

UNIVERSITE D'AIX-MARSEILLE

ED 463- Sciences du Mouvement Humain

Faculté des Sciences du Sport

Institut des Sciences du Mouvement E.-J. Marey. UMR 7287 CNRS & AMU

Groupe PSA

Thèse présentée pour obtenir le grade universitaire de Docteur

Discipline : Sciences du Mouvement Humain

Spécialité : Contrôle Perceptivo-Moteur et Apprentissage

Charlotte JACOBÉ de NAUROIS

Modèles de détection et de prédiction de la somnolence au volant pour des systèmes personnalisés d'aide à la conduite

Soutenue le 27/09/2018 devant le jury :

Alice CAPLIER	Pr, Grenoble-INP	Rapporteur
André DUFOUR	Pr, Université de Strasbourg	Rapporteur
Damien LEGER	PU-PH, Centre du Sommeil de l'Hôtel-Dieu	Examinateur
Bernard CLAVERIE	Pr, Institut Polytechnique de Bordeaux	Examinateur
Jean-Louis VERCHER	DR, CNRS, Aix-Marseille Université	Directeur de thèse
Clément BOUGARD	Dr, Groupe PSA	Co-encadrant

Bonjour, moi Charlotte Jacobé de Naurois,
Vais vous présenter mes data avec émoi.
Je travaille à l'ISM et chez PSA,
Et vais révolutionner tout ça par Small Data !

Qui parmi vous a déjà dormi au volant ?
Un grand nombre d'entre vous, malheureusement.
Heureusement, étant là je change tout ça.
Attachez vos ceintures, et ne somnolez pas.

Ô mes beaux réseaux de neurones artificiels,
Mes petits pour vous, je remuerai terre et ciel :
Physiologie, comportement, sans oublier,
Performance de conduite et activité.
Je vais tout faire pour croître votre puissance.
Le maître mot de tout ça, est leur performance.

Alix Renault et Charlotte Jacobé de Naurois

Remerciements

En premier lieu, je souhaite remercier très chaleureusement mon directeur de recherche, Jean-Louis Vercher. Vous m'avez pris sous votre aile pour m'apprendre et me former à la recherche. Vos connaissances et vos précieux conseils m'ont toujours été d'une très grande aide. Je vous remercie pour la confiance que vous m'avez accordée, votre suivi tout au long de la thèse, vos retours et corrections, votre encouragement, et tous ces échanges scientifiques toujours très enrichissants. J'ai particulièrement apprécié travailler avec vous.

Je remercie également les différents encadrants du groupe PSA : Clément Bougard, Emmanuelle Diaz et Anca Stratulat. Vous avez chacun joué un rôle à un moment voulu dans le cadre de mes travaux. Anca, je te remercie pour ton dévouement pour que la thèse commence et au début de la thèse. Emmanuelle, je te remercie pour avoir pris le flambeau quand Anca est partie. Clément, ce ne fut pas facile de prendre l'encadrement de cette thèse la dernière année, elle aurait pris une tournure différente avec toi peut être dès le début. Je te remercie également pour tes conseils, et surtout, je vous remercie à tous pour vos aides administratives et matériels.

Je désire exprimer toute ma gratitude aux différents membres du jury : Alice Caplier, André Dufour, Bernard Claverie et Damien Léger

Je désire remercier Eric Berton, doyen de la Faculté des Sciences du Sport et directeur de l'Institut des Sciences du Mouvement E.-J. MAREY ; Reinoud Boutsma pour m'avoir accueilli au sein de l'école doctorale. Je remercie très chaleureusement également Nathalie Fenouil et Nathalie Roustan, pour votre bonne humeur et votre aide administrative ; ainsi qu'à l'ensemble des chercheurs, ingénieurs et doctorants du laboratoire d'accueil.

Un grand Merci bien sûr à l'équipe ICS, à Lionel Bringoux, Christophe Bourdin, Fabrice Sarlegna, Marie-Laure Milles, Patricia Romaguère. Vous m'avez tous très bien accueillie au sein de l'équipe et vous avez contribué à l'ambiance joviale au sein de l'équipe. Je vous remercie pour vos conseils scientifiques et votre sympathie. Je remercie également les stagiaires avec qui j'ai pu travailler Laurent Marrou et Loïc de Jouette.

Je désire remercier également Vincent Roussarie pour m'avoir accueillie dans l'équipe SCFH. Merci également à toutes les personnes de PSA de pour votre chaleureux accueil quand j'étais dans les locaux

de PSA : Anca, Emmanuelle, Alix, Morgane, Claire, Francine, Laure, Malisa, Alexandre, Sébastien, Philippe, Jean-Christophe, Vincent, Julien, Pierre, Benjamin, Luciano, Aïssa, Laurent, Gilles, Stéphane, Jérémy, Clément. J'adresse également un grand merci à Mihai pour son soutien des thésards PSA.

La thèse n'aura pas été pareil sans mes amis du laboratoire. Merci à vous car je n'aurai pas autant apprécié travailler à Marseille sans vous : Gwenaëlle, sans nos pauses café, nos repas ensemble, ton soutien, mon expérience à Marseille ne m'aurait pas autant plus ; comme on l'a dit, on travaillera ensemble ; Pascaline pour ta bonne humeur et ton soutien la dernière année ; Alix mon co-bureau pour m'avoir supporté dans le bureau ; Aurore, j'ai beaucoup aimé travailler avec toi, j'espère que l'on aura de nouveau l'occasion de le faire, je suis vraiment très heureuse de t'avoir rencontré au DDI. Je remercie également Elise et Victoria, j'ai beaucoup apprécié les moments passés ensemble au labo ou en dehors.

Je remercie également tous les post-docs et doctorants que j'ai pu rencontrer avec qui j'ai pu passer de bons moments ; Thomas, merci pour ton accueil et ton aide ; Nick, merci pour ta bonne humeur ; Victoria pour les supers moments passés ensemble, Nicolas, pour nos moments pâtisseries ; Penny, Colin, Maxime, Martin, Camille, Delphine, les physios Anabelle, Maxime, Mathieu et Caroline etc... Je tiens à remercier également tous les doctorants que j'ai croisé tout au long de mon cursus et avec qui j'ai pu sympathiser que ce soit au laboratoire ou en formation.

Un grand merci à l'ensemble des participants qui ont gentiment accepté de réaliser la longue expérience sans contrepartie à part un KitKat et de vous être presque endormis au volant. Sans vous, mes travaux n'auraient pas été possible.

Je tiens à remercier du fond du cœur mes parents pour votre soutien sans faille. Je remercie également Etienne, Laurence, Julien, Emma d'amour et le petit Arthur (le plus beau des bébés) pour les formidables moments en famille passée ensemble. Je remercie également mon grand-père George, le docteur de la famille, et sa compagne Michele. J'adresse également une pensée particulière à Huguette et Michel, partis pendant la thèse.

Je souhaite particulièrement remercier Michelle Fayet pour sa précieuse aide à la relecture et à la correction de mon mémoire. J'ai été très heureuse de te « connaître » de nouveau.

Mon dernier remerciement mais pas le moindre s'adresse à Camille. Ton soutien a été inconditionnel dans les méandres de la thèse. Grâce à toi j'ai pu aller jusqu'au bout du chemin.

Avant-propos

Cette thèse de doctorat a été réalisée avec l'Institut des Sciences du Mouvement Etienne-Jules Marey (ISM) et le groupe PSA dans le cadre de l'OpenLab « Automotive Motion Lab ».

Ces travaux ont mené à la rédaction de deux articles et d'un brevet. Les deux articles seront présents dans ce manuscrit.

Ces travaux ont également été présentés oralement à une conférence : Driver Distraction Inattention 2017. Différents posters ont été réalisés pour la journée de l'école doctorale des sciences du mouvement humain et des journées internes à PSA.

En parallèle de ces travaux, une étude a été réalisée sur la reprise en main, ces travaux seront présentés à DDI2018 et DSC2018.

Table des matières

Remerciements	2
Avant-propos	4
Table des illustrations	8
Chapitre 1 : Introduction Générale	11
1.1 Contexte général et problématique	13
1.2 Objectifs et présentation.....	15
Chapitre 2 : Cadre théorique	17
2.1 La somnolence au volant.....	19
2.1.1 État dégradé en conduite et somnolence	19
2.1.2 Pourquoi évaluer la somnolence au volant ?	22
2.1.3 Quelles sont les ressources pour évaluer la somnolence au volant ?	24
2.2 Facteurs influençant la somnolence.....	25
2.2.1 Insuffisance et qualité de sommeil	26
2.2.2 Horaire de conduite	26
2.2.3 Durée de conduite	27
2.2.4 Environnement de conduite	28
2.2.5 Consommation de médicaments et autres substances	29
2.2.6 Trait de personnalité	30
2.2.7 Créneau d'âge	31
2.3 Indicateurs pour estimer la somnolence au volant.....	32
2.3.1 Evaluations subjectives	32
2.3.2 Mesures cognitives ou tests de performance	35
2.3.3 Mesures physiologiques	36
2.3.4 Mesures sensorimotrices	43
2.3.5 Mesures véhicules	46
2.4 Véritable état du conducteur	48
2.5 Nécessité d'avoir une approche hybride.....	48
2.6 Modélisation à l'aide d'apprentissage automatique	50
2.6.1 Modèles de classification et de régression : définitions	51
2.6.2 Étapes de développement d'un modèle	52
2.6.3 Risque de sous et sur-apprentissages	53
2.6.4 Division des jeux de données	55
2.6.5 Application à la détection et la prédiction de la somnolence	59
2.6.6 Une méthode d'apprentissage automatique : les réseaux de neurones artificiels	62
2.7 Synthèse du chapitre.....	66
Chapitre 3 : Méthodologie – Recueil de données	69
3.1 Participants.....	71
3.2 Protocole mis en place	72
3.3 Détermination du niveau de somnolence.....	75
3.4 Analyse de la dynamique de la somnolence	76
Chapitre 4 : Dynamique de la somnolence	79
4.1 Résultats	81
4.1.1 Evolution du niveau de somnolence au cours de l'expérimentation	81
4.1.2 Variables reflétant la dynamique de la somnolence	82

