimulus onset to an MR of either 3.5, 4, 4.5 or 5 s then we subtracted this minimum from the whole epoch and finally averaged the epoch. For the detailed code about AUC calculation either visit the PsPM function GitHub page (https://github.com/bachlab/PsPM/blob/develop/src/pspm_sf_auc.m) or visit our GitHub page.

2.3.7. Response measure method 2 : model-based approach (PsPM)

For the model-based approach, we used the PsPM software developed by (Bach, Daunizeau et al., 2010; Bach et al., 2009, 2013; Bach, Flandin et al., 2010). The software has different implementations of SCR analysis depending on the experimental design. PsPM implements models that use the same idea as the general convolutional hemodynamic response function (HRF) from fMRI BOLD signal analysis (Privratsky et al., 2020). Privratsky et al. (2020) did not use PsPM has is, they rather used some of the functions from PsPM. On the other hand, we did not exactly apply what Privratsky et al. (2020) did, we instead followed the

recommendation from the PsPM manual (https://github.com/bachlab/PsPM/blob/develop/doc/PsPM_Manual.pdf). For HRA-1 and SCRV-4, because both have a fear conditioning experimental design which leads to none fix SCR: because of the anticipation, the time at which the SCR occurs varies between trials. If the neural burst latency after a CS is constant or assumed to be constant, one can use the GLM approach from PsPM, but here the neural burst latency is not constant or assumed not to be constant. In these cases, one should use the dynamic causal model (DCM) from PsPM and this is the case for HRA-1 and SCRV-4. The DCM was also developed from the article from which the HRA-1 and SCRV-4 datasets were taken (Bach, Daunizeau et al., 2010). We therefore decided to use the DCM for both of these datasets. For SCRV-1, we did not use the DCM, but rather the GLM which is similar to the general linear convolutional model (GLCM) that Privratsky et al. (2020) used[1]. Here, because the experimental design of SCRV-1 does not involve anticipation, the neural bursts are considered to have a fix latency and thus can be easily modelled with the GLM algorithm from PsPM. The GLM algorithm from PsPM was developed from the article containing SCRV-1 (Bach et al., 2009). Similar to operationalist metrics, model-based metrics were computed after the filtering, rescaling and motion artifact detection steps.

2.3.8. Evaluation of preprocessing effects on contrasts of interest

We applied all possible combinations to the PS and AUC. This gave us 9 filtering methods × 7 rescaling methods × 4 response ranges × 2 motion artifact algorithm conditions (i.e. without

1 We had to slightly modify the PsPM code for the GLM (i.e. pspm_glm.m, https://github.com/bachlab/PsPM/blob/develop/src/pspm_glm.m), so the GLM function is not exactly the same as the one from PsPM.

and with the motion artifact detection algorithm) for a total of 504 different combinations (Fig. 1). For PsPM, since some of the rescaling methods could not be used, we had 9 filtering methods × 5 or 3 rescaling methods × 2 motion artifact algorithm conditions for a total of 90 or 54 different combinations. As in Privratsky et al. (2020), we wanted to evaluate if the different combinations had any effect on contrasts of interest—the discrimination between the experimental conditions. In the case of our datasets, we calculated the contrast between the aversive CS (CS+) and the neutral CS (CS-) for each preprocessing combination. Then, the effect size from a Wilcoxon signed-rank test between the contrasts (CS+ vs. CS-) was computed and finally we calculated the Akaike Information Criterion (AIC) computed from the PsPM pspm_predval.m function (https://github.com/bachlab/PsPM/blob/develop/src/backroom/pspm_predval.m) which computes a linear mixed-effect regression for each combination. The AIC is computed from these linear mixed-effect regression model with this formula :

$$AIC = n * log\left(\frac{RSS}{n}\right) + 2k$$

where *n* corresponds to the number of observations, *k* is equal to the number of free model parameters and the *RSS* corresponds to the residual sum of squares (Privratsky et al., 2020). The AIC makes it possible to measure if there is any statistical difference between the performance of two different combinations (Bach et al. (2018) and Bach & Friston (2013) cited in Privratsky et al. (2020)). The logic is if the absolute value of the difference between the AIC of two combinations is superior or equal to 6 (i.e. $|\Delta AIC| \geq 6$) then it means that there is a significant difference (i.e. $p < 0.05$) between them with the combination with lower AIC being the best of the two at discriminating between experimental conditions (Privratsky et al., 2020).

To identify the best combination for each filtering method (PM), we used the following logic separately for preprocessing combinations with and without the use of the motion artifact detection step. First, we looked at all the combination to identify the combination with the highest AIC across all PMs. This gave us a combination of PM, MR and rescaling. We then looked at all the combination but this time looking for the highest effect size. This also gave us a combination of PM, MR and rescaling. If the combination with the highest AIC was the same as the combination with the highest effect size, we kept it and used this combination of PM, MR and rescaling with all PMs. For example, if the highest AIC and effect size combination is found for PM2, MR = 4 s and rescaling 4, to do our comparisons between PM will take the AIC from PM1 with MR = 4 s and rescaling 4 and also do the same for PM3 (i.e. PM3 with MR = 4 s and rescaling 4.)[2]. However, if the highest AIC combination and the highest effect size combination are different, we first tested the difference between the respective AIC of both the combination with the highest AIC and the AIC of the combination with the highest effect size. If there was no significant difference between the two, we keep the combination parameters used by the combination with the highest effect size. On the other hand, if there was a significant difference, we then took the combination parameters used by the best model between the two whether it is the combination with the highest AIC or the combination with the highest effect size. We then kept the combination and used the same combination for the other PM. In other words, the combination which yields the best effect size and/or the best AIC is the best combination overall.

2 When the best combination was either PM1 or PM2, it does not tell us which high cutoff frequency to use because none of them use high cutoff frequency. In that case, we used the combination parameters given by either the PM1 or PM2 and used it on PM3 to identify the high cutoff frequency with the highest AIC. Again, if the combination with the highest AIC had a different high cutoff frequency than the combination with the highest effect size, we compared both combination AIC. If there was no significant difference, we kept the combination parameters corresponding to the combination with the highest effect size. If there was a significant difference, we kept the combination parameters from the best model.

3. Results

Results will follow the same order has our objectives. For each dataset, we will first report the output of the motion artifact detection algorithm so readers can have a rough estimate of the difference between the “raw” and “clean” signal that will be fed to the rest of the preprocessing pipeline and analyzed. Next, in line with our first objective, for each metric (e.g. PS), we will present the results relative to each preprocessing combination (i.e. relative to the different filtering and rescaling methods) first without the motion artifact detection algorithm and then when using the motion detection algorithm. This will enable us to compare the impacts of using the algorithm qualitatively (i.e. is the best preprocessing combination the same with and without the motion artifact detection algorithm) and quantitatively (i.e. does the best preprocessing combination with the motion artifact detection algorithm produces better discrimination than the best without it). In line with our second objective, at the end of each dataset, we will then compare the best combination between operationalist approaches (PS vs. AUC) and the best combinations between operationalist and model-based approaches (the best operationalist metric vs. model-based metric) qualitatively (i.e. is the best filtering method the same between different metrics) and quantitatively (i.e. does the best combination for one metric produces better discrimination than the best for another metric). Finally, after having presented all three datasets, we will be able to qualitatively compare the best combinations between datasets and thus give some insights about if optimal preprocessing combinations are dataset specific. All the additional information concerning the effect of filtering on the mean PS, AUC and mean parameter estimates for aversive and neutral events, the discriminate contrast between aversive and neutral events (i.e. the difference between the mean aversive PS and the mean neutral PS, we expect positive values due to the fact that mean aversive PS should be higher than the mean neutral PS)

and information concerning the MR are in the supplementary material for the PS, AUC, GLM or DCM.

3.1 HRA-1

3.1.1. Motion artifact detection algorithm output

In the case of HRA-1, the algorithm identified 200 epochs (67 epochs with a CS+ event and 13 epochs with a CS- event) containing at least one motion artifact out of 3600 event containing epochs (20 participants multiplied by 180 trials per participants). The mean percentage of epochs removed across participants was 8.02% with a standard deviation of 16.49%. The most epochs identified within a single participant was 25 CS+ and 52 CS-, both occurred in the same participant. Note that the experimental protocol of this experience had the participant's arm restrained to minimize the number of motion artifacts (Bach et al. (2009) cited in Bach, Daunizeau et al. (2010)).

3.1.1.1. PS

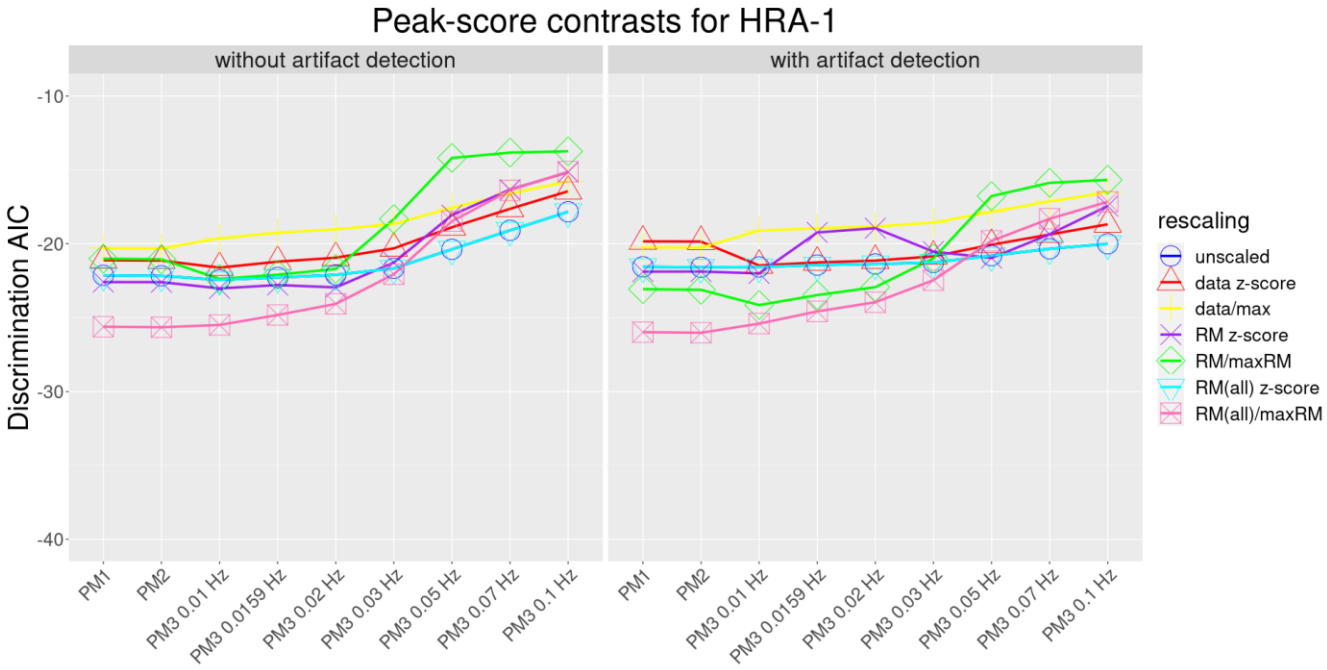
For both without and with motion artifact detection, the AIC for all combinations depending on rescaling method are shown in the Fig. 2, the effect sizes depending on MR time are shown in Fig. 3 and the best combinations using each of the three filtering methods are shown in table 1. There was no significant difference between the three filtering methods' best combination without motion artifact detection (PM1 vs. PM2 $\Delta|AIC| = 0.02$, PM2 vs. PM3 $\Delta|AIC| = 0.48$ and PM1 vs. PM3 $\Delta|AIC| = 0.50$; see table 1). There was also no significant difference between filtering methods' best combination with motion artifact detection (PM1 vs. PM2 $\Delta|AIC| = 0.05$, PM2 vs. PM3 $\Delta|AIC| = 0.62$ and PM1 vs. PM3 $\Delta|AIC| = 0.57$; see table 1).

Since there was no statistically better combination when considering the AIC, we decided to qualitatively consider the effect size to identify the best preprocessing combination without and with motion artifact detection. Using this approach, the low-pass filtering method (PM2)

seemed a slightly better filtering method for both without and with motion artifact detection (see Fig. 3) but the preprocessing combinations used different rescaling methods (i.e. rsm 2 without motion artifact detection and rsm 7 with motion artifact detection; see table1) suggesting that the best preprocessing combination could vary slightly because of the motion artifact detection step. When comparing the best combination for PS without (AIC = -21.13) versus with motion artifact detection (AIC = -26.02), there was no significant difference (Δ|AIC| = 4.89).

Figure 2

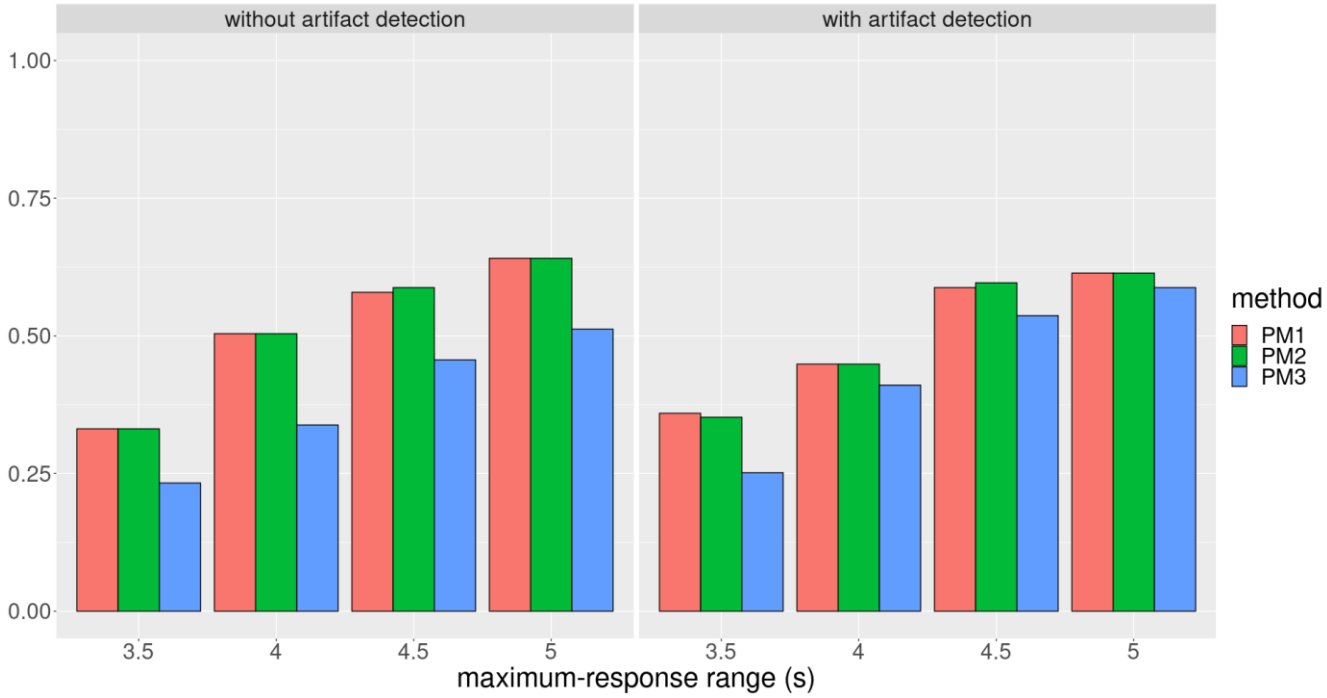
AIC from PS contrast for HRA-1 without and with motion artifact detection



Note. AIC of the linear mixed-effect regression model corresponding to the filter methods and the rescaling. The MR = 5 s and was the same for all the AICs for both without and with motion artifact detection. A Δ|AIC| ≥ 6 meant that the model with the lowest AIC was significantly better than the other one. Note that the lowest AIC in the graph might not be the one used for the comparison if it did not have the best effect size and was not significantly better than the combination with the highest effect size.

Figure 3

Effect size of PS for HRA-1 without and with motion artifact detection



Note. Effect sizes calculated from a Wilcoxon signed-rank test. The effect size of PS increased linearly with the MR without and with motion artifact detection. PM1 and PM2 were extremely similar, but PM2 had in some cases a slightly higher effect size. Overall, the effect size of PS used without motion artifact detection was slightly higher than the effect size of the PS with motion artifact detection across PM and MR.

Table 1

Best combination for each filtering method for PS for HRA-1 without and with motion artifact detection

	PM1	PM2	PM3
without motion artifact detection	rescaling : 2 MR = 5 seconds AIC = -21.11	**rescaling : 2** **MR = 5 seconds** **AIC = -21.13**	rescaling : 2 high cutoff frequency : 0.01 Hz MR = 5 seconds AIC = -21.61

with motion artifact detection	rescaling : 7 MR = 5 seconds AIC = -25.97	**rescaling : 7** **MR = 5 seconds** **AIC = -26.02**	rescaling : 7 high cutoff frequency : 0.01 Hz MR = 5 seconds AIC = -25.40

Note. The best combinations across filtering methods (PM) without and with the motion artifact detection are in bold.

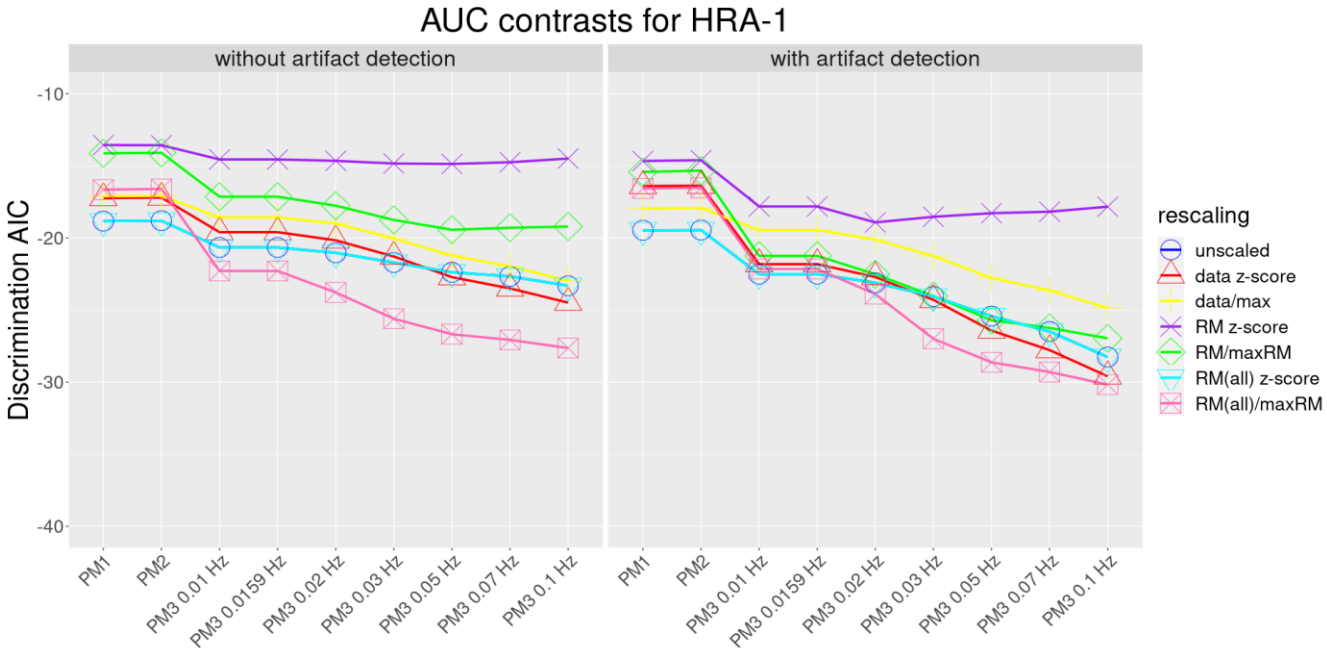
3.1.1.2. AUC

For both without and with motion artifact detection, the AIC for all combinations depending on rescaling method are shown in the Fig. 4, the effect sizes depending on MR time are shown in Fig. 5 and the best combinations using each of the three filtering methods are shown in table 2. Without motion artifact detection, PM3 was significantly better than PM1 and PM2, but PM1 and PM2 were not different (PM3 vs. PM2 Δ|AIC| = 10.08, PM3 vs. PM1 Δ|AIC| = 10.04, PM1 vs. PM2 Δ|AIC| = 0.04; see table 2). With motion artifact detection, again PM3 was significantly better than PM1 and PM2, but again PM1 and PM2 were not different (PM3 vs. PM2 Δ|AIC| = 13.65 and PM3 vs. PM1 Δ|AIC| = 13.60, PM1 vs. PM2 Δ|AIC| = 0.05; see table 2).