4.2	Discussion	85
4.2.1	Variabilité inter-individuelle	85
4.2.2	Effet lié au contexte	86
4.3	Conclusion	86
Chapitre 5 : Méthodologie de la modélisation de la somnolence		89
5.1	Entrées des modèles : création des variables	91
5.1.1	Filtrage et nettoyage des signaux	91
5.1.2	Création des variables agrégées (transformation)	92
5.2	Modélisation : exploitation d'un réseau de neurones artificiels	94
5.2.1	Méthode d'optimisation choisie : algorithme de Levenberg-Marquardt	95
5.2.2	Critères d'arrêt de l'entraînement des réseaux de neurones artificiels	96
5.3	Sortie des modèles à estimer : détection et prédiction de la somnolence.....	97
5.3.1	Niveau de la somnolence au cours de l'expérimentation	97
5.3.2	Prédiction du temps restant avant le niveau de somnolence critique	97
Chapitre 6 : Détection et prédiction de la somnolence au volant en utilisant des réseaux de neurones artificiels		99
6.1	Division des jeux de données	101
6.2	Sources d'information du jeu de données	102
6.3	Problématique du premier article	103
6.4	Detection and prediction of driver drowsiness using artificial neural network models	104
6.4.1	Abstract	104
6.4.2	Introduction	105
6.4.3	Materials and methods	108
6.4.4	Results	113
6.4.5	Discussion	120
6.4.6	Conclusion	124
6.5	Synthèse des résultats.....	125
Chapitre 7 : Généralisation des modèles de détection et prédiction de la somnolence sur des nouveaux conducteurs		127
7.1	Division des jeux de données	129
7.2	Sources du jeu de données	130
7.3	Problématique de ce chapitre	130
7.4	Résultats avec validation croisée inter-.....	131
7.4.1	Détection du niveau de somnolence	131
7.4.2	Prédiction du temps restant avant un niveau dégradé	132
7.5	Discussion	133
7.6	Conclusion	135
Chapitre 8 : Adaptation personnalisée des modèles de détection et prédiction de la somnolence		137
8.1	Apprentissage adaptatif	139
8.2	Division des jeux de données	142
8.3	Sources du jeu de données	143
8.4	Problématique du second article	143
8.5	Adapting artificial neural networks to a specific driver enhances detection and prediction of drowsiness	144
8.5.1	Abstract	144
8.5.2	Introduction	145
8.5.3	Materials and methods	147

8.5.4	Results	157
8.5.5	Discussion	165
8.5.6	Conclusion	168
8.6	Synthèse des résultats du second article	170
Chapitre 9 : Discussion et conclusion générale		171
9.1	Synthèse et avancées de l'ensemble des résultats	173
9.2	Détecter l'état du conducteur de manière continue	174
9.3	Prédire la somnolence, mais pour quoi faire ?.....	178
9.4	Caractère intrusif de la modélisation sur le conducteur	179
9.5	L'ajout d'informations supplémentaires	182
9.6	Transférabilité des connaissances.....	183
9.7	Les limites de l'ensemble de ces travaux	184
9.7.1	Le nombre de participants	185
9.7.2	La vérité de terrain	185
9.7.3	Les conditions de l'expérience	186
9.7.4	Le type de modélisation choisie	186
9.8	Perspectives et futures recherches	187
9.8.1	Intégrer les contre-mesures dans les modèles et utilisation en temps réel	187
9.8.2	Prendre en compte la variabilité intra-individuelle	189
9.8.3	Prendre en compte le niveau d'automatisation du véhicule	189
Références		192

Table des illustrations

Figure 2-1 : Niveau d'éveil en fonction du niveau de performance extraite et adaptée de Reimer, Coughlin, & Mehler, (2009)	21
Figure 2-2 : Différents facteurs potentiels endogènes et exogènes pouvant être liés à la fatigue (extrait de Di Milia et al., 2011).....	25
Figure 2-3 : Variation du niveau de la vigilance selon l'heure de la journée (illustration du rythme circadien)	27
Figure 2-4 : Symboles pouvant apparaître sur la boîte d'un médicament.....	30
Figure 2-5 : Deux RED de formes différentes avec une phase de descente en rose, suivie ou non d'une phase de bas niveau (en jaune sur celle de droite), suivie d'une phase de récupération en vert (extrait de Clarion, 2009)	42
Figure 2-6 : Schématisation de la démarche suivie en apprentissage automatique pour inférer la valeur de y à partir de x. D'après Breiman, (2001).....	52
Figure 2-7 : Pour un problème de régression, illustration du compromis biais-variance : modèle trop simple à gauche (polynôme d'ordre 1), modèle trop complexe à droite (polynôme d'ordre 7) et un modèle intermédiaire entre les deux (polynôme d'ordre 3).	55
Figure 2-8 : Division du jeu de données en jeu de données d'entraînement, de validation et de test. 56	
Figure 2-9 : Une cross-validation à 4 plis : Chaque point appartient à 1 des 4 jeux de test et aux 4 autres jeux d'entraînement.....	57
Figure 2-10 : La stratégie d'exclusion d'une donnée	57
Figure 2-11 : Validation croisée inter- individuelle	58
Figure 2-12 : Mise en correspondance neurone biologique / neurone artificiel	63
Figure 2-13 : Exemple d'un réseau de neurones artificiels à une couche cachée (I_1 à I_n correspond aux entrées du neurone).....	64
Figure 2-14 : Conclusion de la revue de littérature et méthodologie mise en place	67
Figure 2-15 : Les enjeux de ce travail de thèse	67
Figure 3-1 : Dispositif et différentes données enregistrées	73
Figure 3-2 : Vue des trois écrans à l'entrée de la ville et le tableau de bord	74
Figure 4-1 : Moyenne et écart-type du niveau de somnolence au cours du temps	81
Figure 4-2 : Dendrogramme - Arbre de classification hiérarchique ascendant	82
Figure 4-3 : Répartition des participants selon le niveau de somnolence maximum atteint et le temps de conduite restant avant le niveau critique (en minute). Colorisation des participants selon la classification résultant du dendrogramme et k-moyenne (Fig. 4-2).....	83
Figure 4-4 : Moyenne et écart-type de la fréquence de conduite regroupés par groupes.	84
Figure 6-1 : Division des jeux de données pour la détection et prédiction de la somnolence au volant en utilisant des réseaux de neurones artificiels.....	101
Figure 6-2 : Sources de données pour la détection et prédiction de la somnolence au volant en utilisant des réseaux de neurones artificiels.....	102
Figure 6-3: frequency histogram of error distribution (left panel) and correlation (right panel) between real and estimated state, for a model trained with behavioral dataset, driving time and participant information.	116
Figure 6-4: frequency histogram of error distribution (left panel) and correlation (right panel) between real and estimated state, for a model trained with behavioral, car and physiological datasets.	116

Figure 6-5: frequency histogram of error distribution (left panel) and correlation (right panel) between real and estimated times, for a model trained with behavioral dataset, driving time and participant information.	119
Figure 6-6: frequency histogram of error distribution (left panel) and correlation (right panel) between real and estimated times, for a model trained with behavioral, car, and physiological datasets.	119
Figure 7-1 : Division des jeux de données (entraînement / Validation / test) pour la généralisation des modèles de détection et prédiction de la somnolence au volant sur un nouveau conducteur	129
Figure 7-2 : Source de données pour la généralisation des modèles de détection et prédiction de la somnolence au volant sur un nouveau conducteur.....	130
Figure 7-3 : REQM du jeu de validation et du jeu de test pour les différentes sources de données pour la détection du niveau de somnolence. Les étoiles représentent le degré de significativité (NS: $p>0,05$; *: $p<0,05$; **: $p<0,01$; ***: $p<0,001$).	132
Figure 7-4 : REQM du jeu de validation et du jeu de test pour les différentes sources de données pour la détection du niveau de somnolence. Les étoiles représentent le degré de significativité (NS: $p>0,05$; *: $p<0,05$; **: $p<0,01$; ***: $p<0,001$).	133
Figure 8-1 : Schéma explicatif entre Machine Learning Traditionnel et Transfer Learning (adapté de Pan & Yang,(2010)).....	140
Figure 8-2 : Division des jeux de données.....	143
Figure 8-3 : Source de données pour l’adaptation des modèles de détection et prédiction de la somnolence au volant sur un nouveau conducteur.....	143
Figure 8-4 : The static driving simulator by Oktal®. A represents the road scene displayed on the 3 video screens. B represents the dashboard. C is the hardware faceLAB®. D are two of the three electrodes used for ECG, E is the respiratory belt. F are the electrodes for EDA (not used in this study because of important signal loss).....	149
Figure 8-5: Diagram of the scenario with different types of road and the associated traffic	150
Figure 8-6: The training method used to adapt the ANNs. Cylinders represent different datasets. The small numbered circles represent the step of the process defined in part 2.7 Methods for adaptive learning with the ANN. Rectangles represent steps in the process.....	155
Figure 8-7: Root mean square error (RMSE) and standard error of the detected level of drowsiness, based on the AdANN-validation dataset, for different sources of information, before and after adaptation. Stars represent the level of significance (NS: $p>.05$; *: $p<.05$; **: $p<.01$; ***: $p<.001$)..	158
Figure 8-8: mean of RMSE and SD of the predicted time of occurrence of drowsiness level 1.5, based on AdANN-validation dataset, for different sources of information, before and after adaptation. Stars represent the level of significance of the difference in means (NS: $p>.05$; *: $p<.05$; **: $p<.01$; ***: $p<.001$).....	159
Figure 8-9: Detection: Mean and SD of RMSE for the different datasets (ANN-training set, ANN-validation set, AdANN-training set, AdANN-validation set) as a function of amount of data (τ , in minutes) used to adapt the AdANN-training dataset.	160
Figure 8-10: P values for RMSE means compared between the different τ for the Ad-ANN-validation dataset after adaptation.	161
Figure 8-11: Prediction: Mean and SD of RMSE for the different datasets (ANN-training set, ANN-validation set, AdANN-training set, AdANN-validation set) as a function of amount of data (τ , in minutes) used to adapt the AdANN-training dataset.	162
Figure 8-12: P-value for RMSE mean comparing each τ with each other τ for the AD-ANN-validation dataset after adaptation.	163
Figure 8-13: Detection of level of drowsiness: RMSE mean and SD for participants used for the adaptation (A) and other participants (B) never previously encountered by the ANN, before and after	

adaptation with the τ first data recorded on participant A. Best τ is used here. Stars represent the level of significance (NS: $p > .05$; *: $p < .05$; **: $p < .01$; ***: $p < .001$).	164
Figure 8-14: Prediction of impaired driver state: RMSE mean and SD for participants used for the adaptation (A) and other participants never previously encountered (B), before and after adaptation with the τ first data recorded on participant A. Stars represent the level of significance (NS: $p > .05$; *: $p < .05$; **: $p < .01$; ***: $p < .001$).	165
Figure 9-1 : Les différents niveaux d'automatisation.....	177
Tableau 2-1 : La Karolinska Sleepiness Scale.....	33
Tableau 2-2 : Bande de Fréquence EEG, (extrait de Larue, 2010).....	38
Tableau 2-3 : Les différentes mesures véhicules.....	46
Tableau 3-1 : Informations recueillies sur les participants	72
Tableau 3-2 : caractéristiques du parcours.....	74
Tableau 3-3 : Observer Rating of Drowsiness	76
Tableau 4-1 : Tableau de contingence des groupes et de la variable catégorielle « Déjà dormi au volant ».....	85
Tableau 5-1 : Sources de données et fréquences d'échantillonnage	92
Tableau 5-2 : Variables extraites par systèmes d'enregistrement.....	93
Tableau 6-0-1 : All the variables (grouped by source of information, in column) computed for each participant for each minute, used as input for ANNs.....	112
Tableau 6-0-2: Model performance in detecting drowsiness level for the testing dataset: mean square error (MSE), standard deviation, (STD), according to dataset used, with (1) or without (0) driving time, with (1) or without (0) participant information. The worst performance (highest MSE) is highlighted with a * while the best performance (lowest MSE) is highlighted with a #.....	114
Tableau 6-3: Performance of the model in predicting drowsiness level with the testing dataset: mean square error (MSE), standard deviation (STD), according to whether dataset is used with (1) or without driving time (0), participant information, and source of recorded information. The * symbol indicates the worst performance and the # symbol the best performance.	118

Chapitre 1 : Introduction Générale

1.1 Contexte général et problématique

De nos jours, il n'est pas encore possible de s'endormir paisiblement et sereinement dans un véhicule automobile en mouvement, surtout quand on est supposé tenir le volant ! En effet, l'arrivée d'un monde où seuls des véhicules 100% autonomes et 100% sûrs n'est pas encore pour demain. La conduite reste donc encore très ancrée dans nos habitudes quotidiennes. Même si conduire peut apparaître comme une activité ordinaire, routinière, monotone et peu compliquée, car réalisée par une grande majorité de la population, conduire un véhicule est en réalité une activité multitâche, complexe, potentiellement à risque. Afin de garantir le meilleur niveau de performance, elle nécessite une forte mobilisation de ressources physiologiques et cognitives de la part des conducteurs. Ainsi, tout conducteur doit rester, à tout moment, attentif par rapport à l'environnement routier. Afin rester attentif sur la route, des préconisations d'un arrêt toutes les deux heures ou de renoncer à conduire dans un état de fatigue avancé ne sont pourtant pas suffisantes, comme le montrent, hélas, les statistiques d'accidentologie (Philip et al., 2010).¹.