Looking at table 2, the best combination without and with motion artifact detection used the same band-pass filtering method (PM3) and the same rescaling (rsm 7) but they had different MR (5 seconds and 4.5 seconds respectively) and also different high cutoff frequencies (0.05 and 0.1 Hz respectively) suggesting again that the best preprocessing combination could vary slightly because of a motion artifact detection step. This being said, there was no significant difference between the best combination without (AIC = -26.68) and with the motion artifact detection step (AIC = -30.17) (Δ|AIC| = 3.49) when using the AUC.

Figure 4

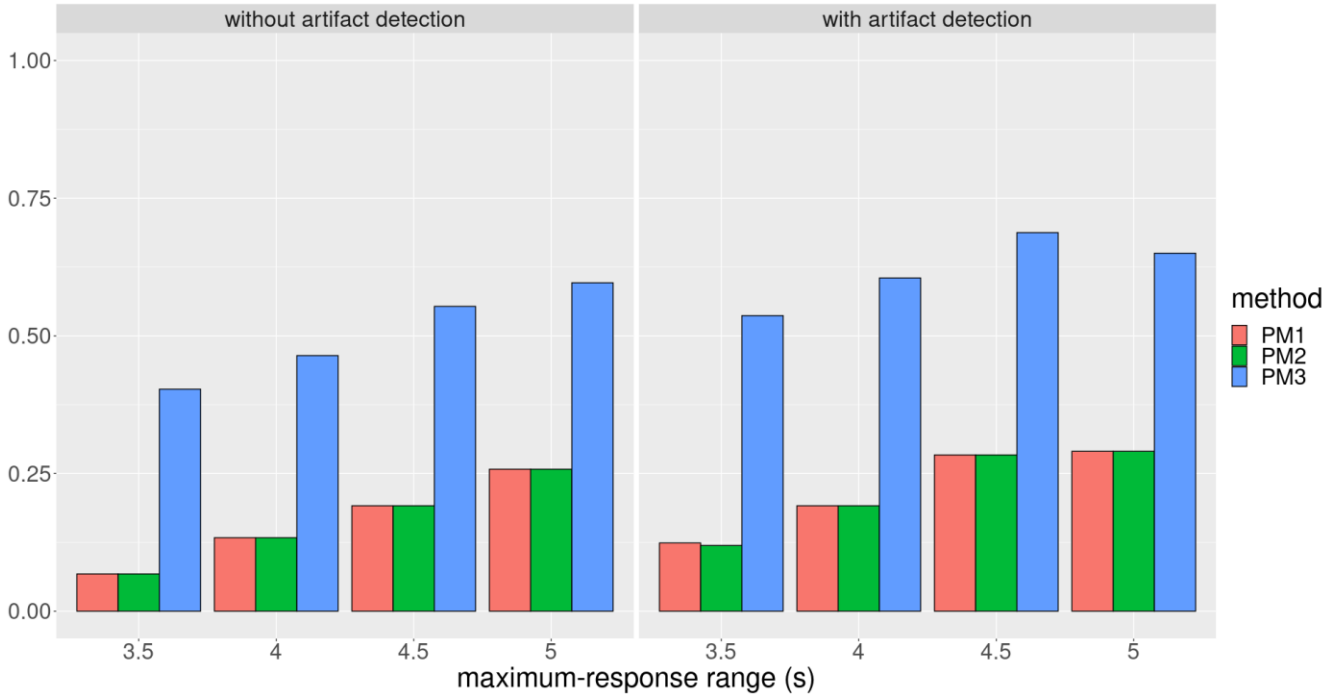
AUC contrasts for HRA-1 without and with motion artifact detection



Note. The AIC of the linear mixed-effect regression model corresponding to the filter methods and the rescaling. The MR = 5 s for without motion artifact detection and MR = 4.5 s with motion artifact detection. A $\Delta|AIC| \geq 6$ meant that the model with the lowest AIC was significantly better than the other one. Note that the lowest AIC in the graph might not be the one used for the comparison if it did not have the best effect size and was not significantly better than the combination with the highest effect size.

Figure 5

Effect size of AUC for HRA-1 without and with the motion artifact detection



Note. Effect sizes calculated from Wilcoxon signed-rank test. The effect size of the AUC increased mostly linearly with the MR without and with motion artifact detection. PM3 clearly has a better effect size across all the MR than PM1 and PM2. PM1 and PM2 were similar except at MR = 3.5 s with motion artifact detection. Overall the effect size of the AUC used with motion artifact detection was higher than those without motion artifact detection for all PMs and MRs.

Table 2

Parameters filter of the best filters for AUC for HRA-1 without and with motion artifact detection

	PM1	PM2	PM3
without motion artifact detection	rescaling : 7 MR = 5 seconds AIC = -16.64	rescaling : 7 MR = 5 seconds AIC = -16.60	**rescaling : 7** **high cutoff frequency : 0.05 Hz** **MR = 5 seconds** **AIC = -26.68**
with motion artifact	rescaling : 7 MR = 4.5 seconds	rescaling : 7 MR = 4.5 seconds	**rescaling : 7** **high cutoff frequency : 0.1 Hz**

detection	AIC = -16.57	AIC = -16.52	**MR = 4.5 seconds** **AIC = -30.17**

Note. The best combinations across filtering methods (PM) without and with the motion artifact detection are in bold.

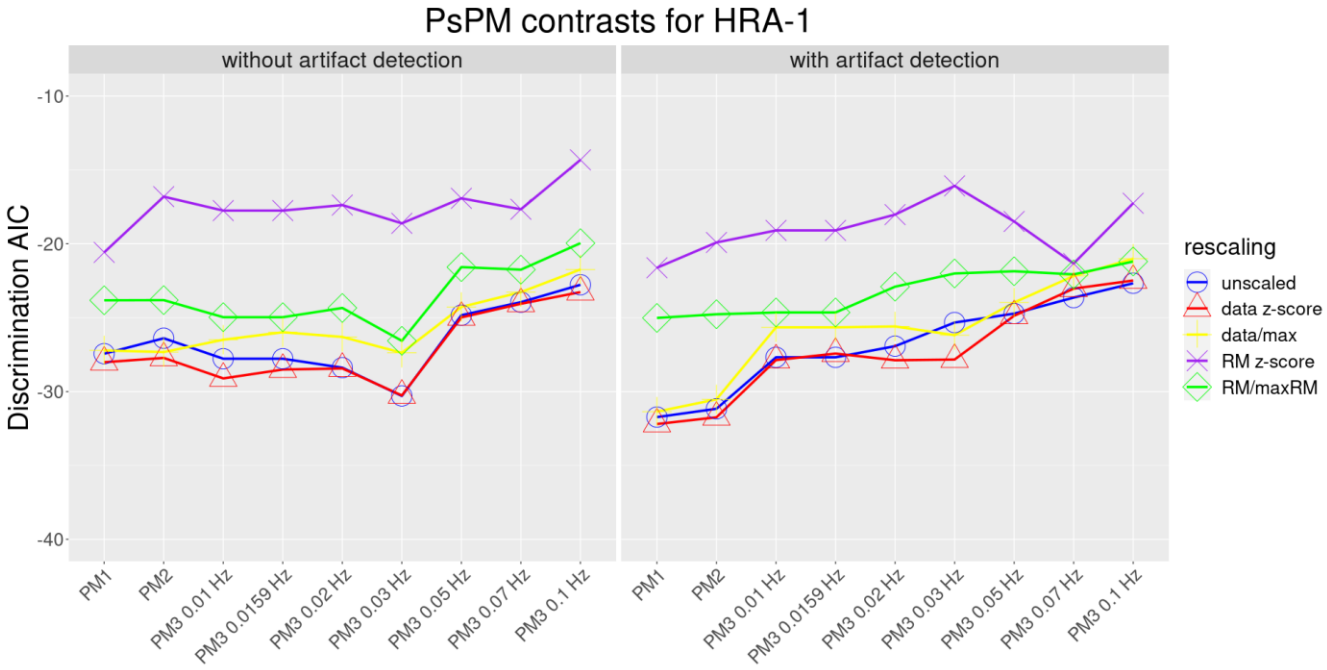
3.1.1.3. Model-based approach

For both without and with motion artifact detection, the AIC for all combinations depending on rescaling method are shown in the Fig. 6, the effect sizes depending on MR time are shown in Fig. 7 and the best combinations using each of the three filtering methods are shown in table 3. There was no significant difference between the filtering methods' best combination without motion artifact detection (PM1 vs. PM2 Δ|AIC| = 0.29, PM2 vs. PM3 Δ|AIC| = 1.39 and PM1 vs. PM3 Δ|AIC| = 1.1; see table 3). There was also no significant difference between the different filtering methods' best combination with motion artifact detection (PM1 vs. PM2 Δ|AIC| = 0.45, PM2 vs. PM3 Δ|AIC| = 3.87 and PM1 vs. PM3 Δ|AIC| = 4.32 see table 3).

Since there was no statistically better combination when considering the AIC, we decided to qualitatively consider the effect size to identify the best preprocessing combination without and with motion artifact detection. Using this approach, the no filtering method (PM1) seemed a slightly better filtering method without motion artifact detection while the low-pass filtering method seemed slightly better when using motion artifact detection (see Fig. 7). However both preprocessing combinations used the same rescaling method (rsm 2; see table3) suggesting that the best preprocessing combination could vary slightly because of a motion artifact detection step. There was no significant difference between the best combination without (AIC = -28.01) and with the motion artifact detection step (AIC = -31.74) (Δ|AIC| = 3.73) when using the model-based metric.

Figure 6

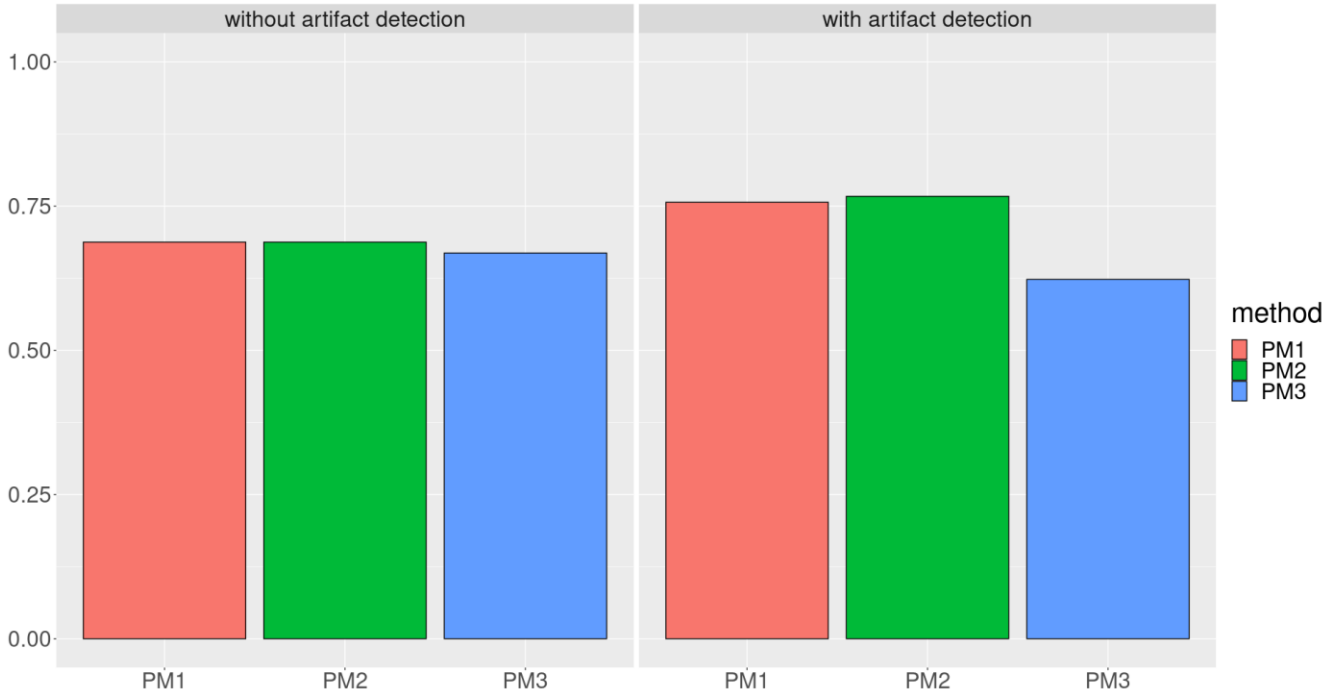
PsPM contrast for HRA-1 without and with motion artifact detection



Note. AIC of the linear mixed-effect regression model corresponding to the filter methods and the rescaling. There was no MR here because the MR was not used by the DCM. A $\Delta|AIC| \geq 6$ meant that the model with the lowest AIC was significantly better than the other one. Note that the lowest AIC in the graph might not be the one used for the comparison if it did not have the best effect size and was not significantly better than the combination with the highest effect size.

Figure 7

Effect size of PsPM for HRA-1 without and with motion artifact detection



Note. Effect size calculated from a Wilcoxon signed-rank test. PM1 and PM2 had similar effect size without motion artifact detection, but PM2 has a slightly better effect size when used with motion artifact detection.

Table 3

Parameters of best filters for PsPM for HRA-1 without and with motion artifact detection

	PM1	PM2	PM3
without motion artifact detection	**rescaling : 2** **AIC = -28.01**	rescaling : 2 AIC = -27.72	rescaling : 2 high cutoff frequency : 0.01 Hz AIC = -29.11
with motion artifact detection	rescaling : 2 AIC = -32.19	**rescaling : 2** **AIC = -31.74**	rescaling : 2 high cutoff frequency : 0.01 Hz AIC = -27.87

Note. The best combinations across filtering methods (PM) without and with the motion artifact detection are in bold.

3.1.2. Best approach with HRA-1 dataset

3.1.2.1 PS vs. AUC

Now, if we compared the best combination for PS (PM2, rsm 7, MR 5 seconds and with motion artifact detection) to the best for the AUC (PM3, high cutoff 0.1 Hz, rsm 7, MR 4.5 seconds and with motion artifact detection), qualitatively, they used different filtering methods and MR. Quantitatively, although the effect size of the AUC (AIC = -30.17) was higher than the PS (AIC = -26.02), there was no significant difference between the two ($\Delta|AIC| = 4.15$).

3.1.2.2 Operationalist vs. model-based approach

If we compared the best combination from the operationalist approach (i.e. the AUC with PM3, high cutoff 0.1 Hz, rsm 7, MR 4.5 seconds and with motion artifact detection) with the best combination from the model-based approach (i.e. PM2, rsm 2, with motion artifact detection), qualitatively, they used different filtering methods and rescaling. Quantitatively, the model-based approach had a better effect size than the AUC but there was no significant difference between the model-based approach (AIC = -31.74) and the AUC (AIC = -30.17) ($\Delta|AIC| = 1.57$).

3.2. SCRV-1

3.2.1. Motion artifact detection algorithm

For the SCRV-1 dataset, motion artifact detection identified 83 epochs (38 epochs with a CS+ event and 45 epochs with a CS- event) with at least one motion artifact out of a total of 3960 epochs (22 participants multiplied by 180 epochs per participants). The mean percentage of epochs removed across participants was 2.06% with a standard deviation of 4.64%. The most epoch identified for the CS+ was 19 and 21 for CS-, again both were from the same participant. Again, the arm of the participants was restrained to minimize the number of potential motion artifacts (Bach et al., 2009).

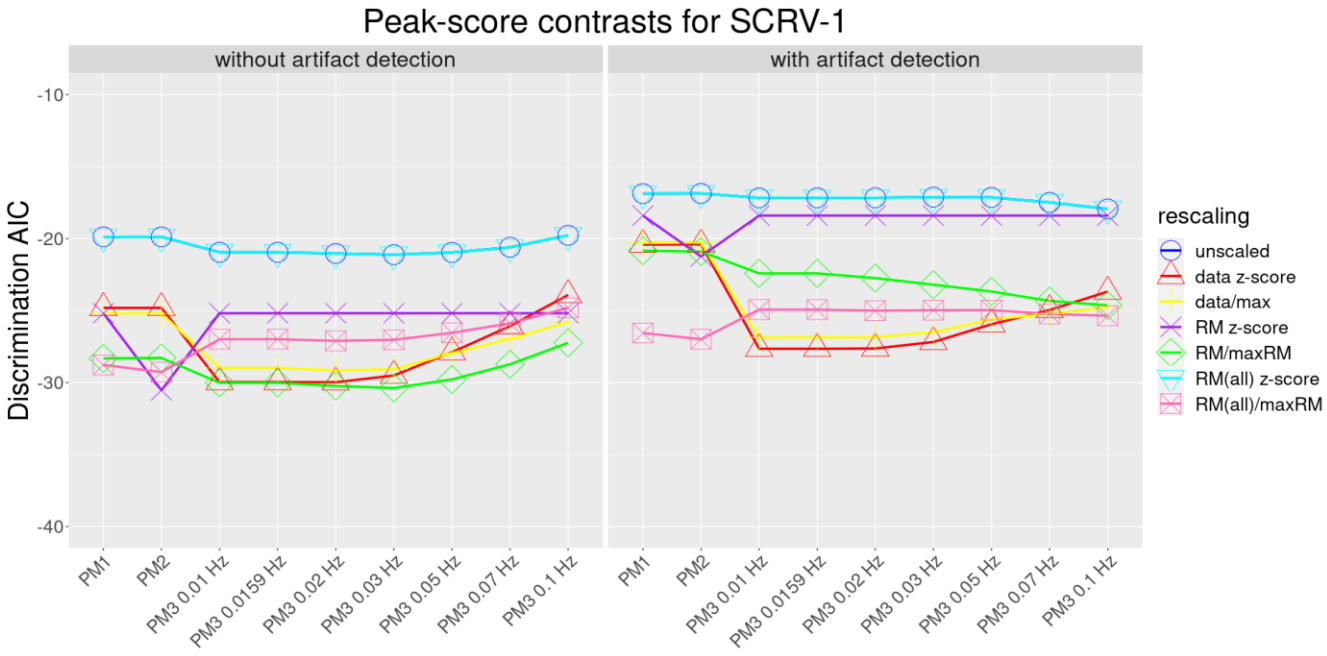
3.2.1.1. PS

For both without and with motion artifact detection, the AIC of all combinations depending on rescaling method are shown in Fig. 8, the effect sizes depending on MR time are shown in Fig. 9 and the best combinations using each of the three filtering methods are shown in table 4. There was no significant difference between the different PM without motion artifact detection (PM1 vs. PM2 $\Delta|AIC| = 0.51$, PM2 vs. PM3 $\Delta|AIC| = 2.18$ and PM1 vs. PM3 $\Delta|AIC| = 1.67$). With motion artifact detection only PM3 was significantly better than PM1 and PM2, whereas PM1 and PM2 were not significantly different (PM1 vs. PM2 $\Delta|AIC| = 0.01$, PM2 vs. PM3 $\Delta|AIC| = 7.25$ and PM1 vs. PM3 $\Delta|AIC| = 7.24$).

Comparing the best combination without motion artifact detection (i.e. PM2, rsm 7, MR 5 seconds) with the best combination with motion artifact detection (i.e. PM3, high cutoff frequency 0.01 Hz, rsm 2, MR 4.5 seconds) and also looking at the effect sizes of the different PM, we came to the same conclusion for the best combination without motion artifact detection (AIC = -29.28) and with motion artifact detection (AIC = -27.65). The difference between both AICs showed no significant difference ($\Delta|AIC| = 1.63$). Therefore, the algorithm did seem to influence the optimal preprocessing approaches.

Figure 8

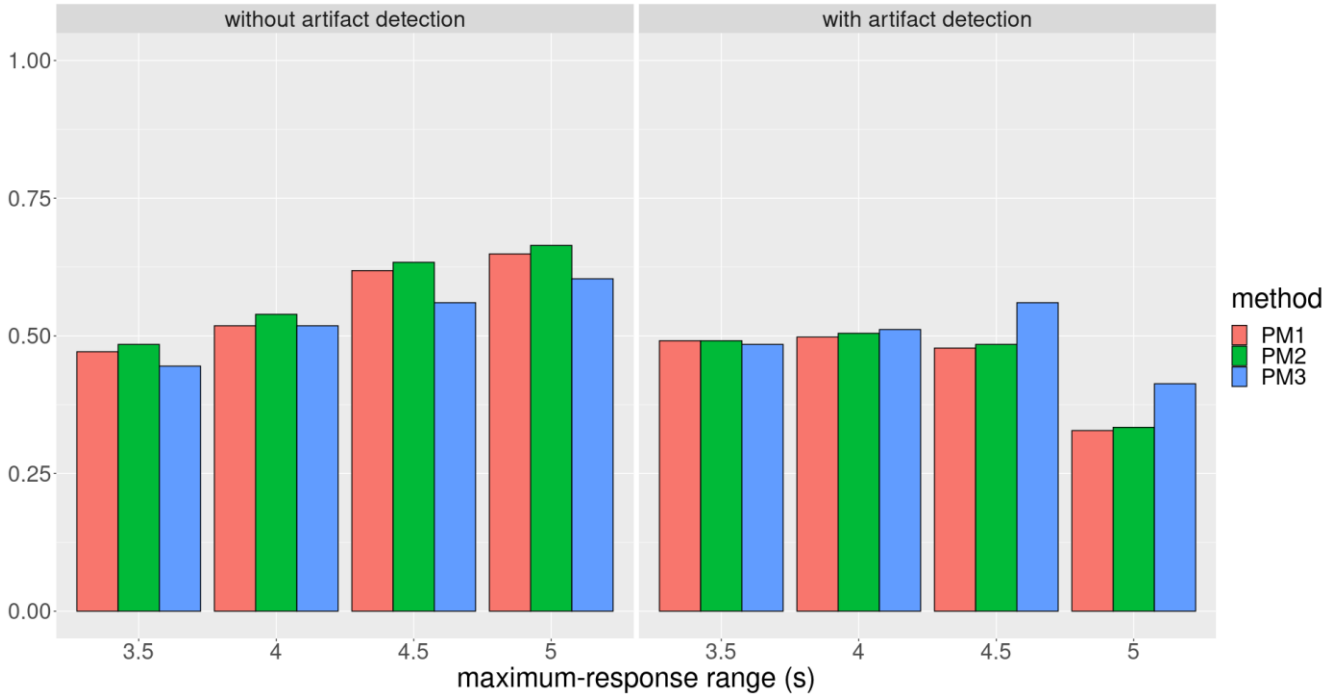
PS contrasts for SCRV-1 without and with motion artifact detection



Note. AIC of the linear mixed-effect regression model corresponding to the filter methods and the rescaling. Without motion artifact detection the MR = 5 s and with motion artifact detection the MR = 4.5 s. A $\Delta|AIC| \geq 6$ meant that the model with the lowest AIC was significantly better than the other one. Note that the lowest AIC in the graph might not be the one used for the comparison if it did not have the best effect size and was not significantly better than the combination with the highest effect size.

Figure 9

Effect size of PS for SCRV-1 dataset without and with motion artifact detection



Note. Effect sizes calculated from a Wilcoxon signed-rank test. Without motion artifact detection, the effect size of PS increased linearly with the MR. With motion artifact detection, it also linearly increased with the MR, but reach maximum at different MR depending on the PM. Overall without motion artifact detection the PS had better effect size than with motion artifact detection.

Table 4

Parameter of the best filters for PS for SCRV-1 without and with motion artifact detection

	PM1	PM2	PM3
without motion artifact detection	rescaling : 7 MR = 5 seconds AIC = -28.77	**rescaling : 7** **MR = 5 seconds** **AIC = -29.28**	rescaling : 7 high cutoff frequency : 0.02 Hz MR = 5 seconds AIC = -27.10
with motion artifact detection	rescaling : 2 MR = 4.5 seconds AIC = -20.41	rescaling : 2 MR = 4.5 seconds AIC = -20.40	**rescaling : 2** **high cutoff frequency : 0.01 Hz** **MR = 4.5 seconds**

			AIC = -27.65

Note. The best combinations across filtering methods (PM) without and with the motion artifact detection are in bold.

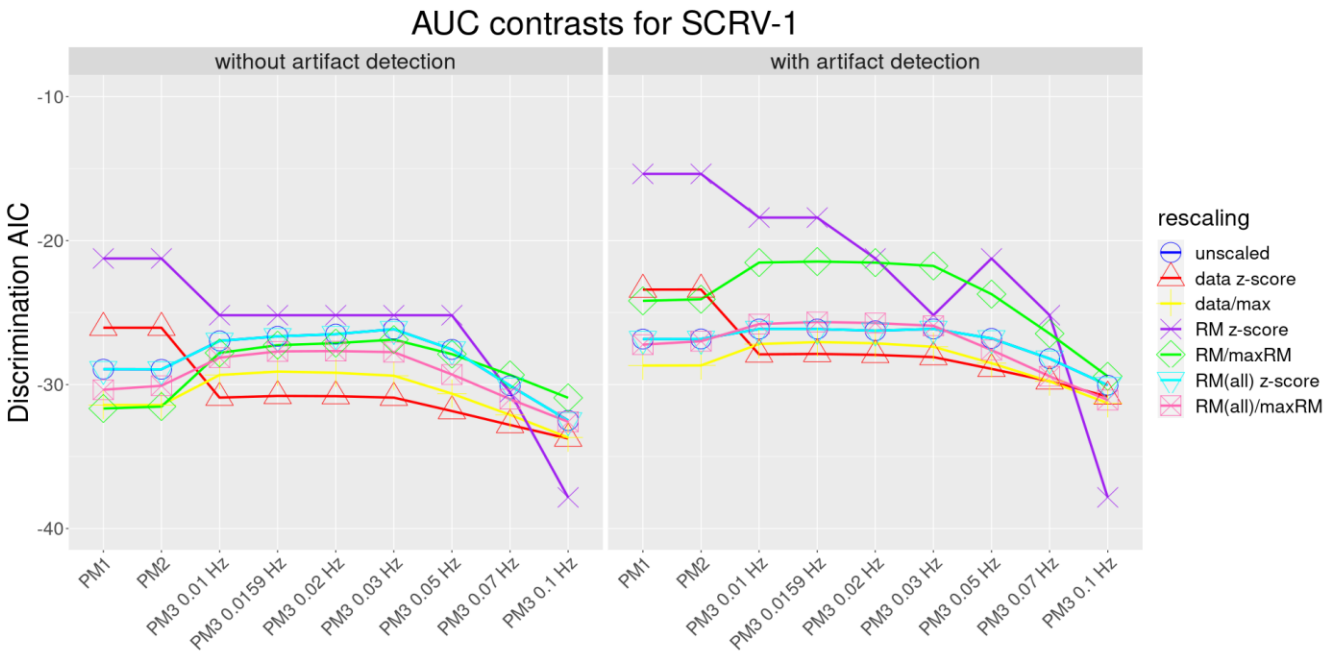
3.2.1.2. AUC

The AIC for all the AUC combinations are shown in Fig. 10 and the bests are resumed in table 5. The effect sizes of the best combinations are shown in Fig. 11. There was no significant difference between the different PM without motion artifact detection (PM1 vs. PM2 Δ|AIC| = 0.01, PM2 vs. PM3 Δ|AIC| = 3.55 and PM1 vs. PM3 Δ|AIC| = 3.56). With motion artifact detection the results were the same, no significant difference between the different PM (PM1 vs. PM2 Δ|AIC| = 0.01, PM2 vs. PM3 Δ|AIC| = 2.62 and PM1 vs. PM3 Δ|AIC| = 2.61).

The best combination without (i.e. PM3, high cutoff frequency 0.1 Hz, rsm 6 and MR 4 s) and with motion artifact detection (i.e. PM3, high cutoff frequency 0.1 Hz, rsm 3 and MR 4 s), looking at the effect size, we also concluded that these were the best combination. Although, the best combination without motion artifact detection (AIC = -32.49) and the best with motion artifact detection (AIC = -31.29) did not differ significantly from each other (Δ|AIC| = 1.2). Again, the algorithm did not seem to influence the optimal preprocessing approaches.

Figure 10

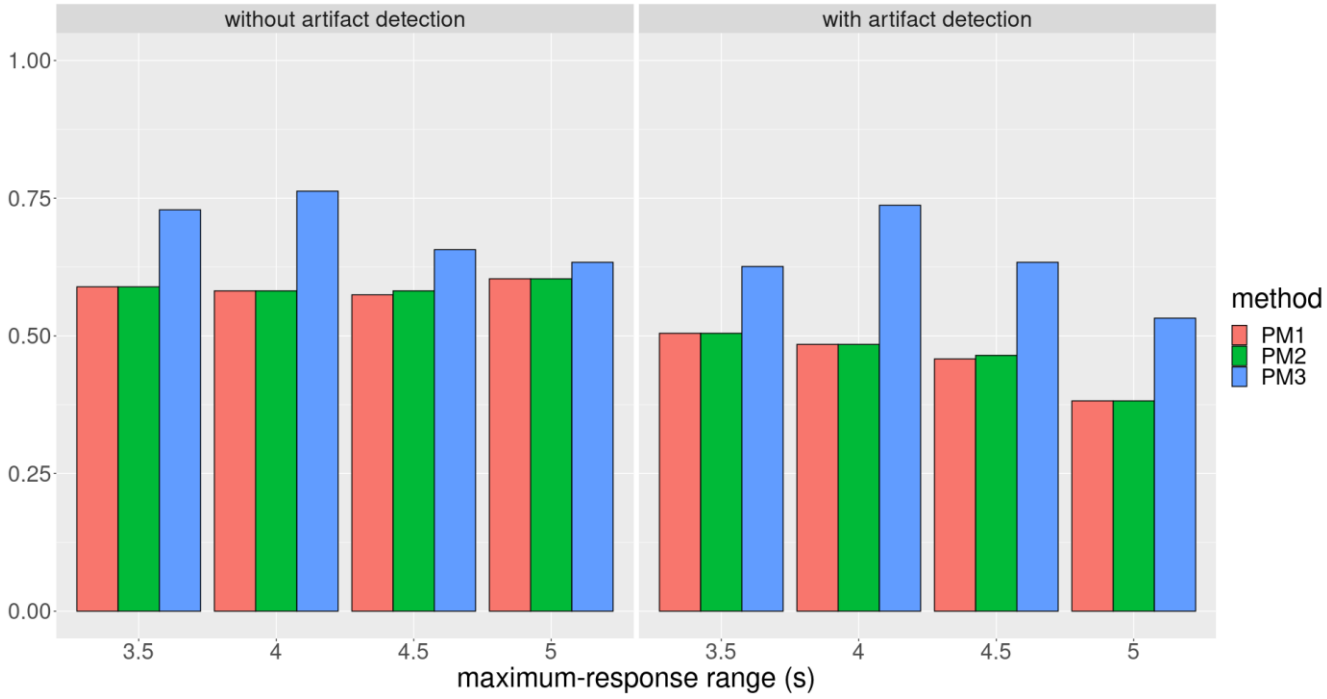
AUC contrasts for SCRV-1 without and with motion artifact detection



Note. AIC of the linear mixed-effect regression model corresponding to the filter methods and the rescaling. The MR = 4 s and was the same for all the AICs for both without and with motion artifact detection. A $\Delta|AIC| \geq 6$ meant that the model with the lowest AIC was significantly better than the other one. Note that the lowest AIC in the graph might not be the one used for the comparison if it did not have the best effect size and was not significantly better than the combination with the highest effect size.

Figure 11

Effect size of AUC for SCRV-1 without and with motion artifact detection



Note. Effect sizes calculated from a Wilcoxon signed-rank test. The effect size increased linearly with MR until MR = 4 s and then decreased for both without and with motion artifact detection. Overall, the effect size for all PMs and across all MRs was better without motion artifact detection, but not by much.

Table 5

Parameter of the best filters for AUC for SCRV-1 without and with motion artifact detection

	PM1	PM2	PM3
without motion artifact detection	rescaling : 6 MR = 4 seconds AIC = -28.93	rescaling : 6 MR = 4 seconds AIC = -28.94	**rescaling : 6** **high cutoff frequency : 0.1 Hz** **MR = 4 seconds** **AIC = -32.49**
with motion artifact detection	rescaling : 3 MR = 4 seconds AIC = -28.68	rescaling : 3 MR = 4 seconds AIC = -28.67	**rescaling : 3** **high cutoff frequency : 0.1 Hz** **MR = 4 seconds** **AIC = -31.29**

Note. The best combinations across filtering methods (PM) without and with the motion artifact detection are in bold.

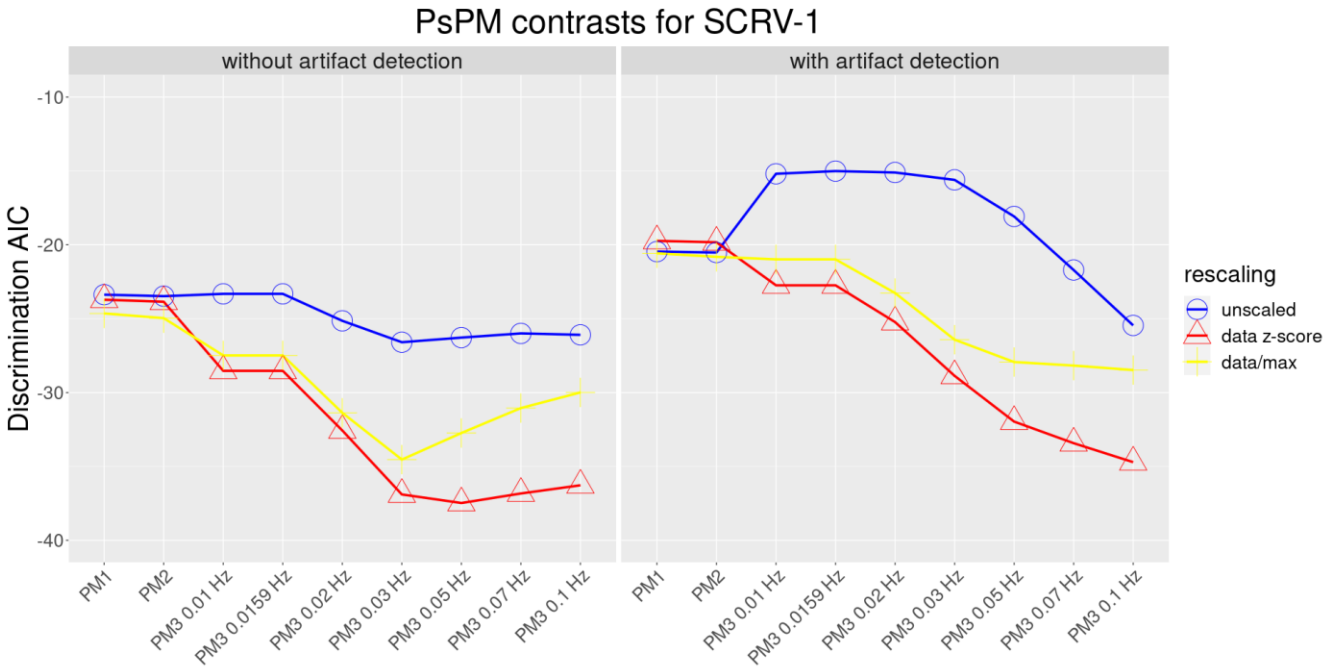
3.2.1.3. Model-based approach

All the different AIC combinations are shown in Fig. 12 and the best combinations are resumed in table 6. The effect sizes of the best combinations are shown in Fig. 13. Without motion artifact detection, PM3 was significantly better than PM1 and PM2, which both were not significantly different from one another (PM1 vs. PM2 Δ|AIC| = 0.32, PM2 vs. PM3 Δ|AIC| = 9.57 and PM1 vs. PM3 Δ|AIC| = 9.89). With motion artifact detection, again PM3 was significantly different than PM1 and PM2, but PM1 and PM2 were not from each other (PM1 vs. PM2 Δ|AIC| = 0.1, PM2 vs. PM3 Δ|AIC| = 14.88 and PM1 vs. PM3 Δ|AIC| = 14.98).

The best combination with and without motion artifact detection used the same PM (i.e. PM3), but with different high cutoff frequency (0.03 and 0.1 Hz, respectively) and rescaling (2 and 3 respectively). Looking at the effect size, we also concluded that these were the best combination. Although, the difference between the best combination without (AIC = -34.53) and the best with motion artifact detection (AIC = 34.71) was not significant (Δ|AIC| = 0.18). The impact of the algorithm did not seem to not influence the optimal preprocessing approaches.