Durant la phase de conduite, et bien que cela soit déconseillé, voire interdit, le conducteur peut penser à autre chose (« *mind wandering* » (Christoff, 2012)), être distrait par son téléphone, se laisser aller et/ou commencer à somnoler. Ces états sont bien évidemment susceptibles d'amener une dégradation des performances de conduite. C'est pourquoi l'estimation en temps réel du véritable état opérationnel du conducteur est un challenge important. Cet état, au sens très large, subit l'influence de nombreux facteurs externes (fluctuation circadienne, fatigue, monotonie, variation du trafic, prise de médicaments, privation de sommeil (Di Milia et al., 2011)). De plus, la variation de cet état du conducteur est propre à chacun (expertise, vieillissement, âge, etc.) et varie en fonction de son état du moment (prise de médicaments, privation de sommeil, stress, etc.) et de l'environnement de conduite (route monotone, trafic, etc.). Par ailleurs, monitorer le conducteur est une manière de récolter des indices de son état général et devient donc une orientation de recherche dont l'objectif est de lutter contre la dégradation de cet état. Or, dans un véhicule routier, il est nécessaire d'utiliser des informations pouvant être enregistrées facilement et surtout de manière non invasive et non intrusive. Les mesures les plus utilisées dans ce domaine concernent, selon la littérature, les mesures physiologiques, les mesures sensorimotrices et les mesures issues du véhicule. De plus, il a été noté une grande variabilité des comportements et des dynamiques de somnolence pour chaque individu.

1

<http://www.autoroutes.fr/FCKeditor/UserFiles/File/Chiffres%20cl%C3%A9s%20accidents%20mortels%202016.pdf>

Pour contrer ce problème de variabilité et fusionner les données, des modèles d'algorithmes d'apprentissage automatique, issus de l'intelligence artificielle, ont souvent été utilisés. Ils permettent de créer une fonction complexe apte à fournir, en sortie, une information relative à l'état du conducteur grâce à des entrées provenant de différents indicateurs, mesurés sur le conducteur. Cependant, ces fonctions, pour être définies, s'établissent et se construisent avec un nombre restreint de conducteurs en raison de l'impossibilité d'effectuer une étude de ce type à grande échelle. Il est par conséquent nécessaire de savoir si ce modèle fonctionnera sur une autre population, c'est-à-dire avec des données issues de conducteurs que le modèle n'a jamais observés. En effet, la conduite est une activité réalisée par une multitude d'individus différents et le modèle abouti devra être transférable à tout nouveau conducteur.

Cette thèse s'intéresse à l'état du conducteur et plus particulièrement à la détection et prédiction d'un état « dégradé » tel que la somnolence lors de la conduite automobile. Dans le cadre de cette étude, le terme de détection fera appel au fait d'estimer à un instant « t » le niveau ou degré de somnolence du conducteur. Quant au terme de prédiction, celui-ci se réfèrera à une estimation temporelle du moment où l'état défini comme dégradé est susceptible de survenir. Plus concrètement, si l'état dégradé arrive à un instant « t' » supérieur à « t », à l'instant « t », l'objectif de la prédiction de somnolence est d'estimer la différence entre « t' et t ». Au-delà de traiter de cette prédiction sur un groupe donné, cette thèse, afin d'envisager un futur modèle généralisable, analysera la notion de transfert et d'adaptation des connaissances, apprises quant à la somnolence, à un nouvel individu.

Il est important de noter que du fait du contexte industriel dans lequel se situe cette thèse, nous avons dû faire des choix, guidés plus par un souci d'optimalité et de faisabilité à terme dans un véhicule que par la recherche de la meilleure solution théorique. Ceci nous a amenés à écarter d'emblée certaines méthodes, outils ou même approches, ayant fait leurs preuves dans la littérature scientifique, mais dont l'applicabilité pratique semble à ce jour illusoire.

1.2 Objectifs et présentation

La somnolence peut être influencée par différents facteurs internes ou externes à l'individu. Ce changement d'état, lié aux fluctuations du niveau de somnolence peut se traduire par des modifications de divers indicateurs. Il est donc possible d'évaluer ce niveau de somnolence à partir de ces indicateurs, et ce, par différents moyens. L'un de ces moyens, qui sera le support de ce travail, est appelé apprentissage automatique ou « machine learning ». Ces notions soulignées ici seront définies et détaillées dans le **chapitre 2**.

Le rationnel qui a guidé le choix des indicateurs retenus ainsi que la méthodologie ayant servi à recueillir les données de cette thèse sera décrite dans le **chapitre 3**. Des résultats sur la dynamique de la somnolence seront ensuite présentés et discutés dans le **chapitre 4**. Ils seront suivis de la méthodologie définissant la modélisation de la détection et de la prédiction de la somnolence dans le **chapitre 5**.

Ensuite, les premiers résultats amenant à une première publication seront présentés dans le **chapitre 6**. Celle-ci met en avant deux aspects : la faisabilité de la prédiction de l'arrivée de la somnolence et la meilleure précision des modèles avec des informations contextuelles telles que le temps de conduite et des informations personnelles. Néanmoins, il est à noter que les modèles développés et testés ne permettent pas de savoir, à cette étape, s'ils fonctionneront sur un nouvel individu, non vu par le modèle lors de la phase d'entraînement de celui-ci.

Dans le **chapitre 7**, les modèles développés précédemment seront alors testés sur un nouvel individu qui n'a pas, cette fois-ci, servi à la modélisation. Cependant, dans ce cas, il est à noter que les performances de généralisation se dégradent très fortement, c'est-à-dire, que les modèles ne sont pas performants sur des nouveaux individus n'ayant pas servi à la définition et à l'entraînement du modèle.

Ce problème nous amènera, dans le **chapitre 8**, à proposer une nouvelle méthodologie : l'utilisation de modèles adaptifs. Le modèle général est entraîné sur un groupe d'individus, il sera ensuite personnalisé sur un nouveau conducteur, en utilisant les premières données de conduite. Ce modèle permettra d'améliorer de manière significative les performances de détection et de prédiction de la somnolence au volant pour ce nouveau conducteur.

Enfin une discussion générale, une conclusion et des perspectives seront développées dans le **chapitres 9**.

Chapitre 2 : Cadre théorique

2.1 La somnolence au volant

2.1.1 État dégradé en conduite et somnolence

La conduite d'un véhicule impose que la personne au volant soit apte à effectuer cette tâche tout au long de son parcours. Le conducteur doit donc rester constamment maître de son véhicule, mais ses capacités à conduire peuvent être altérées si celui-ci se trouve dans un état défini comme « dégradé ». Cependant, il est difficilement envisageable de compter seulement sur le ressenti du conducteur concernant ces capacités à conduire. En effet, son état de somnolence (tout comme l'alcoolémie) peuvent affecter ses capacités de jugement ou de conduite (Dong, Hu, Uchimura, & Murayama, 2011). Ainsi, quantifier, détecter ou même prédire un état dégradé du conducteur de manière fiable est un challenge qui a généré de nombreuses études d'un point de vue aussi bien fondamental qu'appliqué dans le but de réduire le nombre d'accidents causés par un état dégradé du conducteur. Néanmoins, la question derrière l'état du conducteur est un thème complexe et pas toujours clairement défini. Déterminer l'état du conducteur est donc une thématique de recherche encore très présente. De plus, avec l'arrivée des véhicules autonomes, la question de surveiller le conducteur afin de s'assurer de ses capacités à superviser la route ou à reprendre la main si nécessaire est de plus en plus présente.

Concernant cette problématique de la supervision ou surveillance de l'état du conducteur, il s'agit tout d'abord de définir ce qu'on entend par état dégradé, puis de savoir à l'aide de quelles informations pourra-t-on quantifier cet état du conducteur, le mesurer et même le prédire, et enfin de savoir à l'aide de quels moyens ces informations seront utilisées. De plus, tout système doit être fiable (robuste), produire un minimum d'erreurs (positives ou négatives) et fonctionner sur l'ensemble des conducteurs. Cette question sera abordée sous un angle interdisciplinaire : les neurosciences et la physiologie humaine pour caractériser l'état du conducteur mais également l'informatique pour la création de modèles.

Dans ce chapitre nous mettrons tout d'abord en valeur quelles sont les différentes formes d'état dégradé du conducteur en lien avec l'inattention et nous définirons la somnolence, l'état dégradé choisi pour ces travaux. Après une mise en exergue de l'intérêt de mesurer la somnolence au volant, différents facteurs et mesures permettant de caractériser la somnolence de manière indirecte seront détaillés. Enfin, une modélisation à l'aide d'apprentissage automatique étant utilisée dans ces travaux, nous présenterons ce qu'est l'apprentissage automatique et comment l'appliquer à la modélisation de la somnolence au volant.

Derrière la notion d'état dégradé pouvant impacter négativement les performances de conduite se cachent divers états cognitifs tels que la baisse d'attention, la charge de travail, le stress, les émotions, l'hypovigilance, la somnolence, la distraction (Hockey, 2003). Selon la revue de littérature de Dong, Hu, Uchimura, & Murayama, (2011), il existe différentes formes d'inattention telles que les distractions :

- cognitives : lorsque le conducteur pense à autre chose qu'à la tâche de conduite (« mind wandering » en anglais) ;
- sonores : lorsque le conducteur est perturbé par un bruit ;
- biomécaniques : lorsque le conducteur exerce une autre tâche que conduire ;
- visuelles : lorsque le conducteur regarde autre chose que la route ;
- la fatigue physique ;
- la fatigue mentale.

Tous ces états sont en relation et peuvent être imbriqués. Le terme de « *Operator functional State* », ou état fonctionnel de l'opérateur, a été utilisé dans la littérature pour fédérer l'ensemble des états en lien avec la performance d'exécution de la tâche opérationnelle. Selon les études, divers termes peuvent s'y rapporter : la conscience de la situation (*situation awareness*), la charge de travail (*mental workload*) et la fatigue mentale (Yin & Zhang, 2017). Il existe des relations entre ces différents concepts et les performances de conduite. La présente thèse est particulièrement centrée sur la somnolence et, plus précisément, sur l'estimation et la prédiction du niveau de somnolence.

Nous distinguerons, à ce stade, la fatigue physique ou neuromusculaire (caractérisée quant à elle par une réduction de la force musculaire) de la fatigue mentale. Cette dernière est un processus graduel et cumulatif pouvant se traduire par une diminution des efforts, de l'attention et de la vigilance. Au cours de ce manuscrit, lorsque le terme fatigue sera utilisé, cette terminologie fera donc référence à la fatigue mentale et à la fatigue nerveuse centrale plutôt qu'à la fatigue physique et musculaire. Cette forme de fatigue centrale est la plus dangereuse car plus difficile à détecter par l'opérateur, et pouvant amener la somnolence et augmenter la probabilité d'accidents (Croo, Bandmann, Mackay, Rumar, & Vollenhoven, 2001 dans Dong, Hu, Uchimura, & Murayama, 2011).

Généralement, on décrit deux états extrêmes de vigilance, définissant les limites d'un continuum : d'une part le sommeil, en tant que période de repos, et d'autre part, l'état de veille ou de vigilance active, état d'activation de l'organisme permettant la meilleure efficacité comportementale. La vigilance est, quant à elle, le processus qui gère, ajuste et soutient l'activité de traitement de l'information (les ressources attentionnelles) (Wickens & Hollands, 2000). La performance est alors régulée par trois mécanismes énergétiques : l'éveil, l'effort et l'activation (Pribram & McGuinness, 1975). Cette dernière reflète classiquement « le degré d'activité de l'organisme » selon un continuum allant du sommeil à la veille, ou de la veille attentive au stress. De plus, d'après le modèle de Yerkes

Dodson (Cohen, 2011), la veille attentive coïncide au moment défini comme étant celui du maximum d'efficacité mentale (performance optimale). Il ne peut être observable ni pour des niveaux faibles d'activation ni pour des niveaux élevés. L'efficacité attentionnelle accompagne en fait un niveau d'activation optimum et non maximum.

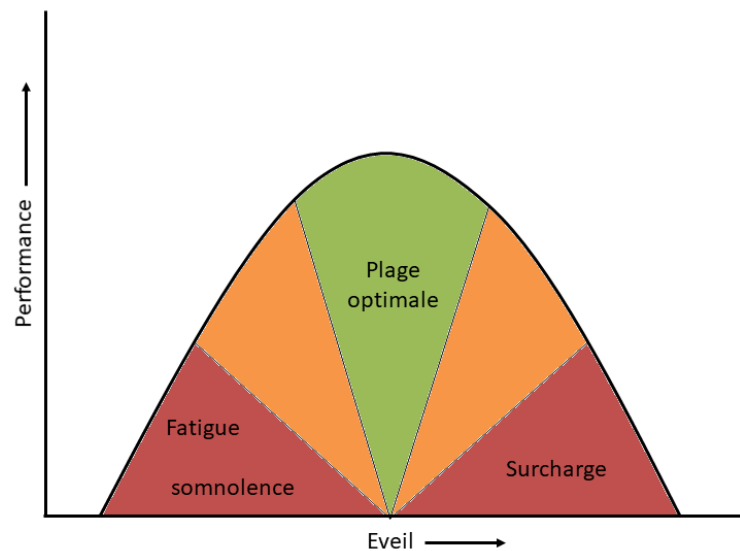


Figure 2-1 : Niveau d'éveil en fonction du niveau de performance extraite et adaptée de Reimer, Coughlin, & Mehler, (2009)

La complexité de ces états et la recherche encore très active sur la « somnolence » peuvent s'expliquer par le fait que celle-ci peut être vue comme un événement multidimensionnel et complexe, il n'existe donc pas de consensus concernant sa définition. En anglais, de nombreux termes font référence à ce concept : *sleepiness, drowsiness, somnolence, sleep propensity, fatigue, tiredness, the ability to fall asleep, sleepability, the ability to stay awake, objective sleepiness, subjective sleepiness* ou *manifest sleepiness* (M. Johns, 1998). Même si fatigue et somnolence sont quelques fois employés l'un pour l'autre, nous voulons distinguer l'état de fatigue mentale de la somnolence. La fatigue est un processus cumulatif et graduel associé à une sensation d'affaiblissement physique ou moral (Grandjean 1979). Elle peut résulter d'un travail physique ou mental soutenu. Quant à la somnolence, elle a été définie par Carskadon & Dement, (1982) comme la tendance à s'endormir (« *the tendency to fall asleep* »). Cette définition amène à penser que la somnolence est un état intermédiaire entre l'état d'une personne alerte et réveillée et l'état de sommeil. Par conséquent, il est possible de considérer qu'il existe un ou plusieurs états intermédiaires de somnolence. Ces deux termes (fatigue et somnolence) sont parfois confondus car leurs effets peuvent se révéler très semblables : diminution de la vigilance, du temps de réaction, de la coordination motrice (Brown, 1994 ; Lyznicki, Doege, Davis

et Williams, 1998). Il est possible de distinguer la fatigue de la somnolence par le fait que la somnolence fluctue bien plus rapidement. Un conducteur peut être fatigué sans être somnolent mais des conditions qui produisent la fatigue sont susceptibles de provoquer également de la somnolence (Stutts, Wilkins, Vaughn, 1999).

Dans le cadre de ce travail, nous considérons l'état du conducteur comme un état continu allant d'un état optimal, lorsque la personne est alerte et se trouve dans des conditions optimales pour la conduite, jusqu'à très somnolent, lorsque la personne est dans un état proche du sommeil associé à des micro sommeils de quelques secondes. Dans la littérature, différents termes sont utilisés pour définir ces « états intermédiaires » : l'utilisation de préfixes comme « peu », « modérément », « très », « extrêmement » ou bien des groupes nominaux comme « *driving without awareness* » « *highway hypnose* », traduits par « conduire sans conscience de la situation » ou par « l'hypnose de l'autoroute ». Ces deux derniers termes décrivent des symptômes de somnolence se produisant quand la volonté du conducteur de maintenir les yeux ouverts surpasse le besoin physiologique de fermer les yeux. Ce phénomène est décrit pour la première fois en 1929 (Miles, 1929). Presque 20% des personnes sont amenées un jour à avoir des états d'inconscience lors de la conduite sur autoroute monotone et ces états se produisent généralement avec des évènements de micro-sommeil (Briest, Karrer-Gauß, & Schleicher, 2006).

2.1.2 Pourquoi évaluer la somnolence au volant ?

La conduite est une activité complexe car le conducteur doit se déplacer dans un environnement en perpétuelle évolution (Neboit, 1980). Il doit ainsi à la fois prendre en compte divers éléments de l'environnement en temps réel et chercher les informations pertinentes pour réagir rapidement et agir en conséquence. Il existe donc une interaction dynamique et complexe entre le conducteur, le véhicule et l'environnement. Ces opérations nécessitent de la part du conducteur des processus cognitifs complexes d'où la dangerosité de toute baisse de vigilance. Par ailleurs, l'activité de conduite réclame aussi des actions totalement automatisées mais également des activités consommatrices de ressources cognitives. Pendant la conduite, le conducteur doit en effet gérer différents processus :

- **le niveau opérationnel** concernant la régulation longitudinale et latérale en agissant sur le volant et les pédales d'accélération et freinage,
- **le niveau tactique** permettant de prendre des décisions sur les actions à mener dans les instants qui suivent,
- **le niveau stratégique** concernant la planification et le choix de l'itinéraire.

Cependant, les ressources cognitives nécessaires à la tâche de conduite peuvent dépendre de plusieurs facteurs comme le contexte par exemple. En effet, conduire sur autoroute par faible trafic ou traverser une intersection ne requièrent pas le même niveau d'attention de la part du conducteur.

La somnolence au volant est dangereuse car, au même titre que les erreurs d'inattention, elle induit des modifications du comportement du conducteur. Sur autoroute, la somnolence est le facteur numéro un d'accidents mortels, représentant près d'1/3 des accidents et est reconnue comme étant la première cause d'accident en France en 2016². Dans ce contexte, les conséquences négatives de la somnolence sont nombreuses (Wylie, Shultz, Mitler, & Mackie, 1996). La première est une baisse de l'attention et de l'activité psychophysique qui pourront entraîner une augmentation du temps de traitement de l'information et de prise de décision, et donc une augmentation du temps de réaction. Il est possible d'observer également une réponse plus variable et moins efficace du conducteur face à des événements inattendus et donc une baisse de sa motivation et de ses performances de conduite dans ce contexte. Quand un conducteur devient somnolent, les symptômes suivants peuvent être observés (Horne & Reyner, 1999):

- bâillements répétés ;
- confusion des pensées devenues brumeuses ;
- sensation d'être déprimé et irritable ;
- réactions et réponses plus lentes ;
- rêveries ;
- difficulté à garder les yeux ouverts assortie d'une sensation de brûlure des yeux ;
- contrôle du volant plus lent ;
- difficulté à maintenir son niveau de concentration ;
- balancement de la tête et/ou du corps ;
- véhicule errant sur la route ou allant dans une autre voie ;
- variations importantes de l'angle au volant ;
- respiration devenant peu profonde ;
- variation du rythme cardiaque ;
- inattention à l'environnement (piétons, panneaux, etc.) ;
- difficulté à stabiliser sa vitesse ;
- difficulté pour maintenir la trajectoire ;
- changement de voie impromptu.

Les conducteurs peuvent présenter certains de ces symptômes, pas d'autres, et ce à des degrés variables.

2

<http://www.autoroutes.fr/FCKeditor/UserFiles/File/Chiffres%20cl%C3%A9s%20accidents%20mortels%202016.pdf>

2.1.3 Quelles sont les ressources pour évaluer la somnolence au volant ?

Tout conducteur n'est pas amené à avoir un épisode de somnolence lorsqu'il conduit. Ainsi, afin de caractériser l'apparition de la somnolence au volant, nous pouvons nous intéresser d'une part, à différents facteurs influençant cette dernière et d'autre part, à différents marqueurs de la somnolence au volant. Parmi les multiples facteurs d'influence de la somnolence (Dababneh & Gindy, 2015), nous nous attacherons à identifier les principaux, afin de mieux cerner leur impact direct sur les capacités de conduite, et de s'en affranchir au cours des passations expérimentales. D'autre part, lorsque l'état du conducteur varie, cela peut se traduire par des variations sur différents indicateurs ou mesures pouvant être appréhendés pour estimer le niveau de somnolence au volant d'un véhicule (Dong, Hu, Uchimura, & Murayama, 2011). Les principaux facteurs puis les principaux marqueurs de la somnolence seront donc décrits par la suite.

La somnolence peut être vue comme un état intermédiaire entre l'état d'une personne alerte/éveillée et l'état de sommeil. Il apparaît qu'il n'y a pas de critère simple, concret et général pour évaluer le niveau de somnolence d'un conducteur automobile. Une approche multifactorielle et personnalisée semble donc la plus prometteuse.