Figure 12

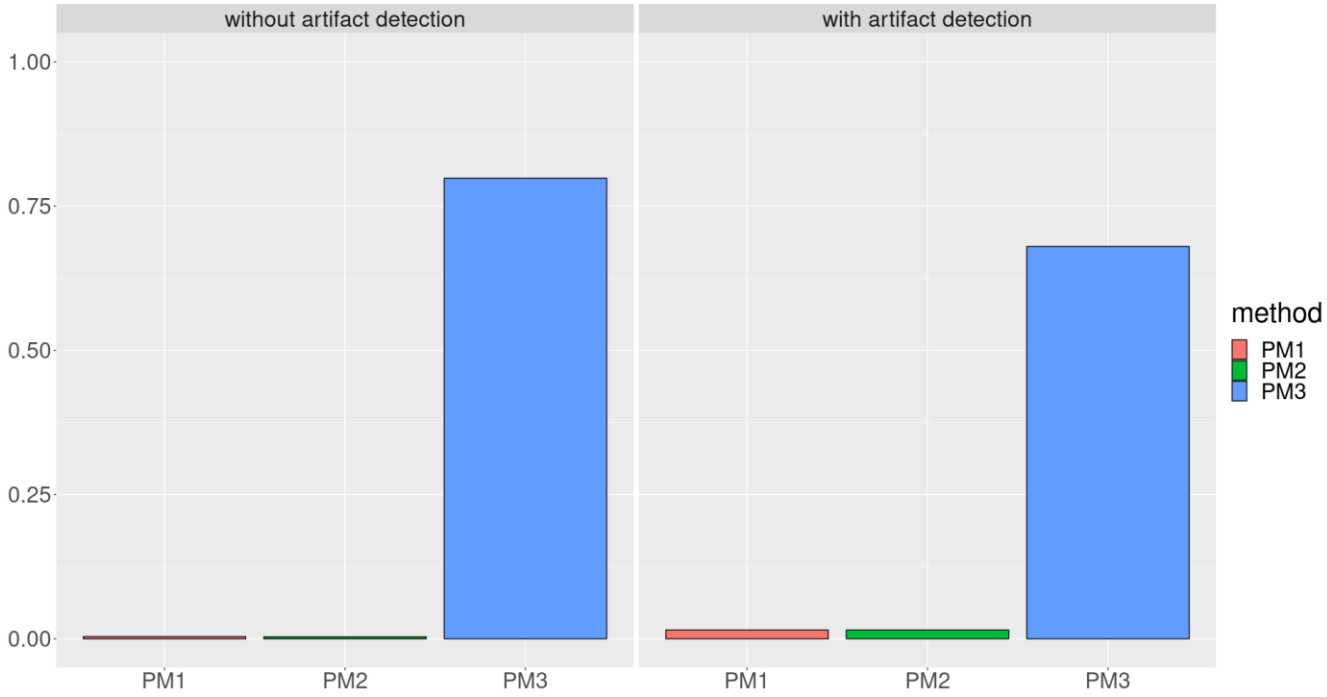
PsPM contrasts for SCRV-1 without and with motion artifact detection



Note. AIC of the linear mixed-effect regression model corresponding to the filter methods and the rescaling. The GLM did not use MR so there was no MR. A $\Delta|AIC| \geq 6$ meant that the model with the lowest AIC was significantly better than the other one. Note that the lowest AIC in the graph might not be the one used for the comparison if it did not have the best effect size and was not significantly better than the combination with the highest effect size.

Figure 13

Effect size of PsPM for SCRV-1 without and with motion artifact detection



Note. Effect sizes calculated from a Wilcoxon signed-rank test. PM1 and PM2 had similar effect size within the same motion artifact detection category. The effect size for PM3 was higher without motion artifact detection than with motion artifact detection. Overall, PM3 had hugely higher effect size than PM1 and PM2.

Table 6

Parameter of the best filters for PsPM for SCRV-1 without and with motion artifact detection

	PM1	PM2	PM3
without motion artifact detection	rescaling : 3 AIC = -24.64	rescaling : 3 AIC = -24.96	**rescaling : 3** **high cutoff frequency : 0.03 Hz** **AIC = -34.53**
with motion artifact detection	rescaling : 2 AIC = -19.73	rescaling : 2 AIC = -19.83	**rescaling : 2** **high cutoff frequency : 0.1 Hz** **AIC = -34.71**

Note. The best combinations across filtering methods (PM) without and with the motion artifact detection are in bold.

3.2.2. Best approach with SCRV-1 dataset

3.2.2.1 PS vs. AUC

Comparing the best combination for the PS (PM2, rsm 7, MR = 5 seconds and without motion artifact detection) to the best for the AUC (PM3, high cutoff 0.1 Hz, rsm 6, MR = 4 seconds, without motion artifact detection), qualitatively, they used different filtering methods, rescaling and MR. Quantitatively, there was no significant difference between the PS (AIC = -29.28) and the AUC (AIC = -32.49) ($\Delta|AIC| = 3.21$). Looking at their respective effect size, the AUC had a better effect size than the PS. Again, it seemed that the AUC had a slight advantage over the PS.

3.2.2.2 Operationalist vs. model-based approach

Comparing the operationalist combination (AUC, PM3, high cutoff of 0.1 Hz, rsm 6, MR = 4 s, without motion artifact detection) to the best model-based approach combination (PM3, high cutoff 0.03 Hz, rsm 3, without motion artifact detection), qualitatively, they used different cutoff frequencies and rescaling. Quantitatively, we got that there was no significant difference between the model-based approach (AIC = -34.53) and the AUC (AIC = -32.49) ($\Delta|AIC| = 2.04$). Looking at the effect size of both told us that the model-based approach had a slightly better effect size than the AUC.

3.3. SCRV-4

3.3.1. motion artifact detection algorithm

For the SCRV-4 dataset, motion artifact detection identified 13 epochs (6 epochs with a CS+ event and 7 epochs with a CS- event) containing at least one motion artifact out of a total of 1984 epochs (31 participants multiplied by 64 epochs per participants). The mean percentage of epochs removed across participants was 1.81% with a standard deviation of 3.61%. The most epoch identified by motion artifact detection for CS+ was 2 and 3 for CS-. This time they came

from different participants. Once again, the arm of the participants was restrained to minimize the number of potential motion artifacts (Bach et al. (2009) cited in Bach, Daunizeau et al. (2010)).

For both the operationalist and model-based approaches (i.e. the DCM), none were capable of significantly differentiating between aversive and neutral event (see supplementary material). Thus, any sort of conclusion on which combination was the best was not very informative. All the results are still shown in the supplementary material. On the other hand, in the original article, the DCM did manage to differentiate significantly between aversive and neutral event. This might be due to the fact that we did not preprocess and did not compute the DCM the same way has in the original article.

3.4. Overall combination

The best PS combination in HRA-1 (PM2, rsm 7, MR 5 seconds and with motion artifact detection) and in SCRV-1 (PM2, rsm 7, MR = 5 seconds and without motion artifact detection) both used a low-pass filtering method, but the usage of motion artifact detection changed. The best AUC combination in HRA-1 (PM3, high cutoff 0.1 Hz, rsm 7, MR 4.5 seconds and with motion artifact detection) and in SCRV-1 (PM3, high cutoff 0.1 Hz, rsm 6, MR = 4 seconds, without motion artifact detection) both used a band-pass filtering method, but the rescaling method, the MR and the utilization of motion artifact detection were all different. In the case of the model-based approach, the best PsPM combination in HRA-1 (PM2, rsm 2, with motion artifact detection) and in the SCRV-1 (PM3, high cutoff 0.03 Hz, rsm 3, without motion artifact detection) were different.

If we look at the best combination for the HRA-1 and SCRV-1 datasets, in each case it was the model-based approach that showed better discrimination (albeit not statistically significantly different), but they did not use the same filtering method nor the same parameters.

Hence, even though the model-based approach seemed to offer slightly better discrimination, the preprocessing combination to use before extracting this metric seemed to depend on the dataset.

4. Discussion

SCR signal preprocessing is still a challenge despite the many attempts at establishing standards (Boucsein et al., 2012; Braithwaite et al., 2013; Figner & Murphy, 2011; Privratsky et al., 2020). This article aimed at adding to the empirical evidence provided by Privratsky et al. (2020) on the effects of preprocessing approaches on the ability to distinguish SCR responses from different experimental conditions. Specifically, similarly to Privratsky et al. (2020), we investigated the impacts of filtering and resampling on operationalist (PS) and model-based metrics but also tested 1) the inclusion of the motion artifact detection step (Taylor et al., 2015), 2) using the AUC as part of the operationalist metric and 3) using three different datasets to investigate if optimal preprocessing combinations are dataset specific. Our results suggest that adding the artifact detection to the preprocessing pipeline did not significantly impact the ability to distinguish experimental conditions, at least in the datasets we used. We also found some evidence that the AUC is a slightly better operationalist metric than PS but that the model-based approaches seem slightly better at contrasting experimental conditions than operationalist approaches. Finally, our data also show that the optimal preprocessing combination seems to be dataset specific.

4.1. Motion artifact detection algorithm effect

While SCR signals can easily be affected by artifacts from movements during the experiment, most previous work that looked at SCR preprocessing have not included an artifact detection and removal step (Bach, Daunizeau et al., 2010; Bach et al., 2009; Privratsky et al., 2020). Furthermore, to our knowledge, the effect of the motion artifact detection algorithm proposed by Taylor et al. (2015) has not been tested with SCR signal that was previously

preprocessed using different types of filters. Again, we seek to gather information on the influence of the algorithm on the optimal preprocessing approaches and not to test whether it should be included or not. From our results, we can see that, in our datasets, using an automatic artifact detection and removal algorithm generally did not improve the ability to distinguish SCR responses from different experimental conditions. Indeed, while for AUC, it slightly increased the effect size of the contrast between aversive and neutral events the difference was marginal and for PS, it generally did not improve the effect size. Similarly, in the case of model-based approach, motion artifact detection did not improve the effect size for the GLM, but it did marginally improve it for the DCM. Even though our data suggest that an artifact detection algorithm might not seem useful, one has to keep in mind that the three datasets we used only contained a small number of motion artifacts. Its usefulness would probably be more evident in experiments with more ecological experimental designs where participants are freely moving, producing larger numbers of artifacts. Because of the low number of motion artifacts in the datasets we used, future research should be conducted with both low and high artifact numbers to better assess the usefulness of motion artifact detection. This being said, by showing that the motion artifact detection algorithm step did not negatively affect SCR data, we provide evidence that this step should probably be used in preprocessing SCR datasets to obtain cleaner data when running statistics. This would be especially useful for future research that would like to use more ecological experimental designs where participants can move. Thus, we think that an automatic artifact detection step does not negatively impact the preprocessing and should probably be used.

4.2. The best metric to analyze SCR

To find out which is the best metric we first identified the best operationalist metric (i.e. PS vs. AUC) and then compared the operationalist to the model-based approaches.

4.2.1. PS vs. AUC

First of all, the preprocessing combination that led to the best discrimination for PS and AUC were different. Across the different datasets, the best combinations for PS included a low-pass filtering while for AUC they included band-pass filtering. Privratsky et al. (2020) found that the best combination for the PS was a band-pass filter which is different from our findings. We think that because Privratsky et al. (2020) dataset was taken during an fMRI session, the low-pass filter used was not enough thus making the band pass the best for them, but not for us. Another possibility is that the optimal filtering approach is datasets specific and thus researchers should always filter their data but test which filtering method works best for their datasets. Our results also show that, when looking at the best combination for each operationalist metric, the AUC seemed to have a slight advantage over the PS across all the dataset although the difference in effect size was not significantly better than the PS. This finding is in line with the work of Shukla et al. (2021) that found the AUC to be better than PS when analyzing SCR data. This advantage may arise from the fact that PS only uses two data points while the AUC uses more data points which may make it more robust and less prone to extreme values.

4.2.2. Operationalist vs. model-based approaches

By comparing for the first time preprocessing pipelines between model-based and AUC approaches, our work complement the findings from Privratsky et al. (2020). Overall, our result showed that, although there was no significant difference between the two, the best combinations for the model-based approach (i.e. the DCM and GLM) had slightly better effect sizes compared to the best combinations for the AUC (the best operationalist approach). Privratsky et al. (2020) found that the model-based approach was significantly better at distinguishing experimental conditions than the PS approach. Although our results go in the same direction then Privratsky et al. (2020), the lack of statistical difference in our datasets may be related to the fact that, as mentioned before, AUC seems to be a better operationalist method than PS. The fact that we did

not find significant difference between the AUC and the model-based approach might be a testament to the strength of the AUC as a good solution between an overly simple (i.e. PS) and very complex (i.e. PsPM) approach. Using increasingly complex approaches may provide better statistical power, but it comes at a price: these approaches are more complicated to use and require extensive computing time (e.g. DCM). Hence, researchers should take this into account when choosing which metric to use in their research.

4.3. Dataset specificity

Privratsky et al. (2020) recommended to use a band-pass filter for both the PS and the model-based approach as it was the filtering method that produced the best discrimination between experimental conditions in their dataset. Our results, on the other hand, suggest that the type of filtering used in the optimal preprocessing pipeline varies between datasets. For example, with the model-based approach, a low-pass filter was the best for HRA-1, while for SCRV-1 a band-pass filter was the best. They also used different rescaling methods (i.e. rsm 2 for HRA-1 and rsm 3 for SCRV-1) and did also not use motion artifact detection the same way (i.e. HRA-1 used motion artifact detection whereas SCRV-1 did not). Similarly, with the PS, for HRA-1 the best preprocessing combination used a low-pass filter while for SCRV-1, the low-pass filter was still the best, but the performance of the band-pass filter was much closer to the low-pass filter than for HRA-1. Furthermore, HRA-1 used motion artifact detection while SCRV-1 did not. The opposite trend was observed for the AUC: for SCRV-1 the best filtering method was a band-pass filter while for HRA-1 the low-pass filter's performance was much closer to one of the band-pass filter, in both cases the rescaling used was different (i.e. rsm 7 for HRA-1 and rsm 6 for SCRV-1) and again both did not use motion artifact detection the same way (i.e. HRA-1 used motion artifact detection whereas SCRV-1 did not). Overall, our results suggest that the characteristics of the dataset could play an important role in determining the optimal combination of filtering and

rescaling methods to use when preprocessing the signal. This idea is in line with Privratsky et al. (2020) suggestion that datasets with different ITI might behave differently with different filters, which was the case with the datasets we used (HRA-1 ITI : 7, 9 or 11 s; SCRV-1 ISI : 3, 9 or 19 s; SCRV-4 ITI : 14, 19 or 23 s). By comparing optimal preprocessing combinations between different datasets, we provide additional arguments for the importance of carefully considering how to preprocess each dataset. The fact that the optimal preprocessing combination seems to be dataset specific suggests that researchers should try to identify what the best combination for their dataset is. Although, this might sound daunting and time consuming, the code we wrote to test the different combinations in this paper and that is freely available could be very useful for future researchers. Moreover, unfortunately our analysis does not allow us to know *why* some specific combination of preprocessing parameters seem to be better for some datasets. One possibility is that the combinations are related to the difference in signal acquisition procedures and parameters between each dataset. This is an interesting question that should be investigated in future research. Overall, the data and code provided here are aimed at promoting the importance of preprocessing SCR signal. Even though no single best combination showed up, across all the dataset and across all metrics, filters were always used showing us the importance of filtering the SCR, which is also what Privratsky et al. (2020) had found.

4.4. Limitations

First of all, we did not test the effectiveness of the motion artifact detection algorithm but rather its usefulness in a typical SCR signal preprocessing pipeline. Future research could use datasets containing a large number of motion artifacts to test motion artifact detection algorithm capabilities with different filtering methods and test how the addition of this step during the preprocessing may produce stronger statistical results. Secondly, how we computed the AUC was based on the approach proposed in Bach, Friston et al. (2010). Maybe other ways of computing

the AUC, such as the one described in Boucsein (2012) or even the simple Riemann sum might have yield different results. Future work could compare the different ways to compute the AUC. Thirdly, of the three datasets we used, two had extremely similar experimental designs (i.e. HRA-1 and SCRV-4) and all datasets, including the dataset from Privratsky et al. (2020), compared neutral to aversive experimental conditions. In future research, more diverse datasets should be tested with our code or at least with different preprocessing combinations. In addition, further work is needed to try to identify if and how different characteristics (e.g. ITI) influence the type of preprocessing producing optimal statistical results.

4.6. Recommendations

Our results both confirm and complement those of Privratsky et al. (2020). Foremost, our data confirm the importance of preprocessing SCR signal before running statistical analysis. Researchers should use available tools (including the ones we provide with this article) to test what preprocessing combination works best with their datasets. In addition, although we acknowledge the force and the advantages of model-based approaches, we also recognize the simplicity of the implementation of the operationalist approach, especially the AUC which can be used to circumvent some of the criticism towards the operationalist approach.

References

Bach, D. R., Castegnetti, G., Korn, C. W., Gerster, S., Melinscak, F., & Moser, T. (2018). Psychophysiological modeling: Current state and future directions. *Psychophysiology*, *55*(11), e13214. https://doi.org/10.1111/psyp.13209

Bach, D. R., Daunizeau, J., Friston, K. J., & Dolan, R. J. (2010). Dynamic causal modelling of anticipatory skin conductance responses. *Biological Psychology*, *85*(1), 163–170. https://doi.org/10.1016/j.biopsycho.2010.06.007

Bach, D. R., Flandin, G., Friston, K. J., & Dolan, R. J. (2009). Time-series analysis for rapid event-related skin conductance responses. *Journal of Neuroscience Methods*, *184*(2), 224–234. https://doi.org/10.1016/j.jneumeth.2009.08.005

Bach, D. R., Flandin, G., Friston, K. J., & Dolan, R. J. (2010). Modelling event-related skin conductance responses. *International Journal of Psychophysiology*, *75*(3), 349–356. https://doi.org/10.1016/j.ijpsycho.2010.01.005

Bach, D. R., & Friston, K. J. (2013). Model-based analysis of skin conductance responses: Towards causal models in psychophysiology. *Psychophysiology*, *50*(1), 15–22. https://doi.org/10.1111/j.1469-8986.2012.01483.x

Bach, D. R., Friston, K. J., & Dolan, R. J. (2010). Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations. *International Journal of Psychophysiology*, *76*(1), 52–55. https://doi.org/10.1016/j.ijpsycho.2010.01.011

Bach, D. R., Friston, K. J., & Dolan, R. J. (2013). An improved algorithm for model-based analysis of evoked skin conductance responses. *Biological Psychology*, *94*(3), 490–497. https://doi.org/10.1016/j.biopsycho.2013.09.010

Bilodeau-Houle, A., Morand-Beaulieu, S., Bouchard, V., & Marin, M.-F. (2023). Parent–child physiological concordance predicts stronger observational fear learning in children with a

less secure relationship with their parent. *Journal of Experimental Child Psychology, 226*, 105553. https://doi.org/10.1016/j.jecp.2022.105553

Boucsein, W. (2012). *Electrodermal Activity*. Springer US. https://doi.org/10.1007/978-1-4614-1126-0

Boucsein, W., Fowles, D. C., Grimnes, S., Gershon, B.-S., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*, 1017–1034.