2.2 Facteurs influençant la somnolence

Étant donné la complexité de l'être humain, il est difficile d'isoler particulièrement un facteur unique pouvant influencer la somnolence dans l'objectif de l'étudier. Ces facteurs peuvent être distingués selon leur typologie : interne à l'individu ou externe c'est-à-dire, dans ce second cas, propre à la tâche de conduite. Les facteurs internes à l'individu vont principalement entraîner, pour le conducteur, de la fatigue et de la somnolence diurne excessives (Dababneh & Gindy, 2015). Une revue de littérature (Di Milia et al., 2011) dont est tirée la Figure 2-2 a répertorié un grand nombre de facteurs pouvant être liés à la fatigue. Cependant, il est important de noter qu'il n'existe pas de relation univoque directe entre la fatigue, la somnolence et ces différents facteurs. En effet, un conducteur peut être somnolent sans être fatigué mais également fatigué sans être somnolent (Stutts, Wilkins, & Vaughn, 1999). Une description permettra de décrire l'influence spécifique des principaux facteurs sur l'état de somnolence du conducteur tels que l'insuffisance et la qualité de sommeil, l'heure et la durée de conduite, l'environnement de conduite, ainsi que des facteurs internes tels que la prise de médicaments ou d'autres substances, les traits de personnalité et le créneau d'âge.

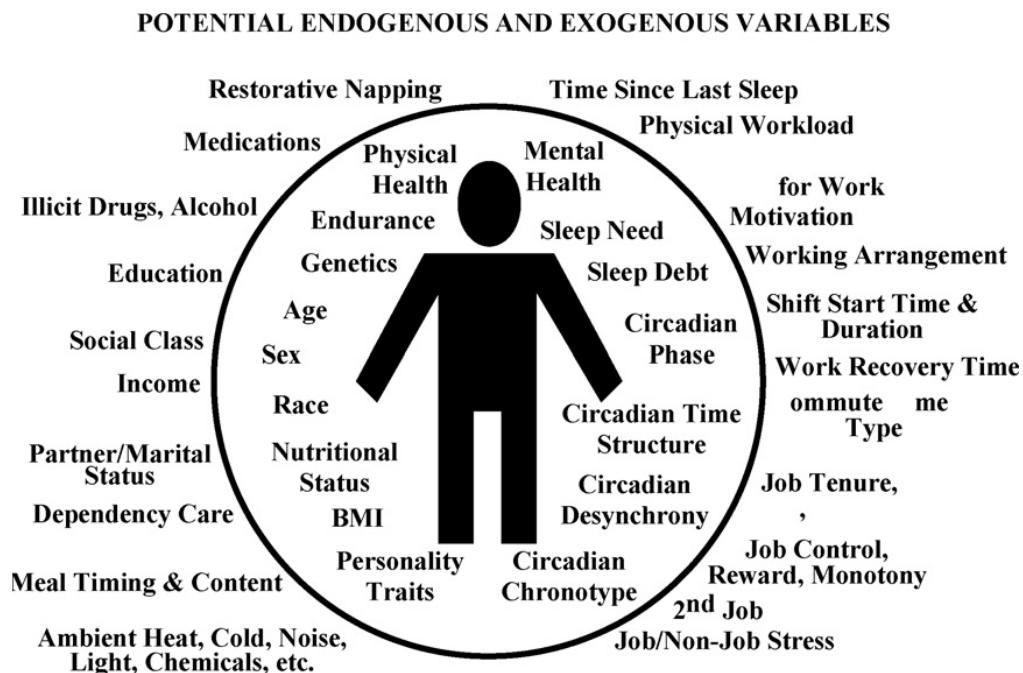


Figure 2-2 : Différents facteurs potentiels endogènes et exogènes pouvant être liés à la fatigue (extrait de Di Milia et al., 2011).

2.2.1 Insuffisance et qualité de sommeil

De toute évidence, l'insuffisance du temps de sommeil peut provoquer une somnolence diurne excessive. De plus, lorsqu'un individu ne dort pas suffisamment de manière chronique, cela peut engendrer un niveau d'éveil insuffisant le jour et augmenter le risque d'accident de conduite. Il en est de même pour la qualité de sommeil. Maia, Grandner, Findley, & Gurubhagavatula, (2013) ont montré en effet que le risque de connaître un épisode de somnolence au volant était multiplié par deux lorsqu'elles dorment 6 heures ou moins, et même par quatre lorsque les personnes dorment moins de 5 heures. De plus, il a été montré qu'une privation même partielle de sommeil (pendant laquelle le conducteur était autorisé à dormir de 3h à 7h) augmente la somnolence estimée de manière subjective et diminue les performances de conduite (Otmani, Pebayle, Roge, & Muzet, 2005). De plus, la qualité et l'insuffisance du temps de sommeil peuvent générer une fatigue dite chronique, qui découle de facteurs tels que des mauvaises conditions de travail avec par exemple des horaires de travail excessifs, des emplois du temps non optimaux et un climat organisationnel stressant (Matthews, 2000). Un sommeil perturbé par une insuffisance du temps de sommeil ou une mauvaise qualité de sommeil altéreront le rythme circadien. Cette notion va donc être développée, en relation avec l'horaire de conduite.

2.2.2 Horaire de conduite

La revue de littérature de Smiley, (1998) a également mis en valeur que l'horaire de conduite est un facteur clé pouvant provoquer une somnolence. En effet, deux pics d'accidents liés à la somnolence ont été observés. Un première plage horaire est identifiée en seconde partie de nuit, entre 2 heures et 6 heures, et dans une moindre mesure, une seconde est observée en début d'après-midi, entre 14 heures et 16 heures (Pack et al., 1995). En effet, les accidents sont 3 fois plus fréquents dans la plage horaire de l'après-midi qu'avant 10 heures ou après 19 heures (Horne & Reyner, 1999). Le fait que l'horaire de conduite ait une telle influence sur la fréquence des accidents s'explique en grand partie par le rythme circadien de la vigilance. En effet, la vigilance et le niveau d'éveil des êtres humains fluctuent selon une horloge biologique, située au niveau des noyaux suprachiasmatiques, proches de l'hypothalamus. Cette horloge biologique est influencée par et se régule grâce à la lumière du jour et plus principalement grâce à l'alternance jour/nuit. On l'appelle fréquemment horloge circadienne, à l'origine du rythme circadien de la vigilance, car la durée est proche de 24h. La Figure 2-3 illustre la variation du niveau de la vigilance au cours de la journée. Deux plages horaires de baisse de vigilance plus importantes sont observées :

- entre 3 heures et 6 heures, premier et plus important vallon sur la courbe figure 2-3,
- entre 14 heures et 16 heures, deuxième vallon moins important.

Par ailleurs, il est nécessaire de préciser que le niveau relatif d'éveil et de vigilance en fonction de ce rythme circadien peut être propre à chaque individu. Le questionnaire de typologie circadienne de Horne et Östberg (Horne & Östberg, 1975) ayant pour objectif de déterminer le chronotype des individus (caractérisant une personne comme « du matin » ou « plutôt du matin », « intermédiaire », « plutôt du soir » ou « du soir ») peut ainsi être utilisé dans une première approche.

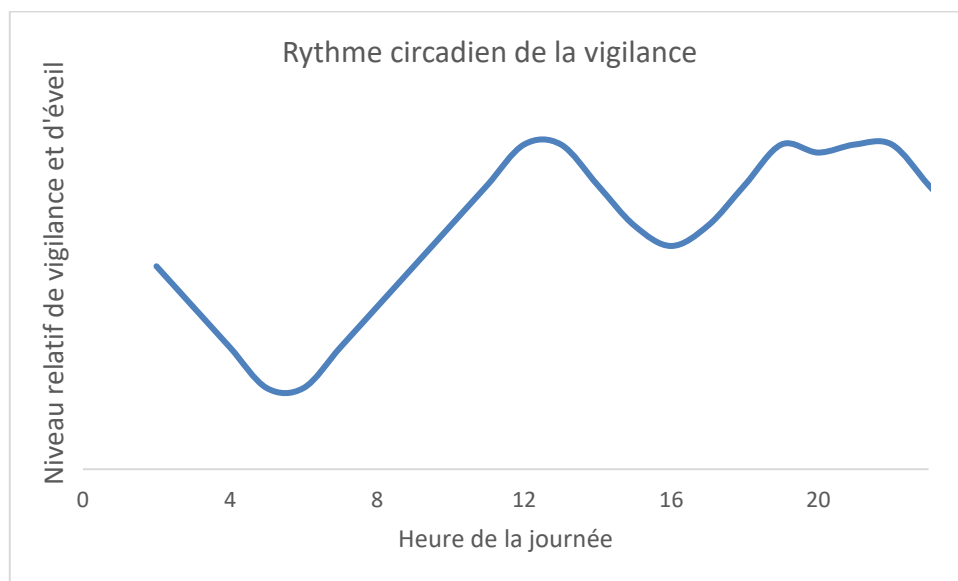


Figure 2-3 : Variation du niveau de la vigilance selon l'heure de la journée (illustration du rythme circadien)

2.2.3 Durée de conduite

Une autre notion temporelle, la durée de conduite, est aussi un facteur important de la somnolence au volant. Selon Smiley, (1998), la durée de la tâche, dans notre cas, la durée de conduite est connue comme induisant de la fatigue et une détérioration des performances de conduite. En effet, le temps de conduite contribue à augmenter la somnolence estimée de manière subjective et diminue les performances de conduite (Otmani, Pebayle, Roge, & Muzet, 2005). Cependant, il est important de noter que paradoxalement la durée de conduite n'est pas un facteur forcément accidentogène. En effet, il a été observé, notamment en Finlande, que 60 % des accidents survenaient pendant la première heure de conduite (Summala & Mikkola, 1994). De plus, une étude par questionnaire sur des usagers fréquents de la route et portant sur les accidents liés à la somnolence a montré que 44,5% des accidents étaient sur un trajet long et 53,3% sur un trajet court (Philip et al., 2010). Pourtant, en France,

la communication de la Sécurité Routière porte en grande partie sur la durée de conduite, réputée ne pas devoir dépasser deux heures sans faire une pause, et ce sans aucune mention des autres facteurs de risque. Les préconisations de s'arrêter toutes les deux heures ne sont donc pas suffisantes pour lutter efficacement contre la somnolence, qui peut à l'évidence apparaître plus tôt.

2.2.4 Environnement de conduite

Au-delà de ces facteurs temporels, divers éléments relatifs à l'environnement de conduite peuvent également influencer l'état du conducteur, devant s'adapter à un environnement changeant. La monotonie, le type de route et le trafic sont les trois principaux éléments contextuels pouvant accentuer le risque de somnolence.