Braithwaite, D. J. J., Watson, D. D. G., Jones, R., & Rowe, M. (2013). A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments. *Psychophysiology, 49*, 1017–1034.

Cuve, H. C. J., Harper, J., Catmur, C., & Bird, G. (2023). Coherence and divergence in autonomic-subjective affective space. *Psychophysiology*, *n/a*(n/a), e14262. https://doi.org/10.1111/psyp.14262

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), Article 1. https://doi.org/10.1038/s41592-018-0235-4

Fahrenberg, J., Walschburger, P., Foerster, F., Myrtek, M., & Müller, W. (1983). An Evaluation of Trait, State, and Reaction Aspects of Activation Processes. *Psychophysiology*, *20*(2), 188–195. https://doi.org/10.1111/j.1469-8986.1983.tb03286.x

Figner, B., & Murphy, R. O. (2011). Using skin conductance in judgment and decision making research. In *A handbook of process tracing methods for decision research* (p. 33).

Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*. https://www.frontiersin.org/articles/10.3389/fnins.2013.00267

Harnett, N. G., Fani, N., Carter, S., Sanchez, L. D., Rowland, G. E., Davie, W. M., Guzman, C., Lebois, L. A. M., Ely, T. D., van Rooij, S. J. H., Seligowski, A. V., Winters, S., Grasser, L. R., Musey, P. I., Seamon, M. J., House, S. L., Beaudoin, F. L., An, X., Zeng, D., … Ressler, K. J. (2023). Structural inequities contribute to racial/ethnic differences in neurophysiological tone, but not threat reactivity, after trauma exposure. *Molecular Psychiatry*, 1–10. https://doi.org/10.1038/s41380-023-01971-x

Koppold, A., Kastrinogiannis, A., Kuhn, M., & Lonsdorf, T. B. (2022). Watching with Argus eyes: Characterization of emotional and physiological responding in adults exposed to childhood maltreatment and/or recent adversity. *Psychophysiology*, *n/a*(n/a), e14253. https://doi.org/10.1111/psyp.14253

Lang, P., Bradley, M., & Cuthbert, B. (2005). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical Report*. https://cir.nii.ac.jp/crid/1573950399053852928

Liberman, L., & Dubovi, I. (2022). The effect of the modality principle to support learning with virtual reality: An eye-tracking and electrodermal activity study. *Journal of Computer Assisted Learning*, *1*(11). https://doi.org/10.1111/jcal.12763

Lim, C. L., Rennie, C., Barry, R. J., Bahramali, H., Lazzaro, I., Manor, B., & Gordon, E. (1997). Decomposing skin conductance into tonic and phasic components. *International Journal of Psychophysiology*, *25*(2), 97–109. https://doi.org/10.1016/S0167-8760(96)00713-1

Moretta, T., Kaess, M., & Koenig, J. (2023). A comparative evaluation of resting state proxies of sympathetic and parasympathetic nervous system activity in adolescent major depression. *Journal of Neural Transmission*, *130*(2), 135–144. https://doi.org/10.1007/s00702-022-02577-3

Posada-Quintero, H. F., & Chon, K. H. (2020). Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review. *Sensors*, *20*(2), 479. https://doi.org/10.3390/s20020479

Privratsky, A. A., Bush, K. A., Bach, D. R., Hahn, E. M., & Cisler, J. M. (2020). Filtering and model-based analysis independently improve skin-conductance response measures in the fMRI environment: Validation in a sample of women with PTSD. *International Journal of Psychophysiology*, *158*, 86–95. https://doi.org/10.1016/j.ijpsycho.2020.09.015

Shukla, J., Barreda-Ángeles, M., Oliver, J., Nandi, G. C., & Puig, D. (2021). Feature Extraction and Selection for Emotion Recognition from Electrodermal Activity. *IEEE Transactions on Affective Computing*, *12*(4), 857–869. https://doi.org/10.1109/TAFFC.2019.2901673

Sklivanioti Greenfield, M., Wang, Y., & Msghina, M. (2022). Similarities and differences in the induction and regulation of the negative emotions fear and disgust: A functional near infrared spectroscopy study. *Scandinavian Journal of Psychology*, *63*(6), 581–593. https://doi.org/10.1111/sjop.12836

Staib, M., Castegnetti, G., & Bach, D. R. (2015). Optimising a model-based approach to inferring fear learning from skin conductance responses. *Journal of Neuroscience Methods*, *255*, 131–138. https://doi.org/10.1016/j.jneumeth.2015.08.009

Starita, F., Pirazzini, G., Ricci, G., Garofalo, S., Dalbagno, D., Degni, L. A. E., Di Pellegrino, G., Magosso, E., & Ursino, M. (2022). Theta and alpha power track the acquisition and

reversal of threat predictions and correlate with skin conductance response. *Psychophysiology*, *n/a*(n/a), e14247. https://doi.org/10.1111/psyp.14247

Stefaniak, A. R., Blaxton, J. M., & Bergeman, C. S. (2021). Age Differences in Types and Perceptions of Daily Stress. *The International Journal of Aging and Human Development*, *94*(2), 215–233. https://doi.org/10.1177/00914150211001588

Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., & Picard, R. (2015). Automatic identification of artifacts in electrodermal activity data. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1934–1937. https://doi.org/10.1109/EMBC.2015.7318762

Wehrli, J. M., Xia, Y., Gerster, S., & Bach, D. R. (2022). Measuring human trace fear conditioning. *Psychophysiology*, *59*(12), e14119. https://doi.org/10.1111/psyp.14119

Supplementary

HRA-1 and SCRV-4 data acquisition

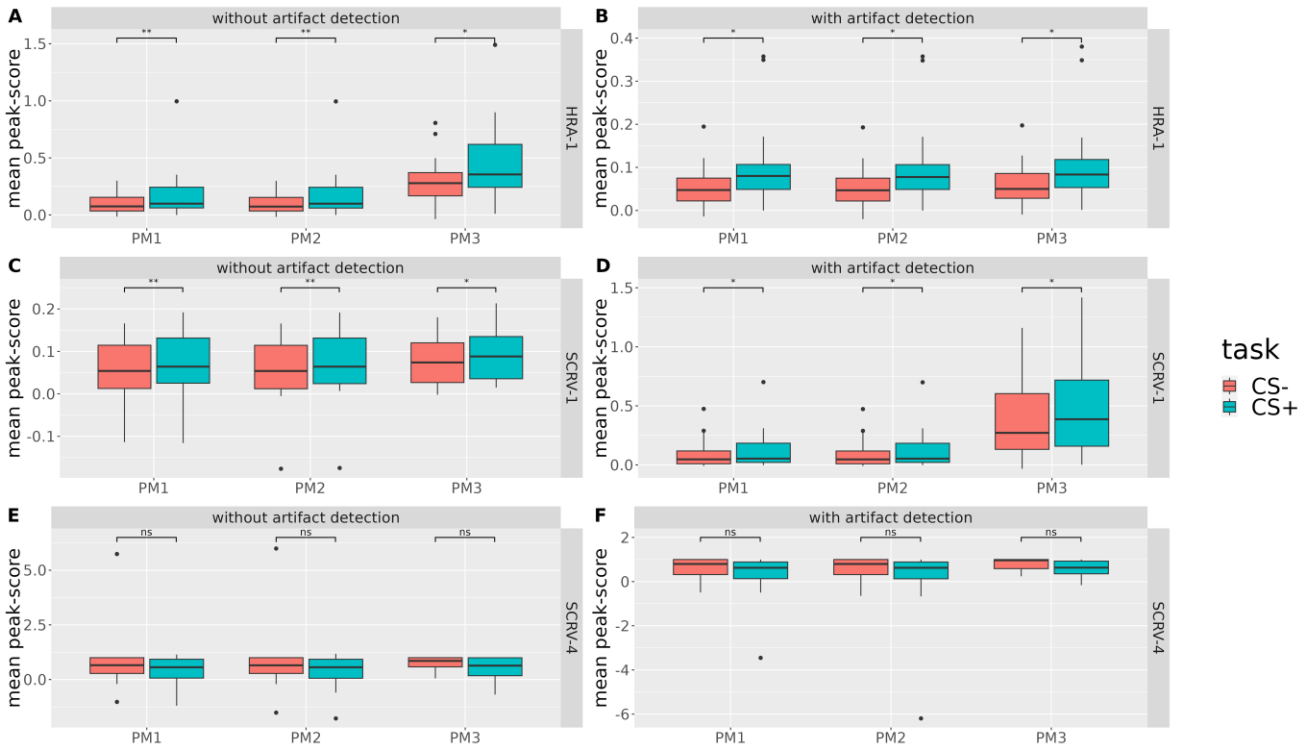
Bach, Daunizeau et al. (2010) recorded skin conductance with electrodes place on the thenar/hypothenar of the non-dominant hand of participants. They used a 100 Hz sampling rate with a constant voltage of 2.5 V. The temperature was between 18 °C and 25 °C and had a relative humidity between 31% and 51% for both experiment (Bach, Daunizeau et al., 2010). For more details, we invite the reader to consult Bach, Daunizeau et al. (2010).

SCRV-1 data acquisition

Skin conductance was recorded with electrodes place on the thenar and hypothenar of the non-dominant hand, although this time Bach et al. (2009) used a motion-restraining armrest to limit motion artifacts. They reported that their lowest sampling rate was 36.7 Hz, but all samplings in the dataset are at 100 Hz, and use a constant voltage of 2.5 V (Bach et al., 2009). Finally, the data was acquired in a magnetically shielded room (MSR) where the temperature varied between 18 °C and 24 °C and the relative humidity varied between 30% and 63% (Bach et al., 2009).

Supplementary Figure 1

Mean of PS across different datasets without and with motion artifact detection



Note. Mean PS elicited by aversive stimuli were significantly different than the mean PS elicited by neutral stimuli in HRA-1 (A and B) and SCRV-1 (C and D) without and with motion artifact detection. For SCRV-4 (E and F) no significant differences were observed, which was similar with the original results of the article. For SCRV-1, in the original article there was no significant difference between the aversive and neutral stimuli. This might be due to the difference in signal preprocessing and PS computation method between our work and the original article. The differences in the scale of the graphics were due to the fact that each uses different preprocessing combinations which result in different scale.

* : $p < 0.05$

** : $p < 0.005$

Supplementary Table 1

Significance levels and effect sizes Wilcoxon signed-rank test of the PS with different processing methods between mean aversive and mean neutral conditions without motion artifact detection for all the datasets.

	PM1	PM2	PM3
HRA-1	Aversive (Mdn = 0.10) Neutral (Mdn = 0.08) p = 0.004 r = 0.64	Aversive (Mdn = 0.10) Neutral (Mdn = 0.07) p = 0.004 r = 0.64	Aversive (Mdn = 0.36) Neutral (Mdn = 0.28) p = 0.02 r = 0.51
SCRV-1	Aversive (Mdn = 0.06) Neutral (Mdn = 0.05) p = 0.002 r = 0.65	Aversive (Mdn = 0.06) Neutral (Mdn = 0.05) p = 0.002 r = 0.66	Aversive (Mdn = 0.09) Neutral (Mdn = 0.07) p = 0.005 r = 0.60
SCRV-4	Aversive (Mdn = 0.56) Neutral (Mdn = 0.66) p = 0.95 r = 0.01	Aversive (Mdn = 0.56) Neutral (Mdn = 0.65) p = 0.96 r = 0.009	Aversive (Mdn = 0.64) Neutral (Mdn = 0.85) p = 0.98 r = 0.005

Note. For all the processing methods for both HRA-1 and SCRV-1, the difference between the mean PS of aversive SCR and mean PS of neutral SCR were significant. For HRA-1 and SCRV-1, the effect sizes were all above r = 0.5. HRA-1 and SCRV-1 had similar effect sizes except for PM3, SCRV-1 had a higher effect size. There was no significant difference between the mean PS of aversive SCR and the mean PS of neutral SCR for SCRV-4, which was in line with the original article results.

Supplementary Table 2

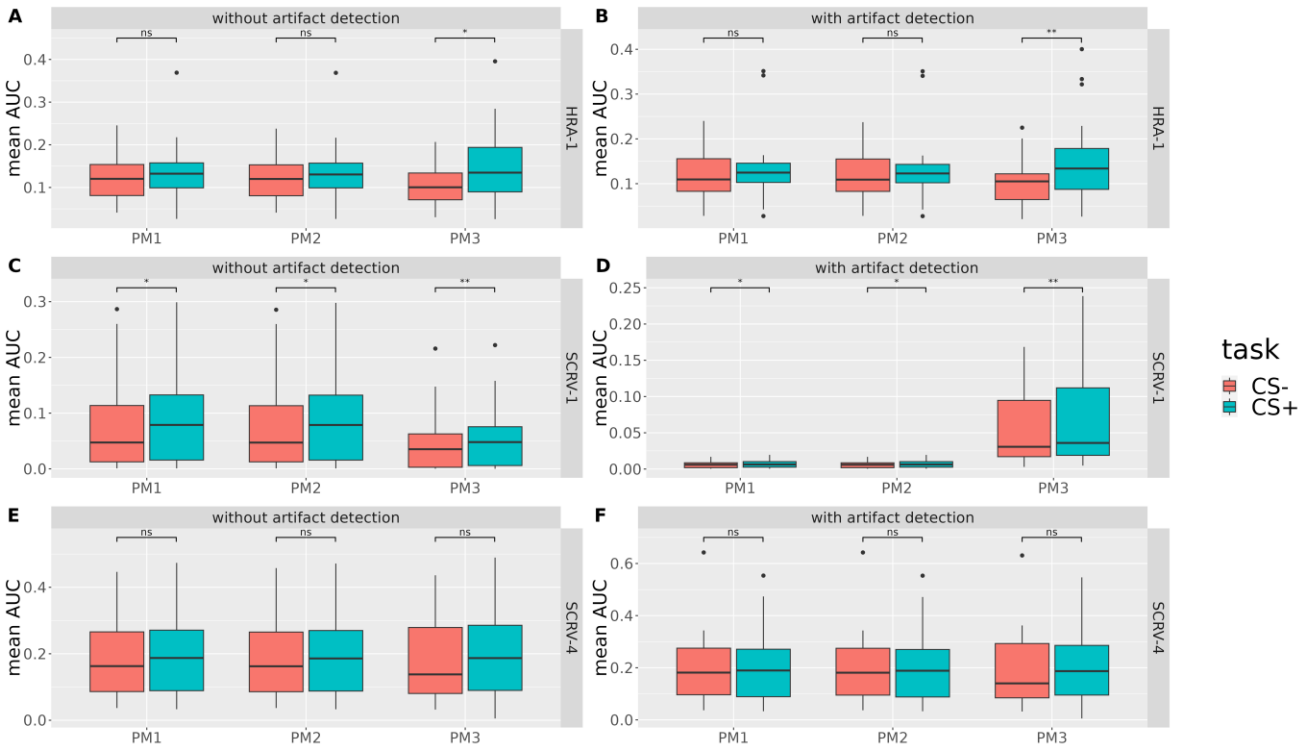
Significance levels and effect sizes Wilcoxon signed-rank test of the PS with different processing methods between mean aversive and mean neutral conditions with motion artifact detection for all the datasets.

	PM1	PM2	PM3
HRA-1	Aversive (Mdn = 0.08) Neutral (Mdn = 0.05) $p = 0.006$ $r = 0.61$	Aversive (Mdn = 0.08) Neutral (Mdn = 0.05) $p = 0.006$ $r = 0.61$	Aversive (Mdn = 0.08) Neutral (Mdn = 0.05) $p = 0.009$ $r = 0.59$
SCRV-1	Aversive (Mdn = 0.05) Neutral (Mdn = 0.05) $p = 0.03$ $r = 0.48$	Aversive (Mdn = 0.05) Neutral (Mdn = 0.05) $p = 0.02$ $r = 0.48$	Aversive (Mdn = 0.39) Neutral (Mdn = 0.27) $p = 0.009$ $r = 0.56$
SCRV-4	Aversive (Mdn = 0.63) Neutral (Mdn = 0.80) $p = 0.98$ $r = 0.005$	Aversive (Mdn = 0.63) Neutral (Mdn = 0.80) $p = 0.98$ $r = 0.005$	Aversive (Mdn = 0.63) Neutral (Mdn = 0.97) $p = 1$ $r = 0.0006$

Note. For all the processing methods for both HRA-1 and SCRV-1, the difference between the mean PS of aversive SCR and mean PS of neutral SCR were significant. For HRA-1 and SCRV-1, the effect sizes were all above $r = 0.47$. Overall, HRA-1 had higher effect size than SCRV-1. There was no significant difference between the mean PS of aversive SCR and the mean PS neutral SCR for SCRV-4, which was in line with the original article results.

Supplementary Figure 2

Mean AUC across different datasets without and with motion artifact detection



Note. Mean AUC elicited by aversive stimuli were significantly different than the mean AUC elicited by neutral stimuli in HRA-1 (A and B) and SCRV-1 (C and D) without and with motion artifact detection. For SCRV-4 (E and F) no significant differences were observed. The differences in the scale of the graphics were due to the fact that each uses different preprocessing combinations which result in different scale.

* : $p < 0.05$

** : $p < 0.005$

Supplementary Table 3

Significance levels and effect sizes Wilcoxon signed-rank test of the AUC with different processing methods between mean aversive and mean neutral conditions without motion artifact detection for all the datasets.

	PM1		PM2		PM3	
HRA-1	Aversive (Mdn =	Neutral (Mdn =	Aversive (Mdn =	Neutral (Mdn =	Aversive (Mdn =	Neutral (Mdn =

	0.13) 0.12) $p = 0.25$ $r = 0.26$	0.13) 0.12) $p = 0.25$ $r = 0.26$	0.13) 0.10) $p = 0.008$ $r = 0.60$
SCRV-1	Aversive (Mdn = 0.08) Neutral (Mdn = 0.05) $p = 0.006$ $r = 0.58$	Aversive (Mdn = 0.08) Neutral (Mdn = 0.05) $p = 0.006$ $r = 0.58$	Aversive (Mdn = 0.05) Neutral (Mdn = 0.04) $p = 0.0003$ $r = 0.76$
SCRV-4	Aversive (Mdn = 0.19) Neutral (Mdn = 0.16) $p = 0.3$ $r = 0.18$	Aversive (Mdn = 0.19) Neutral (Mdn = 0.16) $p = 0.3$ $r = 0.18$	Aversive (Mdn = 0.19) Neutral (Mdn = 0.14) $p = 0.12$ $r = 0.28$

Note. Only for PM3 the difference between the mean AUC of aversive SCR was significantly different from the mean AUC of neutral SCR in the case of HRA-1. For all the processing methods for SCRV-1, the difference between the mean aversive SCR and neutral SCR were significant. The effect sizes of SCRV-1 were all above $r = 0.57$. Overall, SCRV-1 had higher effect size than HRA-1. There was no significant difference between the mean aversive SCR and neutral SCR for SCRV-4.