La monotonie, liée au type de route, provoque un manque de variété et par conséquent d'intérêt, une tendance routinière des actions et du visuel, et amène une situation appelée « *highway hypnosis* », ou hypnose de l'autoroute (Williams & Shor, 1970). Ce phénomène a été mis en évidence pour la première fois en 1929. En effet, lors de tâches importantes comme la conduite, le conducteur désirent rester éveillé, pouvait finir par dormir les yeux ouverts (Miles, 1929). Il devient alors inconscient par rapport à la situation de conduite. En anglais, le terme « *driving without awareness* » a été employé pour qualifier la conduite sans conscience de la situation pour la première fois par Brown, (1991). La monotonie de la route amène donc un phénomène comparable à l'hypnose car les conducteurs fixent un point de la route à l'horizon et se révèlent incapables de réagir correctement à des situations potentiellement dangereuses. Au final, ce phénomène a été interprété comme une réduction de la vigilance et donc un état intermédiaire entre l'état d'éveil et celui de sommeil (Schlutz, 2006). Or, les infrastructures routières actuelles sur les autoroutes, souvent composées de longues courbes et grandes lignes droites, amènent fréquemment à cette sensation de monotonie perçue (Brandt et al., 2004). La monotonie va accélérer le processus de baisse de l'état de vigilance du conducteur et ainsi entraîner un risque important de somnolence, toutefois dans ce cas indépendante de l'état de fatigue (Larue, 2010). Ainsi, une étude par questionnaire a mis en valeur le fait que la somnolence était impliquée dans 19 % des accidents ayant lieu en ville, 21,9 % sur des routes départementales/régionales et 57,7 % sur des autoroutes. Lorsque l'on s'intéresse aux presque-accidents (un accident évité de justesse), cette fois la somnolence est impliquée dans 77,6 % des situations rapportées sur autoroute, contre seulement 1 % en ville et 15,2 % sur route départementales/régionales (Philip et al., 2010).

Bien entendu, le fait que le conducteur conduise dans un environnement dans lequel d'autres véhicules sont également présents complexifie la tâche de conduite. Il est donc nécessaire que celui-

ci s'adapte à cette situation qui est continuellement en évolution. Une situation avec trafic faible sur autoroute ne va pas réclamer le même niveau de ressources attentionnelles que la traversée d'une intersection chargée en trafic (Neboit, 1980). Il a également été observé qu'en conditions réelles dans un environnement urbain ou rural avec fort trafic, le conducteur, plus mobilisé d'un point de vue attentionnel, est forcément plus alerte (Barr, Yang, Hanowski, & Olson, 2011).

Après avoir mis en valeur des facteurs liés à la tâche de conduite, des facteurs internes au conducteur tels que la prise de substances, les traits de personnalité et l'âge vont être décrits.

2.2.5 Consommation de médicaments et autres substances

La substance indéniablement réputée comme dangereuse pour la conduite, même prise à plus ou moins longue échéance a conduite, est l'alcool. Au-delà de ces effets directs sur les capacités de conduite, la consommation d'alcool a également une forte influence sur le niveau de somnolence. En effet, toute consommation d'alcool diminue le temps à l'endormissement, mais perturbe aussi la qualité du sommeil car la consommation d'alcool induit une diminution des durées passées dans chacun des stades de sommeil : lent, profond et paradoxal (Ebrahim, Shapiro, Williams, & Fenwick, 2013). La qualité du sommeil étant perturbée, l'une des répercussions pouvant avoir des conséquences dangereuses au volant est la somnolence diurne.

Par ailleurs, d'autres substances comme la caféine ou la théine, entraînent l'effet inverse : le maintien d'un effet d'éveil, ce qui pourrait expliquer pourquoi près de la moitié des conducteurs boivent du café pour lutter contre la somnolence au volant (Anund, Kecklund, Peters, & Åkerstedt, 2008). La consommation de caféine augmente le temps de latence pour s'endormir avec et sans privation de sommeil (Lumley, Roehrs, Asker, Zorick, & Roth, 1987).

Certaines substances chimiques ainsi que certains médicaments peuvent aussi influencer l'état du conducteur en amenant une somnolence diurne excessive. Partant de ce constat, la législation a évolué et oblige les industriels du secteur à mentionner ce risque sur les emballages et les notices de leurs médicaments, sous la forme des symboles présentés sur la Figure 2-4. Parmi les substances ou médicaments les plus dangereux en ce sens, on trouve les sédatifs hypnotiques, les anxiolytiques, les antidépresseurs, les neuroleptiques, les antihistaminiques, les amphétamines, les opiacés et les produits dérivés de la cocaïne (Amato, 2011). La connaissance et la compréhension de ces symboles ne garantissent pas cependant que le conducteur soit conscient de l'effet de ces médicaments sur sa performance et son niveau de vigilance.



Figure 2-4 : Symboles pouvant apparaître sur la boîte d'un médicament

2.2.6 Trait de personnalité

Les traits de personnalité des conducteurs ont été étudiés en corrélation avec la somnolence et la fatigue afin de déterminer leurs influences potentielles sur la conduite. Trois dimensions ont alors été identifiées comme facteurs favorisant l'hypovigilance : l'extraversion, la recherche de sensation et la dépendance au champ (Thiffault & Bergeron, 1997). Tout d'abord, les personnes extraverties sont davantage à la recherche de stimulations provenant de l'environnement (Eysenck, 1967). Quant aux personnes à la recherche de sensations, elles ressentent le besoin de vivre des expériences variées, de prendre des risques physiques et/ou sociaux (Zuckerman, 1979). Ainsi, en raison de l'absence de nouveautés et de la monotonie de la conduite, les personnes recherchant des expériences sont plus amenées à faire des écarts de route et celles à la recherche de sensations fortes, ont également tendance à davantage s'endormir au volant (Thiffault & Bergeron, 2003a). Quant aux individus dépendants du champ, ils ont généralement des difficultés pour repérer un stimulus cible ou pertinent parmi de nombreux stimuli, non essentiels au sein d'une scène visuelle, pouvant induire une mauvaise prise d'information (Thiffault & Bergeron, 1997). Dans un environnement monotone, des personnes possédant ces traits de personnalités pourraient ainsi avoir tendance à être plus somnolentes et/ou à être somnolentes plus rapidement que les autres. La notion de conduite dynamique semble aussi être liée à la tendance à s'endormir dans des conditions monotones (Karrer, Vöhringer-Kuhnt, Baumgarten, & Briest, 2004). Néanmoins, ce facteur semble difficile à identifier à grande échelle car cela nécessiterait des passations, des questionnaires de personnalité à chaque conducteur, ce qui serait très contraignant d'un point de vue pratique.

2.2.7 Créneau d'âge

L'âge peut être considéré comme un facteur influençant la somnolence au volant. Une étude par questionnaire concernant les accidents liés à la somnolence a montré que les personnes âgées entre 18 et 30 ans présentaient un risque 1,42 fois plus important que les autres (Philip et al., 2010). Toutefois, le manque d'expérience de conduite, tout comme que l'âge, peut également expliquer cette augmentation de risque. Ces résultats confirment en effet ceux d'une étude précédemment menée par ces mêmes auteurs (Philip et al., 2004) indiquant que le temps de réaction (variable utilisée pour déterminer le niveau d'éveil de la personne) d'une population jeune (20 - 25 ans) est plus court que celui d'une population plus âgée (52 - 63 ans) après une nuit de 8 heures de sommeil, mais que ce temps de réaction est également plus altéré par une privation de sommeil de 24 heures. Les performances de la population plus âgée ne subissent que peu de modifications. Ainsi, nous pourrions en déduire que lorsque la qualité et la quantité de sommeil sont altérées, une population jeune sera plus amenée à être somnolente. Il a en outre été démontré qu'une population jeune est plus susceptible d'avoir des accidents causés par la somnolence très tôt le matin. Cet état de fait peut s'expliquer par les habitudes de vie des jeunes conducteurs, par exemple une conduite plus fréquente pendant une plage horaire très matinale où la vigilance diminue généralement, du fait des rythmes circadiens (Di Milia et al., 2011). Toutefois, cette affirmation doit être contrastée car chez une population plus âgée, les accidents (de tous types, pas seulement automobiles) sont en grande partie liés à l'état de santé physique et mentale, du fait de maladies chroniques ou de troubles du sommeil, alors que chez une population jeune, les accidents seraient plutôt dus à l'usage de substances comme le cannabis ou l'alcool, et à un sommeil de trop courte durée (Richard, Thélot, & Beck, 2013). De plus, il a également été observé qu'une population plus âgée se dit plus sensible à des défauts d'attention (attention sélective et/ou attention partagée) qu'une population plus jeune (Gabaude, 2010).

Bien que les facteurs d'influence abordés ne soient pas exhaustifs, ils présentent tous un impact non négligeable sur le risque de somnolence. Cependant, leur implication sur les variations de l'état réel du conducteur en véhicule reste à déterminer

2.3 Indicateurs pour estimer la somnolence au volant

La communauté scientifique a très tôt cherché à mesurer, de façon subjective et qualitative, ou objective et quantitative, le niveau de somnolence, en particulier (en ce qui nous concerne) dans un véhicule (réel ou simulé) lors d'une tâche de conduite. En raison de la complexité de cette notion de somnolence, plusieurs sources de données extrêmement variées ont été envisagées. Il est possible de les regrouper en différentes catégories :

- **évaluations subjectives,**
- **mesures cognitives ou test de performance,**
- **mesures physiologiques,**
- **mesures sensorimotrices,**
- **mesures provenant du véhicule.**

Ces différentes mesures seront décrites plus amplement dans les parties suivantes, en précisant leurs avantages et inconvénients respectifs.

2.3.1 Evaluations subjectives

Parmi les mesures subjectives, il est possible de les distinguer en deux catégories : les mesures auto-subjectives et les mesures subjectives évaluées par une tierce personne.

2.3.1.1 Mesures auto-subjectives

Les mesures auto-subjectives, également appelées parfois mesures psychologiques, consistent à demander à l'individu d'auto-évaluer au moyen d'un questionnaire son niveau de somnolence.

2.3.1.1.1 Variables pertinentes

L'évaluation subjective la plus couramment utilisée est l'échelle de somnolence de Karolinska (*Karolinska Sleepiness Scale* (KSS)) qui consiste à demander aux participants à leur étude d'évaluer leur niveau de somnolence sur une échelle de 1 à 9 (Åkerstedt & Gillberg (1990) ; Shahid, Wilkinson, Marcu, & Shapiro, 2011). Au cours de ces dernières années, la KSS a été l'un des instruments de mesures pour la détection de la somnolence par auto-évaluation les plus utilisés (Alhazmi, 2013; Daza et al., 2014; Friedrichs & Yang, 2010; Krajewski, Batliner, & Golz, 2009; Lee, Lee, & Chung, 2016; L. Li et al., 2014; Murata & Naitoh, 2015). L'échelle de somnolence de Stanford, comprise entre 1 et 7, est parfois utilisée (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973). Wylie et al. (1996) ont démontré la sensibilité et l'applicabilité de cette mesure dans le contexte automobile. La fiabilité et validité de ces mesures auto-subjectives ont également été démontrées (Kaida et al. 2006). Généralement, lors de la

conduite, le temps entre chaque évaluation de son propre niveau de somnolence est de 15 minutes, et cela particulièrement pour la KSS.

Par ailleurs, il existe des mesures qui permettent de déterminer la somnolence diurne générale telle que l'échelle de somnolence d'Epworth (M.W. Johns, 1991). Le rôle de cette dernière est d'évaluer le niveau de somnolence dans différentes situations de la vie courante. Ce test considéré comme fiable permet une auto-évaluation rapide du niveau de somnolence sur une journée d'activité.

Tableau 2-1 : La Karolinska Sleepiness Scale

Niveau	Description
1	Extrêmement éveillé
2	Très éveillé
3	Eveillé
4	Assez éveillé
5	Mi- éveillé, mi- somnolent
6	Signes de somnolence
7	Somnolent mais sans difficulté à rester éveillé
8	Somnolent avec difficulté à rester éveillé
9	Très somnolent luttant contre le sommeil.

2.3.1.1.2 Limites et avantages

Le principal avantage des mesures auto-subjectives est leur facilité de mise en place, c'est pourquoi elles ont été très souvent étudiées et utilisées dans le cadre d'études en conduite automobile (Alhazmi, 2013; Friedrichs & Yang, 2010; Murata & Naitoh, 2015). Une relation entre les clignements des yeux ou la position latérale et la KSS a été observée, mais de grandes différences individuelles sont également observées (Ingre, Åkerstedt, Peters, Anund, & Kecklund, 2006). Cependant, elles présentent plusieurs limites. Tout d'abord, ces mesures sont intrusives vis-à-vis de la tâche de conduite lorsqu'elles sont utilisées durant l'exécution de la tâche. De plus, il est parfois difficile d'auto-évaluer son propre niveau de somnolence, elles risquent d'être bruitées et peu fiables. De plus, cette méthode permet d'estimer l'état du conducteur au maximum toutes les 15 minutes car en augmentant cette fréquence, le conducteur est gêné trop fréquemment et cela risque de le maintenir éveillé et donc de biaiser la mesure. Avec ce choix de fréquence, il n'est toutefois pas possible de connaître en temps réel de brusques variations de l'état de somnolence du conducteur. Lorsque l'expérience dure plus de

3 heures sur une route monotone, il a été démontré que la KSS n'était pas appropriée en raison de la difficulté des participants à juger de leur propre état (Friedrichs & Yang, 2010; Schmidt et al., 2009). Finalement, l'évaluation auto-subjective n'est par définition pas une mesure objective et lorsque la tâche est monotone, il se peut que l'évaluation individuelle diffère de l'état physiologique de la personne (Brown, 1997). Ainsi, se baser seulement sur le ressenti du conducteur ne serait pas totalement fiable, celui-ci n'étant pas nécessairement conscient de son état dégradé de vigilance.

2.3.1.2 Evaluation subjective par une personne tierce

Une autre méthode subjective utilisée dans l'objectif de connaître l'état du conducteur se base sur son évaluation, par une tierce personne, à partir de vidéos filmant le comportement des conducteurs, sur la base de critères et d'échelles établis et validés.

2.3.1.2.1 Variables pertinentes

Cette méthode se base sur la subjectivité d'experts et non sur la subjectivité d'une auto-évaluation comme celle citée précédemment. Cette évaluation par « *vidéo coding* » se base sur l'éthologie qui permet de caractériser objectivement les comportements (Hancock & Parasuraman, 1992). Afin d'augmenter la fiabilité, cette méthode peut utiliser des critères, des échelles de notation et peut être effectuée par plusieurs observateurs en parallèle (Daza et al., 2011). L'échelle « *Trained observer rating* » définit l'état du conducteur de manière continue allant de « alerte » à « extrêmement somnolent » (Li, Seignez, Lambert, & Loonis, 2014; Wierwille & Ellsworth, 1994). Wierwille & Ellsworth (1994) ont élaboré une technique dans laquelle les expérimentateurs sont capables d'estimer le niveau de somnolence en se basant essentiellement sur des caractéristiques faciales telles que le tonus facial, les fermetures des paupières et des maniérismes visibles sur une vidéo. Ils ont mis en valeur une fiabilité inter et intra-individuelle de cette méthodologie. Parfois, l'échelle a été discrétisée en différents niveaux :

- **sur deux niveaux** (alerte ou somnolent) (McDonald, Lee, Schwarz, & Brown, 2013) ;
- **sur trois niveaux** : pas somnolent, moyennement somnolent, très somnolent (J. Wang, Li, & Lu, 2014) ;
- **sur cinq niveau**: alerte, peu somnolent, modérément somnolent, très somnolent et extrêmement somnolent pour la « *Observer Rating of Drowsiness* » (Belz, Robinson, & Casali, 2001; Rost et al., 2015).

2.3.1.2.2 Limites et avantages

Le principal avantage de cette méthodologie est qu'elle permet de prendre en compte l'ensemble des comportements, ce qui en fait une évaluation considérée comme fiable et complète de l'état du conducteur. Cependant, cette analyse est longue et nécessite généralement plusieurs observateurs possédant un certain niveau d'expertise ou au moins une formation à la méthode. Elle ne peut également être mesurée en temps réel dans un véhicule et nécessite un post-traitement coûteux. Cette forme d'étiquetage a d'ailleurs été décrite comme étant plus riche que celle utilisant l'électroencéphalographie et le PERCLOS (Millemann et al., 2001). Ces deux mesures/sources seront décrites par la suite.

Les auto-évaluations et les évaluations par une personne tierce sont formellement des mesures subjectives. Le développement de mesures objectives, quantitatives, est nécessaire, comme par exemple les mesures liées à des tests de performance.

2.3.2 Mesures cognitives ou tests de performance

Cette mesure fait appel à des indicateurs de performances sur une double tâche ou une tâche simple (en plus de la conduite), telles que le nombre d'erreurs ou le temps de réaction.

2.3.2.1 Variables pertinentes

Analyser la somnolence à partir de tests de performance tels que le temps de réaction ou le nombre d'erreurs lors d'une double tâche a été déjà fait l'objet de différents travaux (Nakayama, Futami, Nakamura, & Boer, 1999; Tango, Calefato, Minin, & Canovi, 2009). En effet, lorsque la vigilance diminue, le temps de réaction nécessaire à la détection d'une cible augmente, tout comme le nombre d'erreurs par omissions et/ou fausses alarmes (Davies & Parasuraman, 1982; Hancock & Parasuraman, 1992). Ces mesures sont également parfois utilisées pour évaluer le niveau de distraction des conducteurs (Liang, Reyes, & Lee, 2007; Tango et al., 2009). Dans le cadre de la conduite automobile, le paradigme de la double tâche a également été adapté en mesurant la réponse du conducteur via une perturbation de ses performances de conduite, objectivée par une déviation de la position sur la voie.

2.3.2.2 Limites et avantages

Les tests de performances peuvent parfois être contreproductifs car ils sont susceptibles d'induire un certain niveau de distraction et les critères d'évaluation retenus ne sont pas toujours disponibles en temps réel. Par ailleurs, les stimuli apparaissant de manière aléatoire dans le temps (car non-prédictibles), il n'est pas possible de connaître l'état du conducteur à un intervalle régulier. Enfin, l'implication dans une double tâche retarde l'arrivée de la somnolence (Grandjean, 1979). Cette mesure ne semble donc pas être pertinente afin d'évaluer l'impact de la somnolence dans le cadre de la conduite dans un environnement réel. Ce type d'approche est en effet trop problématique en raison des différentes perturbations pouvant survenir. Comme pour l'auto-évaluation, ces techniques ne semblent pas opportunes afin de connaître l'état du conducteur en temps réel. Différentes études se sont alors intéressées à l'utilisation de mesures physiologiques.

2.3.3 Mesures physiologiques

Il a été établi que lors de l'observation d'un changement du niveau de vigilance, cet état de fait peut se traduire par des changements d'états psychophysiologiques, et peuvent donc être mis en évidence par des mesures de l'activité du système nerveux central et/ou plus spécifiquement du système nerveux autonome (Boucsein, Haarmann, & Schaefer, 2007). Les signaux physiologiques les plus couramment utilisés pour mesurer l'état fonctionnel du conducteur sont notamment les niveaux d'activité : cérébrale, cardiaque, musculaire, respiratoire et électrodermale (Dong et al., 2011; Liu, Hosking, & Lenné, 2009). Nous verrons cependant dans ce chapitre que si la plupart de ces indicateurs physiologiques covariant bien avec l'état de vigilance (ou de somnolence), ils sont tous très loin de signaler de façon univoque un changement d'état spécifique. De ce fait, aucun indicateur, pris séparément, ne peut être considéré comme suffisant à lui seul.

Il est important de rappeler que le système nerveux de l'homme est classiquement divisé en deux parties : d'un côté le système nerveux central (SNC) et d'un autre, le système nerveux périphérique (SNP). Le SNC est composé de l'encéphale et de la moelle épinière. En ce qui concerne le système nerveux périphérique, il peut se diviser en système nerveux somatique (activation des muscles volontaires) et en système nerveux autonome (SNA) ou neurovégétatif (organes internes et muscles non associées à un contrôle volontaire). Concernant le système nerveux autonome, on distingue les systèmes sympathique et parasympathique. Le rôle du système parasympathique est de maintenir les fonctions corporelles par l'économie des fonctions cardio-vasculaires et respiratoires, notamment via la diminution de la fréquence cardiaque. Parallèlement, le système sympathique est quant à lui chargé des réactions de secours. Par exemple, à la suite d'un stress, le système sympathique

provoque une augmentation de la fréquence cardiaque, une sécrétion de salive et des sueurs. Ces processus sont néanmoins gouvernés par le système nerveux central.

Les activités cérébrale, cardiaque, respiratoire et électrodermale seront décrites par la suite. En revanche, l'activité musculaire, principalement étudiée en relation avec la fatigue musculaire ou physique (Jagannath & Balasubramanian, 2014) et qui se distingue de la fatigue cognitive ou mentale pouvant être mise en relation avec le niveau de somnolence du conducteur, ne fera pas l'objet d'un développement ultérieur.

2.3.3.1 Activité cérébrale

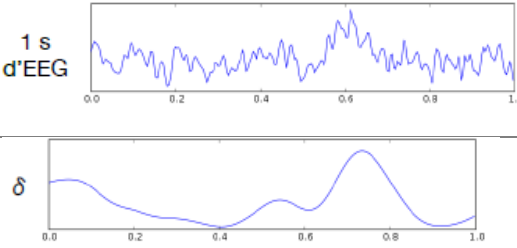
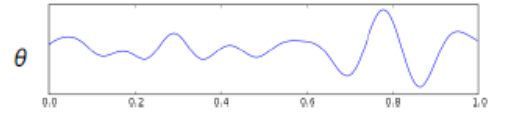
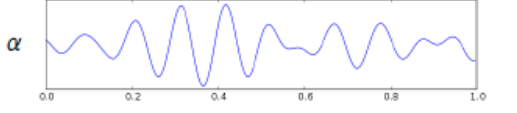
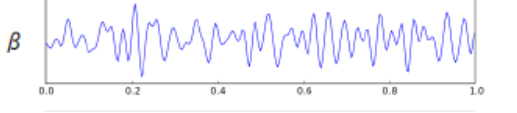
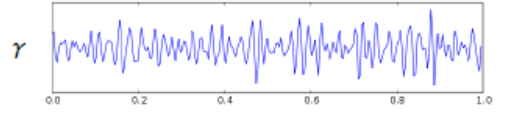
Le standard de référence utilisé pour détecter la somnolence à partir de mesures physiologiques est l'électroencéphalographie (EEG). En effet, l'EEG représente le principal indicateur non-invasif d'activité du système nerveux central au niveau du cortex cérébral, et peut être utilisé pour connaître la transition entre l'état d'éveil et d'endormissement (De Gennaro, Ferrara, Curcio, & Cristiani, 2001). L'EEG consiste à mesurer et analyser l'activité électrique du cerveau recueillie à distance par des électrodes placées sur le cuir chevelu. Dans le but d'estimer l'état dégradé du conducteur, l'analyse se concentre principalement sur la rythmicité de l'EEG dans le domaine fréquentiel et dans une moindre mesure sur les potentiels évoqués (*Event-Related-Potential*, ERP en anglais).