Supplementary Table 4

Significance levels and effect sizes Wilcoxon signed-rank test of the AUC with different processing methods between mean aversive and mean neutral conditions with motion artifact detection for all the datasets.

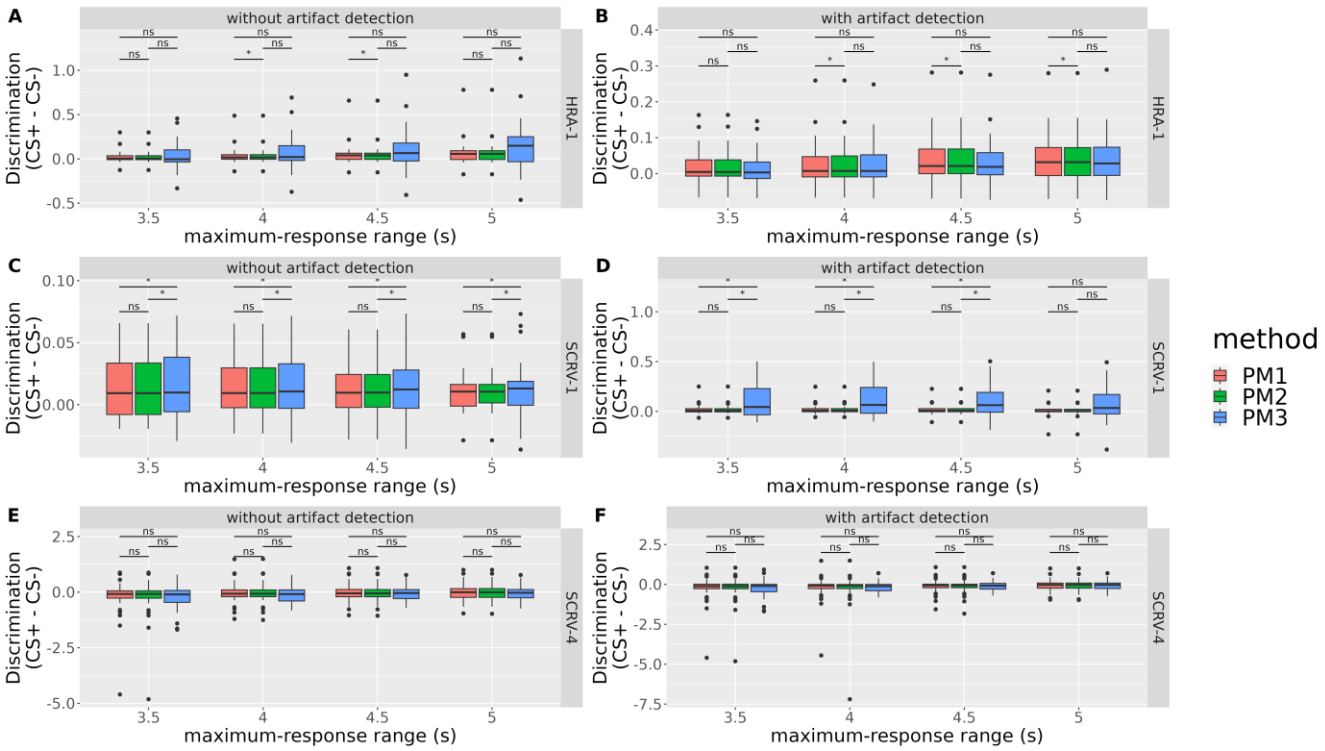
	PM1	PM2	PM3
HRA-1	Aversive (Mdn = 0.12) Neutral (Mdn = 0.11) $p = 0.20$ $r = 0.28$	Aversive (Mdn = 0.12) Neutral (Mdn = 0.11) $p = 0.20$ $r = 0.28$	Aversive (Mdn = 0.13) Neutral (Mdn = 0.11) $p = 0.002$ $r = 0.69$
SCRV-1	Aversive Neutral	Aversive Neutral	Aversive Neutral

	(Mdn = 0.006) (Mdn = 0.006) $p = 0.02$ $r = 0.48$	(Mdn = 0.006) (Mdn = 0.006) $p = 0.02$ $r = 0.48$	(Mdn = 0.04) (Mdn = 0.03) $p = 0.0005$ $r = 0.74$
SCRV-4	Aversive (Mdn = 0.19) Neutral (Mdn = 0.18) $p = 0.49$ $r = 0.12$	Aversive (Mdn = 0.19) Neutral (Mdn = 0.18) $p = 0.49$ $r = 0.12$	Aversive (Mdn = 0.19) Neutral (Mdn = 0.14) $P = 0.2$ $r = 0.23$

Note. Only for PM3 the difference between the mean AUC of aversive SCR was significantly different from the mean AUC of neutral SCR in the case of HRA-1. For all the processing methods for SCRV-1, the difference between the mean aversive SCR and neutral SCR were significant. The effect sizes of SCRV-1 were all above $r = 0.47$. Overall, SCRV-1 had higher effect size than HRA-1. There was no significant difference between the mean aversive SCR and neutral SCR for SCRV-4.

Supplementary Figure 3

PS based discrimination of aversive PS and neutral PS between different processing methods and maximum response range for all datasets without and with motion artifact detection



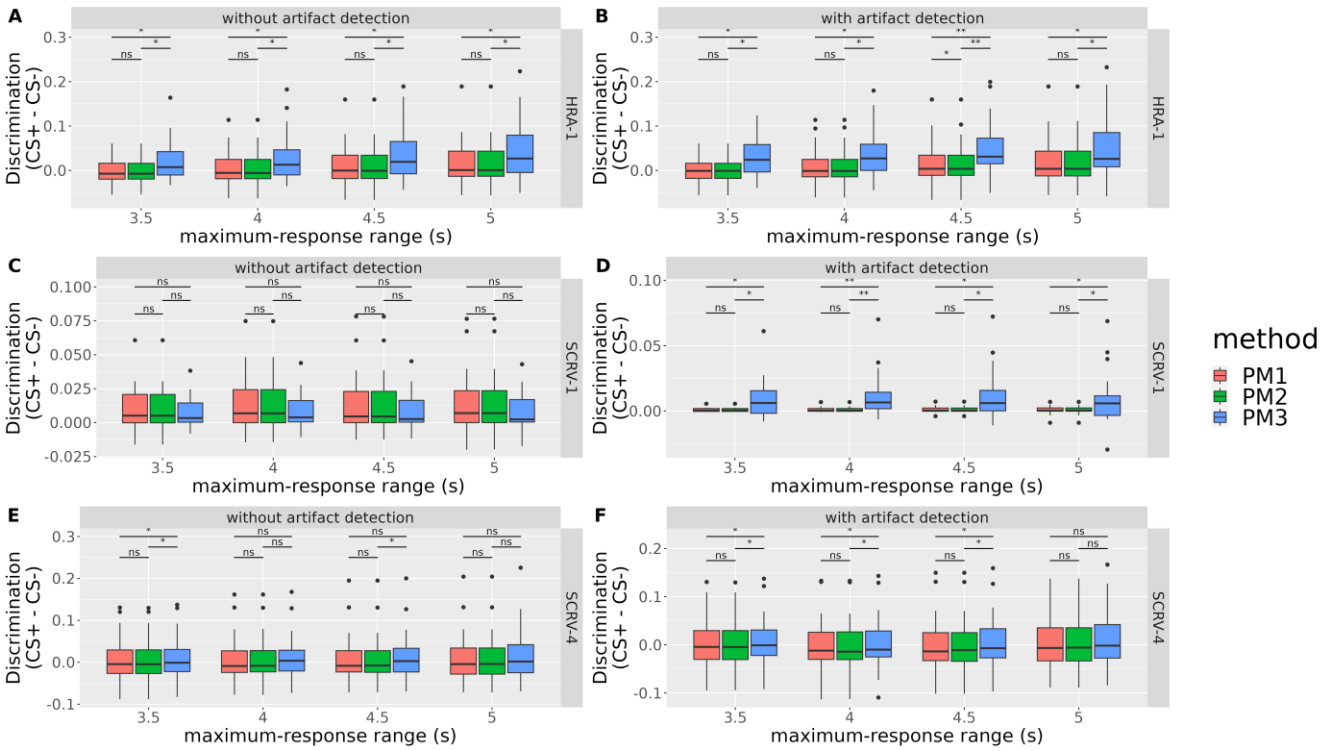
Note. The difference between all the aversive PS and neutral PS for each PM at different MR was computed, then a Wilcoxon signed-rank test was computed to see if there was significant difference in the discrimination capabilities of each PM within the same MR. For HRA-1 (A and B), the only significant differences were between PM1 and PM2, but not at all MRs. For SCRV-1, all the significant differences were between PM3 vs. PM1 and PM3 vs. PM2, again not at all MRs. Finally, there was no significant difference between any PM for any MR for SCRV-4. The differences in the scale of the graphics were due to the fact that each uses different preprocessing combinations which result in different scale.

* : $p < 0.05$

** : $p < 0.005$

Supplementary Figure 4

AUC based discrimination of aversive AUC and neutral AUC between different processing methods and maximum response range for all datasets without and with motion artifact detection



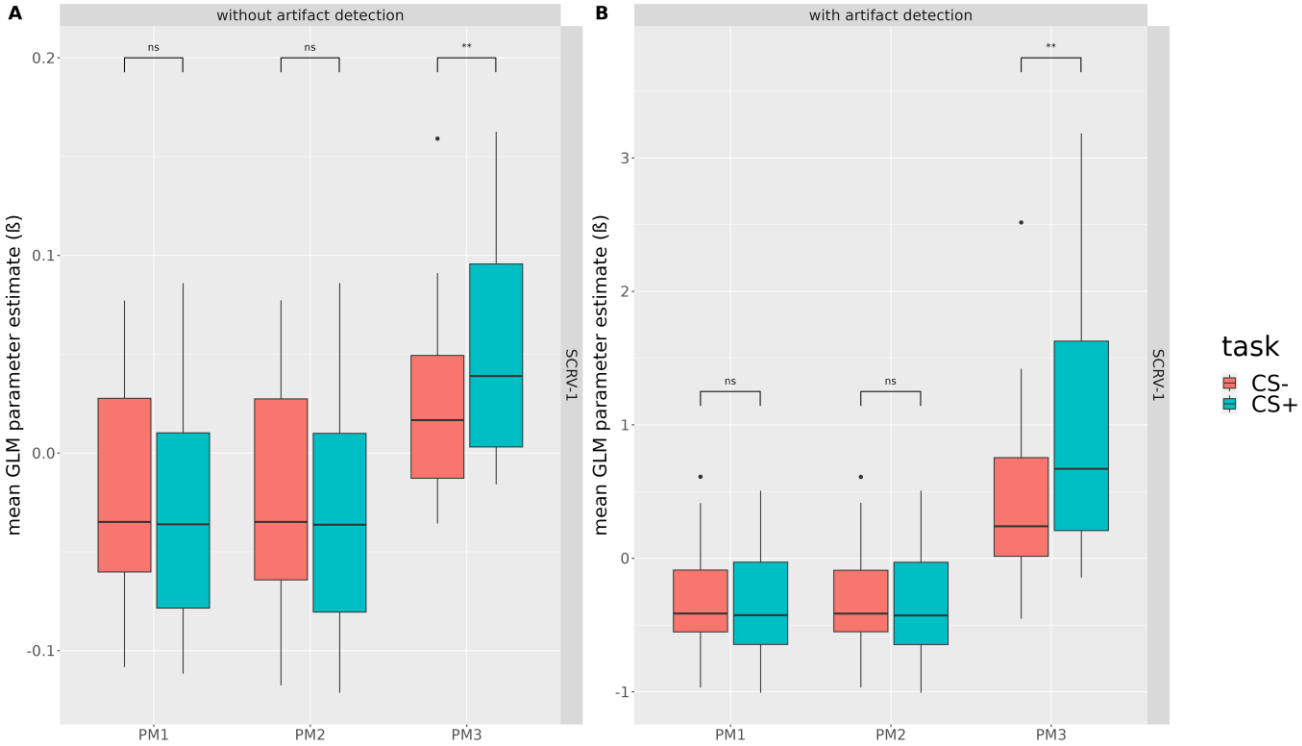
Note. The difference between all the aversive AUC and neutral AUC for each PM at different MR was computed, then a Wilcoxon signed-rank test was computed to see if there was significant difference in the discrimination capabilities of each PM within the same MR. For HRA-1 (A and B), the significant differences were between PM3 vs. PM1 and PM3 vs. PM2 across all PMs and MRs for both without and with motion artifact detection. For SCRV-1, all the significant differences were between PM3 vs. PM1 and PM3 vs. PM2 across all PMs and MRs, but only with motion artifact detection. There was no significant difference without motion artifact detection. Finally, the only significant difference for SCRV-4 were for PM3 vs. PM1 and PM3 vs. PM2, but not for all PMs and not across all MRs. The differences in the scale of the graphics were due to the fact that each uses different preprocessing combinations which result in different scale.

* : $p < 0.05$

** : $p < 0.005$

Supplementary Figure 5

Mean parameter estimate for PsPM for SCRV-1 without and with motion artifact detection



Note. For both without and with motion artifact detection there was only significant difference between aversive estimate parameter and neutral estimate parameter for PM3. The fact that with PM3 mean aversive and neutral parameter estimate had a significant difference between them was similar to the original article results. The differences in the scale of the graphics were due to the fact that each uses different preprocessing combinations which result in different scale.

* : $p < 0.05$

** : $p < 0.005$

Supplementary Table 5

Significance levels and effect sizes Wilcoxon signed-rank test of the mean PsPM estimated parameters for aversive and neutral conditions without motion artifact detection for SCRV-1.

PM1	PM2	PM3

SCRV-1	Aversive (Mdn = -0.036)	Neutral (Mdn = -0.035)	Aversive (Mdn = -0.036)	Neutral (Mdn = -0.035)	Aversive (Mdn = 0.039)	Neutral (Mdn = 0.017)
	p = 0.99 r = 0.0037		p = 0.99 r = 0.0034		p = 0.0002 r = 0.80	

Note. The only significant difference between the mean aversive estimate parameter and the mean neutral estimate parameter was for PM3 and its effect size as very high.

Supplementary Table 6

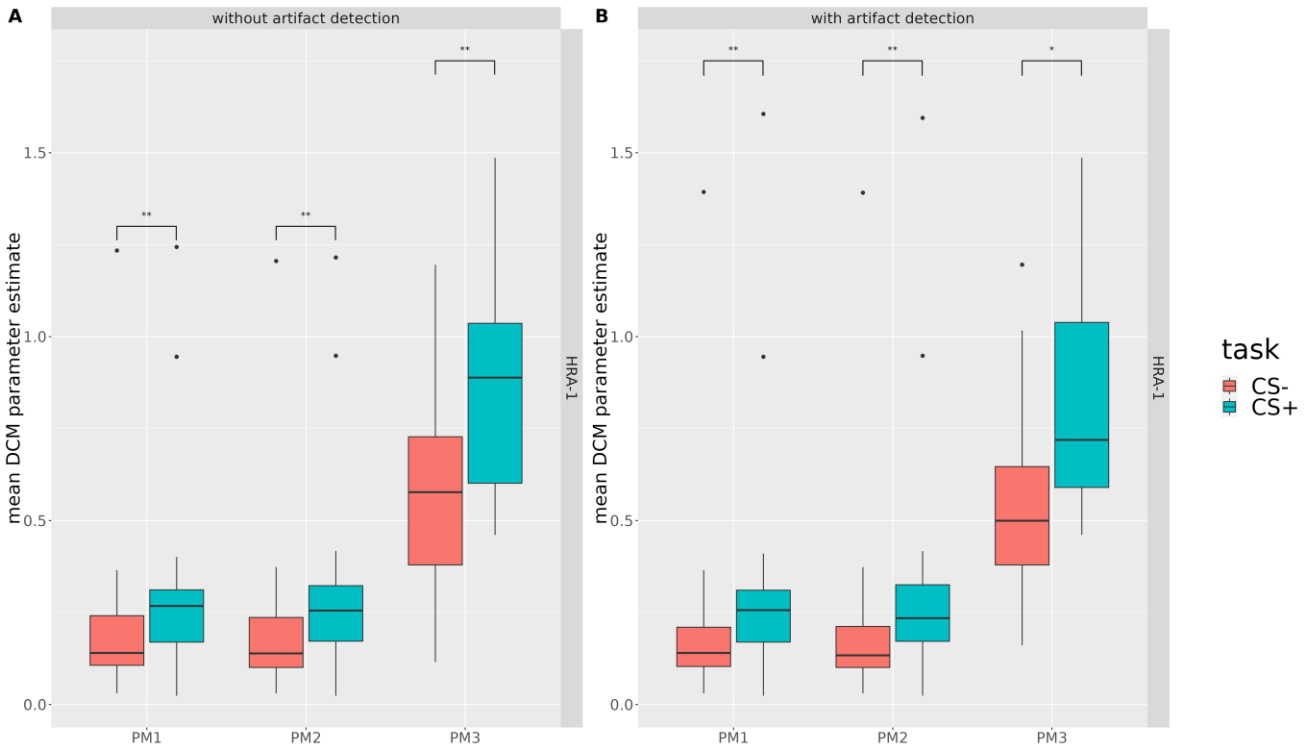
Significance levels and effect sizes Wilcoxon signed-rank test of the mean PsPM estimated parameters for aversive and neutral conditions with motion artifact detection for SCRV-1.

	PM1		PM2		PM3	
SCRV-1	Aversive (Mdn = -0.43)	Neutral (Mdn = -0.41)	Aversive (Mdn = -0.43)	Neutral (Mdn = -0.41)	Aversive (Mdn = 0.67)	Neutral (Mdn = 0.24)
	p = 0.94 r = 0.015		P = 0.94 r = 0.015		p = 0.001 r = 0.68	

Note. The only significant difference between the mean aversive estimate parameter and the mean neutral estimate parameter was for PM3 and its effect size was high.

Supplementary Figure 6

Mean parameter estimate for PsPM for HRA-1 without and with motion artifact detection



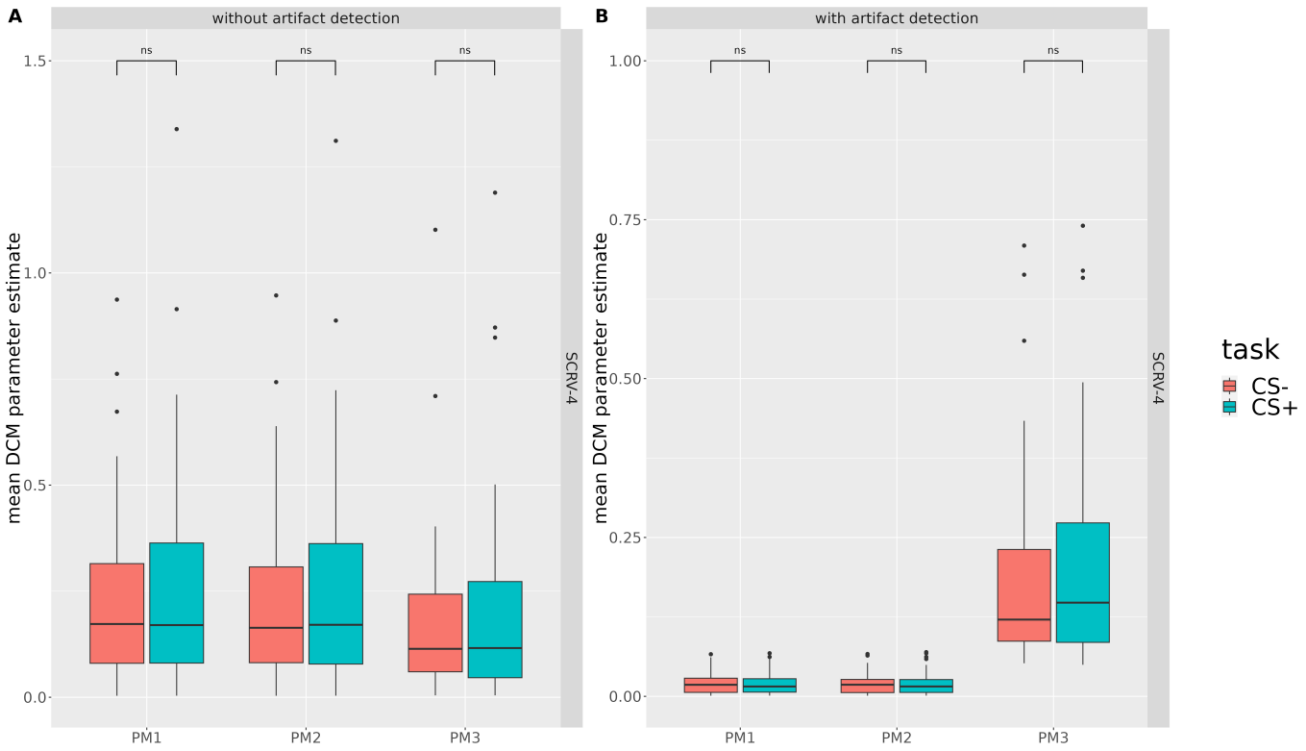
Note. For both without and with motion artifact detection there was significant difference between aversive estimate parameter and neutral estimate parameter for all PMs. The original article did not have different PM, but they still found significant difference between the mean aversive and the mean neutral estimate parameter.

* : $p < 0.05$

** : $p < 0.005$

Supplementary Figure 7

Mean parameter estimate for PsPM for SCRV-4 without and with motion artifact detection



Note. For both without and with motion artifact detection there was no significant difference between aversive estimate parameter and neutral estimate parameter for all PMs. The original article did not have different PM, but did find significant difference between the mean aversive and the neutral estimate parameter. The difference between our results and the original results might be due to a difference in preprocessing or estimate parameter computation. The differences in the scale of the graphics were due to the fact that each uses different preprocessing combinations which result in different scale.

* : $p < 0.05$

** : $p < 0.005$

Supplementary Table 7

Significance levels and effect sizes Wilcoxon signed-rank test of the mean PsPM estimated parameters for aversive and neutral conditions without motion artifact detection for HRA-1.

PM1	PM2	PM3

HRA-1	Aversive (Mdn = 0.27) Neutral (Mdn = 0.14) $p = 0.002$ $r = 0.69$	Aversive (Mdn = 0.26) Neutral (Mdn = 0.14) $p = 0.002$ $r = 0.69$	Aversive (Mdn = 0.89) Neutral (Mdn = 0.58) $P = 0.003$ $r = 0.67$

Note. All the PM had a significant difference between the mean aversive and the mean neutral estimate parameters. The effect sizes of all PMs were above $r = 0.66$.

Supplementary Table 8

Significance levels and effect sizes Wilcoxon signed-rank test of the mean PsPM estimated parameters for aversive and neutral conditions with motion artifact detection for HRA-1.

	PM1	PM2	PM3
HRA-1	Aversive (Mdn = 0.26) Neutral (Mdn = 0.14) $p = 0.0007$ $r = 0.76$	Aversive (Mdn = 0.23) Neutral (Mdn = 0.13) $p = 0.0006$ $r = 0.77$	Aversive (Mdn = 0.72) Neutral (Mdn = 0.50) $p = 0.005$ $r = 0.62$

Note. All the PM had a significant difference between the mean aversive and the mean neutral estimate parameters. The effect sizes of all PMs were above $r = 0.61$.

Supplementary Table 9

Significance levels and effect sizes Wilcoxon signed-rank test of the mean PsPM estimated parameters for aversive and neutral conditions without motion artifact detection for SCRV-4.

	PM1	PM2	PM3
SCRV-4	Aversive (Mdn = 0.015) Neutral (Mdn = 0.018) $p = 0.80$ $r = 0.045$	Aversive (Mdn = 0.016) Neutral (Mdn = 0.018) $p = 0.68$ $r = 0.075$	Aversive (Mdn = 0.15) Neutral (Mdn = 0.12) $p = 0.21$ $r = 0.22$

Note. None of the PM had a significant difference between the mean aversive and neutral parameter estimate. This result was different from the results of the original article, this might be due to the fact that the preprocessing was different and that the mean parameter estimate had been computed differently

Supplementary Table 10

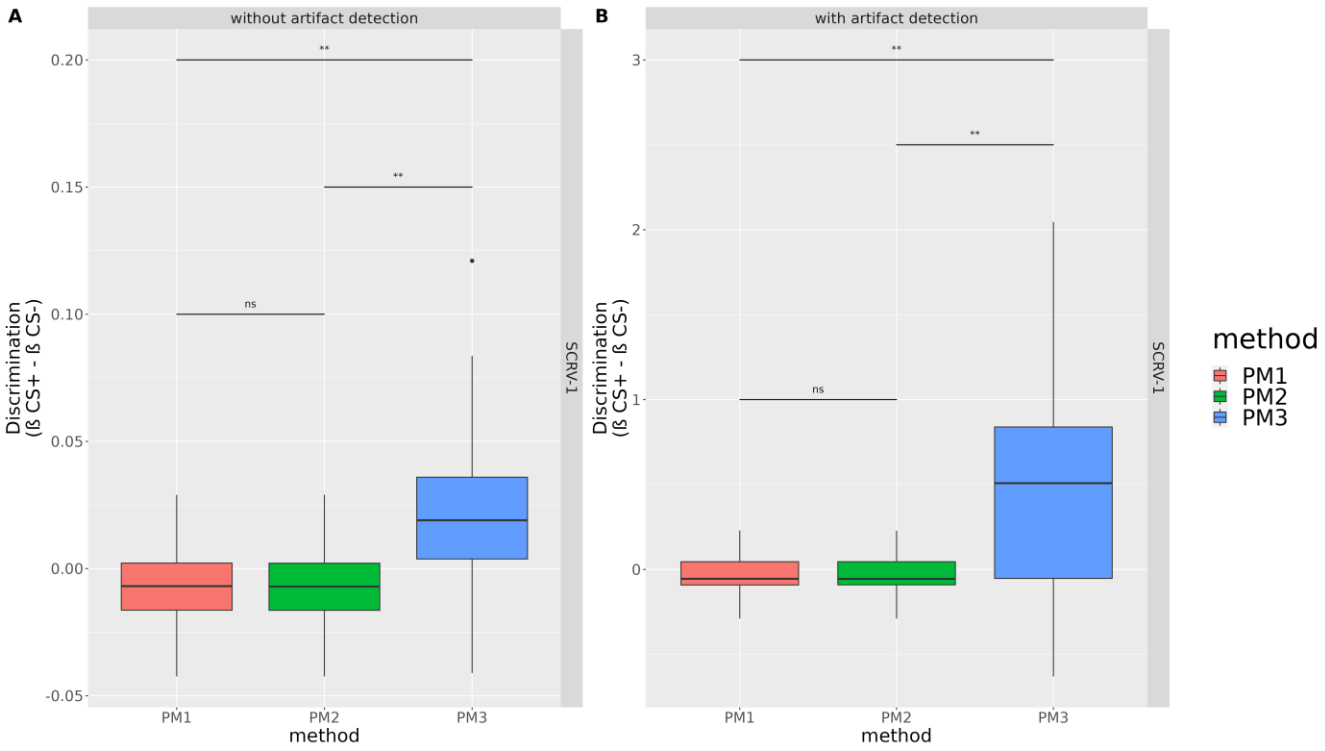
Significance levels and effect sizes Wilcoxon signed-rank test of the mean PsPM estimated parameters for aversive and neutral conditions with motion artifact detection for SCRV-4.

	PM1		PM2		PM3	
SCRV-4	Aversive (Mdn = 0.17)	Neutral (Mdn = 0.17)	Aversive (Mdn = 0.17)	Neutral (Mdn = 0.16)	Aversive (Mdn = 0.12)	Neutral (Mdn = 0.11)
	$p = 0.68$ $r = 0.08$		$p = 0.80$ $r = 0.04$		$p = 0.17$ $r = 0.24$	

Note. None of the PM had a significant difference between the mean aversive and neutral parameter estimate. This result was different from the results of the original article, this might be due to the fact that the preprocessing was different and that the mean parameter estimate had been computed differently

Supplementary Figure 8

PsPM based discrimination of mean aversive and mean neutral parameter estimate for SCRV-1 without and with motion artifact detection



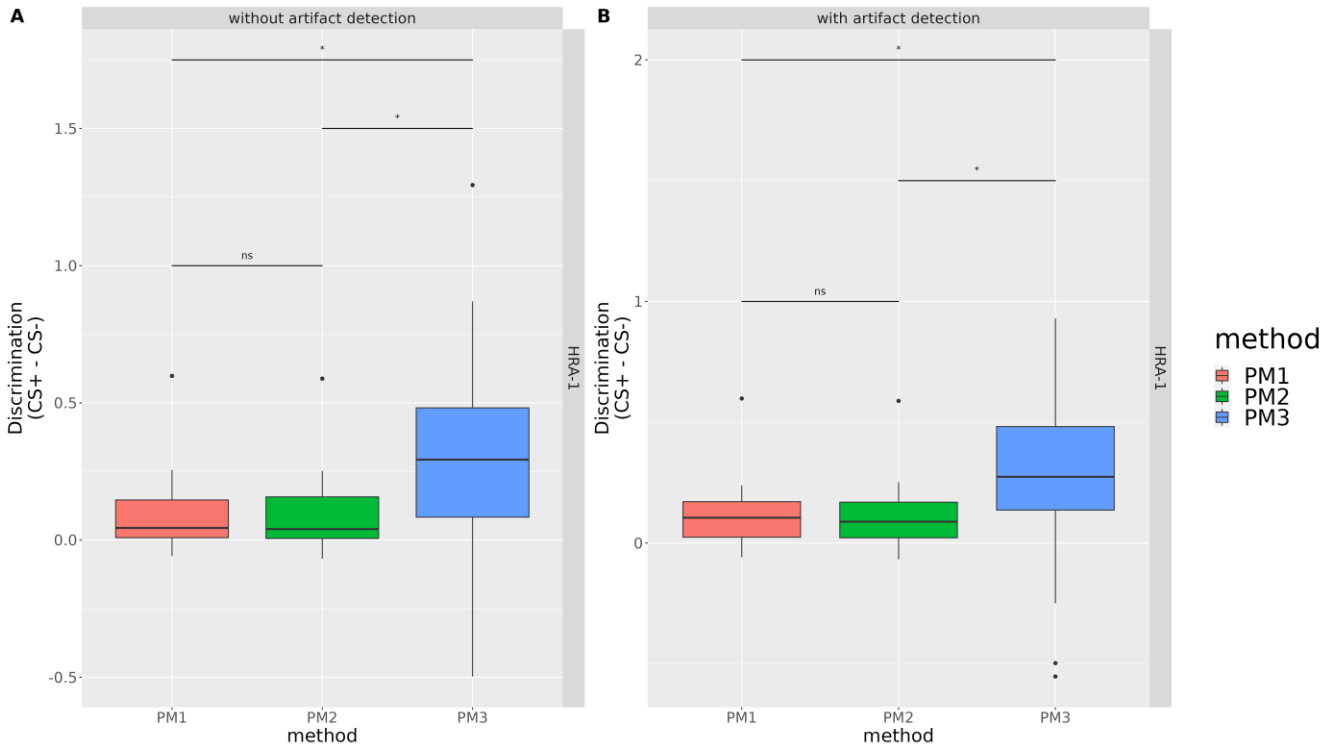
Note. The difference between all the mean aversive parameter estimate and mean neutral parameter estimate for each PM was computed, then a Wilcoxon signed-rank test was computed to see if there was significant difference in the discrimination capabilities of each PM. For both with and without motion artifact detection, only PM3 vs. PM1 and PM3 vs. PM2 had a significant difference. The differences in the scale of the graphics were due to the fact that each uses different preprocessing combinations which result in different scale.

* : $p < 0.05$

** : $p < 0.005$

Supplementary Figure 9

PsPM based discrimination for HRA-1 without and with motion artifact detection



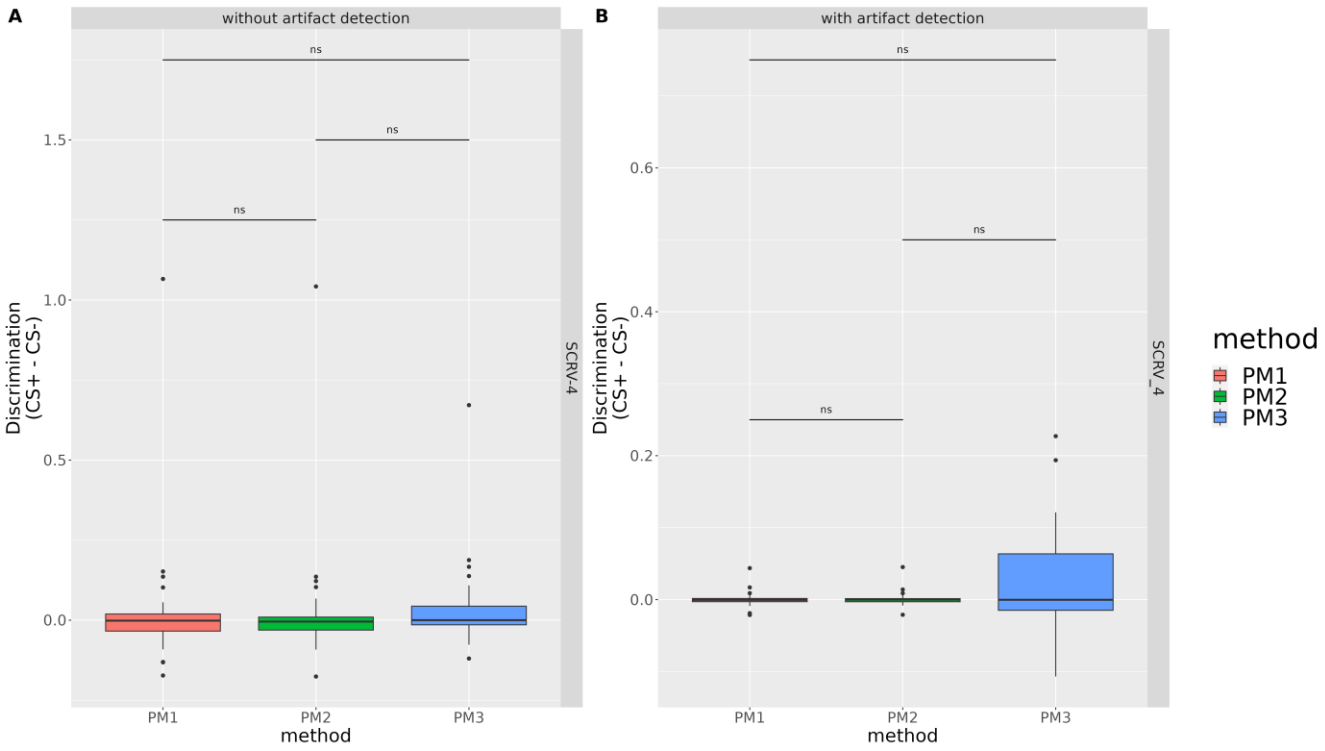
Note. The difference between all the mean aversive parameter estimate and mean neutral parameter estimate for each PM was computed, then a Wilcoxon signed-rank test was computed to see if there was significant difference in the discrimination capabilities of each PM. For both with and without motion artifact detection, only PM3 vs. PM1 and PM3 vs. PM2 had a significant difference. The differences in the scale of the graphics were due to the fact that each uses different preprocessing combinations which result in different scale.

* : $p < 0.05$

** : $p < 0.005$

Supplementary Figure 10

PsPM based discrimination for SCRV-4 without and with motion artifact detection



Note. The difference between all the mean aversive parameter estimate and mean neutral parameter estimate for each PM was computed, then a Wilcoxon signed-rank test was computed to see if there was significant difference in the discrimination capabilities of each PM. For both with and without motion artifact detection, there was no significant difference between the different PM. The differences in the scale of the graphics were due to the fact that each uses different preprocessing combinations which result in different scale.

* : $p < 0.05$

** : $p < 0.005$

Discussion générale

L'utilisation du prétraitement du signal EDA demeure aujourd'hui un défi, et ce malgré la publication de plusieurs recommandations (Boucsein et al., 2012; Braithwaite et al., 2013; Figner & Murphy, 2011; Privratsky et al., 2020). Ce mémoire avait comme objectif général de promouvoir l'importance du prétraitement du signal SCR en présentant des recommandations empiriquement appuyées. Pour ce faire, nous avons reproduit le plus fidèlement possible la méthodologie présentée dans l'article de Privratsky et al. (2020) tout en ajoutant certains éléments supplémentaires : une étape d'identification automatique des artefacts de mouvement, une seconde mesure opérationnaliste (l'AUC) et l'analyse des effets de différentes combinaisons de prétraitement sur plusieurs jeux de données.

Ce mémoire a permis de fournir de nouvelles données sur les effets de différentes étapes de prétraitement sur la capacité à discriminer entre les réponses associées à différentes conditions expérimentales – l'objectif de la plupart des études. Les résultats présentés dans ce mémoire confirment l'importance de l'utilisation du prétraitement dans l'étude du signal SCR (Boucsein et al., 2012; Braithwaite et al., 2013; Figner & Murphy, 2011 cités dans Privratsky et al., 2020), l'importance du filtrage et de la remise à l'échelle. En effet, les meilleures combinaisons de prétraitement identifiées par nos analyses n'incluaient jamais le non-filtrage ou l'absence de remise à l'échelle.

L'ajout d'une étape de détection automatique d'artefact ne semblait pas en général améliorer significativement la capacité à discriminer entre les réponses de deux conditions expérimentales. Cependant, il est important de noter que l'ajout de cette étape ne nuisait pas à cette discrimination. L'absence d'effet est probablement reliée au fait que les jeux de données utilisés dans ce mémoire ne comportaient pas beaucoup d'artefacts de mouvement. Il est possible que l'inclusion de cette étape produise des effets plus importants dans des jeux de données

comportant beaucoup d'artefacts. Le développement d'outils de détections automatiques permettant de réduire le temps de traitement et la subjectivité associée à l'inspection visuelle du signal –comme l'algorithme de Taylor et al. (2015) utilisé dans ce mémoire - devrait aider à promouvoir l'inclusion de cette importante étape dans le prétraitement du signal EDA. Ceci sera d'autant plus important pour le développement de paradigmes plus écologiques dans lesquels les participants pourront se mouvoir et où l'EDA sera mesurée de manière ambulatoire.

Le mémoire avait également comme objectif de tenter d'identifier la meilleure approche parmi les approches opérationnalistes et par modèle afin de quantifier les réponses SCR. Pour y arriver, nous avons tout d'abord comparé deux métriques de l'approche opérationnaliste : le PS et l'AUC. Nos résultats montrent que pour les jeux de données que nous avons utilisés, il n'y avait aucune différence statistiquement significative entre le PS et l'AUC. Toutefois, l'AUC avait en général une meilleure taille d'effet que le PS suggérant que l'AUC pourrait avoir un petit avantage sur le PS. Cette idée s'appuie d'ailleurs sur d'autres données ayant montré la supériorité de l'AUC sur le PS dans un autre jeu de données (Shukla et al., 2021). Nous avons par la suite procédé à la comparaison entre l'AUC et l'approche par modèle. Encore une fois, aucune différence statistiquement significative n'a été obtenue. Cependant, les tailles d'effet de l'approche par modèle étaient généralement plus grandes que celles de l'AUC. D'autres études avaient mentionné la supériorité de l'approche par modèle notamment en montrant que seule l'approche par modèle permettait de trouver des différences significatives entre deux conditions expérimentales (Bach, Daunizeau, et al., 2010; Bach et al., 2009). Ces études n'avaient cependant comparé l'approche par modèle qu'au PS. Ainsi, nos résultats suggèrent que cet avantage pour l'approche par modèle vs l'approche opérationnaliste pourrait être moins marqué lorsque l'on utilise l'AUC. Puisque l'approche par modèle est plus complexe à utiliser et peut nécessiter

d’importantes ressources informatiques (notamment en matière de temps d’analyse), l’utilisation de l’AUC pourrait être une alternative intéressante pour certaines équipes de recherche.

Finalement, nous voulions vérifier si une unique combinaison de prétraitement pouvait être appliquée à différents jeux de données. Nos résultats suggèrent au contraire que les combinaisons de prétraitement soient spécifiques aux jeux de données. Ces résultats rejoignent l’idée avancée par Privratsky et al. (2020) suggérant que les paramètres de prétraitement pourraient être dépendants du design expérimental (p. ex. de la durée des ITI). Ainsi, les données originales présentées dans ce mémoire s’ajoutent aux arguments en faveur de l’importance d’effectuer un prétraitement du signal SCR (Braithwaite et al., 2013; Figner & Murphy, 2011), mais précisent également que les équipes de recherche gagneraient en identifiant la meilleure combinaison à utiliser pour leurs données. En partageant le code ayant servi à faire nos analyses, nous espérons fournir une banque d’outils aux chercheurs et chercheuses afin de promouvoir et faciliter cette étape.

Contributions et retombées

Nous croyons que ce mémoire pourrait avoir plusieurs retombées. Premièrement, nous souhaitons que les résultats issus de ce mémoire éclaircissent au sein de communauté scientifique le rôle et l’importance du prétraitement du signal EDA. En lien avec cette retombée, nous visons que le code qui accompagne ce mémoire (et qui sera éventuellement disponible en ligne gratuitement) encouragera les chercheurs et chercheuses à faire des choix informés au sujet du prétraitement à utiliser dans leurs études. Deuxièmement, une autre contribution de ce mémoire concerne la présentation de données sur l’utilisation d’une étape de détection automatique d’artefacts de mouvement lors du prétraitement. Notre recommandation d’ajouter cette étape ainsi que le code y étant associé pourrait ultimement ouvrir la voie à l’adoption de paradigmes expérimentaux plus écologiques. Troisièmement, il est aussi possible que les connaissances et les

outils issus de ce mémoire puissent être mis à contribution dans d'autres domaines tels que la médecine et l'ingénierie, puisque parfois en médecine et même en génie biomédical l'EDA est utilisée. Par exemple, la compagnie Fitbit LLC qui appartient à Alphabet inc. (c.-à-d. auparavant Google LLC) a introduit en 2020 la première *smartwatch* destinée au grand publique incluant un capteur d'EDA (Kozuch, 2020). L'utilisation à grande échelle de cette technologie produira à terme une quantité importante de données et l'accès à des outils automatisés de prétraitement comme ceux développés et utilisés dans le cadre de ce mémoire pourrait aider à tirer profit de ces grands ensembles de données. Finalement sur le plan méthodologique, puisque nos résultats démontrent que l'AUC semble meilleure que le PS, ceci pourrait potentiellement encourager les chercheurs et chercheuses à adopter l'AUC comme métrique de l'approche opérationnaliste plutôt que le PS qui est la norme en ce moment. De plus, les statistiques présentées dans ce mémoire ont été calculées sur des tests non paramétriques puisque les données des différents jeux de données ne respectaient pas le postulat de normalité. À notre connaissance et parmi tous les articles étudiés dans le cadre de ce mémoire, bien que ces derniers utilisent toujours des tests paramétriques, aucun ne mentionne avoir testé la normalité de leurs données, alors qu'il est possible que les SCR ne suivent pas une distribution normale (Braithwaite et al., 2013). Nous espérons encourager les autres équipes de recherches à toujours vérifier la normalité de leur SCR et à utiliser des tests non paramétriques advenant que les transformations ne fonctionnent pas le cas échéant.

Limites

Plusieurs des limites ont déjà été mentionnées dans la discussion de l'article. Essentiellement, nous avons mentionner dans l'article que : 1) l'efficacité de l'algorithme de détection d'artefact de mouvement que nous avons décidé d'utiliser n'a pas été testée explicitement; 2) l'AUC a été calculé avec une seule approche alors que plusieurs moyens de la

calculer existent; 3) les trois jeux de données utilisés ont des designs expérimentaux assez similaires limitant la généralisation de nos résultats. En plus de ces limites, nous aurions également pu tester plus formellement la capacité à discriminer des différentes approches en comparant nos résultats (p. ex. l'AIC) à des contrastes qui seraient faits en utilisant des données aléatoires (c.-à-d. en faisant des contrastes avec des données dont on aurait randomisé la condition expérimentale entre aversive et neutre). Ceci aurait entre autres permis de tester si différentes approches de prétraitement pourraient donner lieu à des erreurs de type 1. Nous avons pris la décision de ne pas effectuer ce type d'analyse, car nous désirions reproduire la méthodologie utilisée dans Privratsky et al. (2020). De plus, nous n'avons pas obtenu les mêmes résultats que dans l'article original pour le DCM dans le cas du SCRV-4 (Bach, Daunizeau, et al., 2010). En effet, dans l'article original en utilisant cette approche, une différence significative avait été trouvée entre les conditions expérimentales alors que nous n'avons pas observé cette différence. Ceci suggère que d'autres choix dans le prétraitement que ceux que nous avons testés pourraient également améliorer la capacité à discriminer. Cette hypothèse est en adéquation avec l'idée que nous proposons que les chercheurs et chercheuses devraient tester différentes combinaisons de prétraitement (incluant, mais pas limitées à celles que nous avons testées) afin d'identifier la meilleure pour leurs jeux de données. Finalement, notre approche et nos résultats devraient être répliqués en utilisant des échantillons beaucoup plus grands, car les échantillons que nous avons utilisés ne comportaient que 20 à 32 participants. De plus, afin de pouvoir déterminer si nos résultats suggérant que les paramètres de prétraitement soient spécifiques au jeu de données analysées pourraient être expliqués par des différences dans les paramètres d'acquisition des jeux de données il aurait fallu tester pour l'ensemble des permutations possibles des différents paramètres d'acquisition de signal EDA. Toutefois, nous n'avons pas testé avec différents paramètres d'acquisition puisque les données que nous avons utilisées ne représentent

qu’une partie de l’ensemble des différentes façons d’acquérir le signal. Cette question reste cependant intéressante et l’impact des paramètres d’acquisition sur le prétraitement devrait être examiné lors de travaux futurs. De plus, il est important de noter que pour l’instant nous ne pouvons pas être certains que la spécificité des paramètres de prétraitement observée dans notre étude n’est pas due au simple hasard (relié à du bruit statistique).

Conclusion

L’utilisation du prétraitement de l’EDA reste un défi dans les domaines de la psychologie et des neurosciences cognitives (Privratsky et al., 2020) et l’étude présentée dans ce mémoire visait à fournir des recommandations et des outils à la communauté scientifique afin de promouvoir l’importance du prétraitement dans les travaux utilisant le SCR. À la lumière de nos résultats, nous croyons fortement : 1) que les pipelines de prétraitement devraient inclure une étape d’identification automatique des artefacts de mouvements; 2) que bien que l’approche par modèle semble meilleure, si l'on désire utiliser une approche opérationnaliste pour mesurer les réponses, on devrait choisir l’AUC plutôt que le PS; et 3) que le choix des combinaisons de prétraitement à utiliser devrait s’appuyer sur des tests afin de les optimiser en fonction du jeu de données à analyser. Nous espérons par ce mémoire avoir réussi à rendre le prétraitement du signal EDA plus accessible, mais surtout avoir réussi à encourager n’importe qui utilisant l’EDA à s’intéresser au prétraitement.

Références

Bach, D. R., Castegnetti, G., Korn, C. W., Gerster, S., Melinscak, F., & Moser, T. (2018). Psychophysiological modeling: Current state and future directions. *Psychophysiology*, *55*(11), e13214. https://doi.org/10.1111/psyp.13209

Bach, D. R., Daunizeau, J., Friston, K. J., & Dolan, R. J. (2010). Dynamic causal modelling of anticipatory skin conductance responses. *Biological Psychology*, *85*(1), 163–170. https://doi.org/10.1016/j.biopsycho.2010.06.007

Bach, D. R., Flandin, G., Friston, K. J., & Dolan, R. J. (2009). Time-series analysis for rapid event-related skin conductance responses. *Journal of Neuroscience Methods*, *184*(2), 224–234. https://doi.org/10.1016/j.jneumeth.2009.08.005

Bach, D. R., Flandin, G., Friston, K. J., & Dolan, R. J. (2010). Modelling event-related skin conductance responses. *International Journal of Psychophysiology*, *75*(3), 349–356. https://doi.org/10.1016/j.ijpsycho.2010.01.005

Bach, D. R., & Friston, K. J. (2013). Model-based analysis of skin conductance responses: Towards causal models in psychophysiology. *Psychophysiology*, *50*(1), 15–22. https://doi.org/10.1111/j.1469-8986.2012.01483.x

Bach, D. R., Friston, K. J., & Dolan, R. J. (2010). Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations. *International Journal of Psychophysiology*, *76*(1), 52–55. https://doi.org/10.1016/j.ijpsycho.2010.01.011

Bach, D. R., Friston, K. J., & Dolan, R. J. (2013). An improved algorithm for model-based analysis of evoked skin conductance responses. *Biological Psychology*, *94*(3), 490–497. https://doi.org/10.1016/j.biopsycho.2013.09.010

Bilodeau-Houle, A., Morand-Beaulieu, S., Bouchard, V., & Marin, M.-F. (2023). Parent–child physiological concordance predicts stronger observational fear learning in children with a less secure relationship with their parent. *Journal of Experimental Child Psychology*, *226*, 105553. https://doi.org/10.1016/j.jecp.2022.105553

Boucsein, W. (2012). *Electrodermal Activity*. Springer US. https://doi.org/10.1007/978-1-4614-1126-0

Boucsein, W., Fowles, D. C., Grimnes, S., Gershon, B.-S., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*, 1017–1034.

Braithwaite, D. J. J., Watson, D. D. G., Jones, R., & Rowe, M. (2013). A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments. *Psychophysiology*, *49*, 1017–1034.

Chaspari, T., Tsiartas, A., Duker, L. I. S., Cermak, S. A., & Narayanan, S. S. (2016). EDA-gram: Designing electrodermal activity fingerprints for visualization and feature extraction. *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 403–406. https://doi.org/10.1109/EMBC.2016.7590725

Cuve, H. C. J., Harper, J., Catmur, C., & Bird, G. (2023). Coherence and divergence in autonomic-subjective affective space. *Psychophysiology*, *n/a*(n/a), e14262. https://doi.org/10.1111/psyp.14262

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), Article 1. https://doi.org/10.1038/s41592-018-0235-4

Fahrenberg, J., Walschburger, P., Foerster, F., Myrtek, M., & Müller, W. (1983). An Evaluation of Trait, State, and Reaction Aspects of Activation Processes. *Psychophysiology*, *20*(2), 188–195. https://doi.org/10.1111/j.1469-8986.1983.tb03286.x

Figner, B., & Murphy, R. O. (2011). Using skin conductance in judgment and decision making research. In *A handbook of process tracing methods for decision research* (p. 33).

Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis

with MNE-Python. *Frontiers in Neuroscience*, *7*. https://www.frontiersin.org/articles/10.3389/fnins.2013.00267

Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., & Citi, L. (2016). cvxEDA: A Convex Optimization Approach to Electrodermal Activity Processing. *IEEE Transactions on Biomedical Engineering*, *63*(4), 797–804. https://doi.org/10.1109/TBME.2015.2474131

Harnett, N. G., Fani, N., Carter, S., Sanchez, L. D., Rowland, G. E., Davie, W. M., Guzman, C., Lebois, L. A. M., Ely, T. D., van Rooij, S. J. H., Seligowski, A. V., Winters, S., Grasser, L. R., Musey, P. I., Seamon, M. J., House, S. L., Beaudoin, F. L., An, X., Zeng, D., … Ressler, K. J. (2023). Structural inequities contribute to racial/ethnic differences in neurophysiological tone, but not threat reactivity, after trauma exposure. *Molecular Psychiatry*, 1–10. https://doi.org/10.1038/s41380-023-01971-x

Koppold, A., Kastrinogiannis, A., Kuhn, M., & Lonsdorf, T. B. (2022). Watching with Argus eyes: Characterization of emotional and physiological responding in adults exposed to childhood maltreatment and/or recent adversity. *Psychophysiology*, *n/a*(n/a), e14253. https://doi.org/10.1111/psyp.14253

Lakshmi, M. R., Prasad, D. T. V., & Prakash, D. V. C. (2014). Survey on EEG Signal Processing Methods. *International Journal of Advanced Research in Computer Science and Software Engineering*, 9.

Lang, P., Bradley, M., & Cuthbert, B. (2005). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical Report*. https://cir.nii.ac.jp/crid/1573950399053852928

Liberman, L., & Dubovi, I. (2022). The effect of the modality principle to support learning with virtual reality: An eye-tracking and electrodermal activity study. *Journal of Computer Assisted Learning*, *1*(11). https://doi.org/10.1111/jcal.12763

Lim, C. L., Rennie, C., Barry, R. J., Bahramali, H., Lazzaro, I., Manor, B., & Gordon, E. (1997). Decomposing skin conductance into tonic and phasic components. *International Journal of Psychophysiology*, *25*(2), 97–109. https://doi.org/10.1016/S0167-8760(96)00713-1

Liu, T. T. (2016). Noise contributions to the fMRI signal: An overview. *NeuroImage*, *143*, 141–151. https://doi.org/10.1016/j.neuroimage.2016.09.008

Lykken, D. T., & Venables, P. H. (1971). Direct Measurement of Skin Conductance: A Proposal for Standardization. *Psychophysiology*, *8*(5), 656–672. https://doi.org/10.1111/j.1469-8986.1971.tb00501.x

Moretta, T., Kaess, M., & Koenig, J. (2023). A comparative evaluation of resting state proxies of sympathetic and parasympathetic nervous system activity in adolescent major depression. *Journal of Neural Transmission*, *130*(2), 135–144. https://doi.org/10.1007/s00702-022-02577-3

Neumann, E., & Blanton, R. (1970). THE EARLY HISTORY OF ELECTRODERMAL RESEARCH. *Psychophysiology*, *6*(4), 453–475. https://doi.org/10.1111/j.1469-8986.1970.tb01755.x

Posada-Quintero, H. F., & Chon, K. H. (2020). Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review. *Sensors*, *20*(2), 479. https://doi.org/10.3390/s20020479

Posada-Quintero, H. F., Florian, J. P., Orjuela-Cañón, A. D., Aljama-Corrales, T., Charleston-Villalobos, S., & Chon, K. H. (2016). Power Spectral Density Analysis of Electrodermal Activity for Sympathetic Function Assessment. *Annals of Biomedical Engineering*, *44*(10), 3124–3135. https://doi.org/10.1007/s10439-016-1606-6

Privratsky, A. A., Bush, K. A., Bach, D. R., Hahn, E. M., & Cisler, J. M. (2020). Filtering and model-based analysis independently improve skin-conductance response measures in the fMRI environment: Validation in a sample of women with PTSD. *International Journal of Psychophysiology*, *158*, 86–95. https://doi.org/10.1016/j.ijpsycho.2020.09.015

Shukla, J., Barreda-Ángeles, M., Oliver, J., Nandi, G. C., & Puig, D. (2021). Feature Extraction and Selection for Emotion Recognition from Electrodermal Activity. *IEEE Transactions on Affective Computing*, *12*(4), 857–869. https://doi.org/10.1109/TAFFC.2019.2901673

Sklivanioti Greenfield, M., Wang, Y., & Msghina, M. (2022). Similarities and differences in the induction and regulation of the negative emotions fear and disgust: A functional near infrared spectroscopy study. *Scandinavian Journal of Psychology*, *63*(6), 581–593. https://doi.org/10.1111/sjop.12836

Staib, M., Castegnetti, G., & Bach, D. R. (2015). Optimising a model-based approach to inferring fear learning from skin conductance responses. *Journal of Neuroscience Methods*, *255*, 131–138. https://doi.org/10.1016/j.jneumeth.2015.08.009

Starita, F., Pirazzini, G., Ricci, G., Garofalo, S., Dalbagno, D., Degni, L. A. E., Di Pellegrino, G., Magosso, E., & Ursino, M. (2022). Theta and alpha power track the acquisition and reversal of threat predictions and correlate with skin conductance response. *Psychophysiology*, *n/a*(n/a), e14247. https://doi.org/10.1111/psyp.14247

Stefaniak, A. R., Blaxton, J. M., & Bergeman, C. S. (2021). Age Differences in Types and Perceptions of Daily Stress. *The International Journal of Aging and Human Development*, *94*(2), 215–233. https://doi.org/10.1177/00914150211001588

Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., & Picard, R. (2015). Automatic identification of artifacts in electrodermal activity data. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1934–1937. https://doi.org/10.1109/EMBC.2015.7318762

Tronstad, C., Staal, O. M., Sælid, S., & Martinsen, Ø. G. (2015). Model- based filtering for artifact and noise suppression with state estimation for electrodermal activity measurements in real time. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2750–2753. https://doi.org/10.1109/EMBC.2015.7318961

Wehrli, J. M., Xia, Y., Gerster, S., & Bach, D. R. (2022). Measuring human trace fear conditioning. *Psychophysiology*, *59*(12), e14119. https://doi.org/10.1111/psyp.14119