2.3.3.1.1 Variables pertinentes

En tant qu'indicateur le plus direct de l'activité du système nerveux central, l'EEG a été souvent utilisé pour détecter la somnolence (Hajinoroozi, Mao, & Huang, 2015; Lal, Craig, Boord, Kirkup, & Nguyen, 2003; F.-C. Lin, Ko, Chuang, Su, & Lin, 2012; Peiris et al., 2005; Picot, 2009; Yeo, Li, Shen, & Wilder-Smith, 2009; Yeo et al., 2009). Bien que différents indicateurs peuvent être utilisés, les marqueurs EEG relatifs à la somnolence combinent généralement la puissance spectrale mesurée dans la bande alpha (8-12Hz) (indicateur relativement précoce de la somnolence) et dans la bande thêta (4-7Hz) (associée à une somnolence plus importante) (Liu et al., 2009). En effet, la transition de l'activité cérébrale vers des fréquences plus lentes montre le passage d'un état éveillé à un état somnolent. Les ondes alpha peuvent être observées avec une augmentation modérée de la somnolence dans un premier temps dans les régions occipitales et centrales du cerveau (dérivations C3 et C4). Lorsque la somnolence apparaît plus clairement, l'onde thêta prend le relais. Si nous observons une onde delta, nous pouvons en conclure d'ores et déjà que le conducteur est dans un sommeil lent et profond. Un score de somnolence objectif (OSS), dérivé du signal EEG, a d'ailleurs été proposé comme mesure de la somnolence (Thorslund, Anund, Forsman, Gustafsson, & Soerensen, 2004).

Quant aux potentiels évoqués ils ont joué ces dernières années un rôle de plus en plus important afin de comprendre les mécanismes de l'attention (Luck, Woodman, & Vogel, 2000). Une étude a montré que les bandes alpha, bêta, l'amplitude de l'onde P300 du potentiel lié à un évènement (onde de positivité intervenant 300 ms après un stimulus), ou encore des mesures relatives à la variabilité du rythme cardiaque étaient significativement différentes avant et après une tâche de conduite monotone induisant de la fatigue mentale (Zhao, Zhao, Liu, & Zheng, 2012).

Tableau 2-2 : Bande de Fréquence EEG, (extrait de Larue, 2010)

Bandes de Fréquences	Noms Onde	Rôle	1 seconde d'EEG
De 0 à 4 Hz	Onde delta δ	Sommeil lent et profond	
De 4 à 8 Hz	Onde thêta θ	Diminution importante de la vigilance	
De 8 à 12 Hz	Onde Alpha α	Diminution modérée de la vigilance	
De 12 à 30 Hz	Onde Bêta β	État éveillé actif	
Supérieur à 30 Hz	Onde Gamma γ	Tâches cognitives ou fonctions motrices	

2.3.3.1.2 Limites et avantages

Le principal avantage de l'utilisation de l'EEG est que celui-ci peut être considéré comme un indicateur direct de l'activité du système nerveux central. Des chercheurs (Caldwell, Hall, & Erickson, 2002) ont réalisé une comparaison des résultats obtenus en laboratoire et en vol (hélicoptère) quant à l'utilisation d'ondes EEG afin de détecter le niveau de vigilance du pilote. Les résultats montrent que l'EEG est un bon indicateur pouvant mesurer en vol la fatigue (à l'échelle centrale). Une des principales limites de l'utilisation de l'EEG est cependant la difficulté à mesurer et analyser les signaux en situation

écologique car il représente un outil de mesure intrusif pour la conduite et de plus il est facilement bruité par les mouvements du participant ainsi que par l'environnement électromagnétique. Ces limites peuvent se traduire par un rapport signal sur bruit trop faible. Par ailleurs, Peiris et al. (2005) montrent que deux experts indépendants analysant l'EEG pour détecter la somnolence peuvent ne pas donner la même réponse pour le même participant en même temps. De surcroît, l'EEG étant un système de mesure nécessitant un placement précis des électrodes, sa pose est particulièrement délicate. Il serait donc difficilement envisageable de demander à chaque personne avant d'entrer dans sa voiture de porter un casque EEG. Quant aux potentiels évoqués, leur analyse fait nécessairement appel à des stimuli reproductibles et renvoie à des contraintes de traitement bien plus lourdes encore.

2.3.3.2 Activité cardiaque

D'autres mesures liées à l'activité du système nerveux périphérique ont été employées. Lorsque le conducteur devient somnolent, le système nerveux autonome est affecté, ce qui peut être identifié par une variation de la fréquence cardiaque. Cette dernière peut mettre en évidence une augmentation de l'activité sympathique et une réduction de l'activité vagale (Lal & Craig, 2002; Riemersma, Sanders, Wildervanck, & Gaillard, 1977).

2.3.3.2.1 Variables pertinentes

De nombreuses études ont utilisé l'analyse du rythme cardiaque pour l'estimation du niveau de somnolence du conducteur (Chua, McDarby, & Heneghan, 2008; Khushaba, Kodagoda, Lal, & Dissanayake, 2011; G. Li & Chung, 2013; Patel, Lal, Kavanagh, & Rossiter, 2011; Rodriguez Ibañez, García González, Ramos Castro, & Fernández Chimeno, 2011; Vicente, Laguna, Bartra, & Bailón, 2016; Yang, Lin, & Bhattacharya, 2010; C. Zhao et al., 2012). En étudiant l'activité cardiaque, des mesures peuvent être extraites soit dans le domaine temporel (rythme cardiaque moyen, écart-type, différence entre la fréquence cardiaque instantanée la plus élevée et plus faible au cours d'une période de temps, etc.) ou dans le domaine fréquentiel par l'étude de différentes bandes fréquentielles avec l'analyse de la densité spectrale. Concernant le domaine temporel, il a été montré que la fréquence cardiaque moyenne diminue pendant une conduite nocturne prolongée (Riemersma et al., 1977; Yang et al., 2010), mais aussi avec une augmentation du niveau de fatigue et de somnolence (Lal & Craig, 2001; Yang, Lin, & Bhattacharya, 2010). Quant au domaine fréquentiel, il a pour but l'analyse des composantes oscillatoires du signal en analysant la variabilité du rythme cardiaque. Différentes bandes de fréquences sont analysées afin de caractériser la variabilité du rythme cardiaque (Appelhans & Luecken, 2006) :

- très basses fréquences de 0 à 0.04 Hz (VLF) : englobant des mécanismes de régulation tels que la thermorégulation ;
- basses fréquences 0.04 à 0.15 Hz (LF) : assimilée de nos jours à l'activité du système nerveux sympathique (même si cela est encore très discuté) ;
- hautes fréquences 0.4 à 0.40 Hz (HF) : assimilée de nos jours à l'activité parasympathique ;
- très hautes fréquences 0.4 à 3 Hz (VHF).

L'analyse s'est également centrée sur des rapports de fréquences avec principalement LF/HF également appelé le « sympathetic-vagal ratio » qui diminue lorsque le conducteur devient fatigué et/ou somnolent (Appelhans & Luecken, 2006; Calcagnini, Biancalana, Giubilei, Strano, & Cerutti, 1994). Ce rapport permettrait de savoir quel système prédomine le contrôle de l'activité cardiaque.

En outre, des mesures non linéaires telles que l'entropie d'un signal sont également utilisées pour mesurer la somnolence au volant. Une étude a montré que l'entropie approchée de l'ECG, et les bandes de puissance LF et HF sont significativement différentes avant et après la fin de la tâche de conduite (Zhao, Zhao, Liu, & Zheng, 2012). L'entropie de Shannon sur un signal ECG a également été utilisée pour détecter des événements de somnolence ; cette dernière étant plus élevée lorsque le conducteur devient somnolent (Li & Chung, 2013) .

2.3.3.2 Limites et avantages

L'utilisation fréquente du rythme cardiaque s'explique par le fait qu'il peut facilement être mesurable à l'aide d'une montre connectée ou d'une caméra en analysant les variations de couleur de la peau (Bousefsaf, 2014). Néanmoins, là encore, l'activité cardiaque est loin de signer les variations d'un seul état. Par exemple, certaines substances font baisser la fréquence cardiaque alors que d'autres l'accélèrent, notamment la nicotine ou la caféine (D. Smith, Tong, & Leigh, 1977). Le rythme cardiaque peut également varier en fonction d'autres états cognitifs comme la charge de travail et le stress (Healey & Picard, 2005; Keshan, Parimi, & Bichindaritz, 2015; Reimer et al., 2009). Ainsi, il peut être difficile de savoir si une variation du rythme cardiaque est spécifique à la somnolence ou non. La fréquence cardiaque diminue habituellement pendant la conduite et lorsque le conducteur est fatigué (Lal & Craig, 2001), mais le contraire peut également se produire (Apparies, Riniolo, & Porges, 1998).

2.3.3.3 Activité respiratoire

Le couplage entre la respiration et le rythme cardiaque entraîne des variations de la fréquence cardiaque en fonction du cycle respiratoire. Ce phénomène est appelé « arythmie sinusale respiratoire ». Fonction vitale de l'organisme, la respiration est un processus permettant l'échange sanguin de gaz carbonique et d'oxygène à travers les poumons et les alvéoles. La particularité de cette activité est qu'elle peut être contrôlée de manière à la fois inconsciente et volontaire. Généralement, la respiration est mesurée à l'aide de sangles flexibles placées sur la cage thoracique et abdominale. Elle peut également se baser sur la pression de l'air expiré et sur la température de celui-ci.

2.3.3.3.1 Variables pertinentes

Concernant la respiration, l'amplitude du signal, la fréquence (Mehler, Reimer, Coughlin, & Dusek, 2009; Takahashi & Yokoyama, 2011), des mesures concernant la régularité (Lee, Lee, & Chung, 2014) ou également un mix de ces mesures (Ju et al., 2015) ont été utilisés dans le cadre de la somnolence au volant. Le rythme respiratoire augmente avec une charge de travail élevée (Mehler et al., 2009)

2.3.3.3.2 Limites et avantages

Même si cette donnée est facilement mesurable, comme pour les précédentes, la fréquence de respiration varie en fonction d'autres états (stress par exemple (Healey & Picard, 2005)) et n'a jamais été utilisée seule, à notre connaissance, comme mesure de la somnolence, notamment du fait du manque de spécificité en lien avec une variation de l'état de somnolence du conducteur.

2.3.3.4 Activité électrodermale

Une autre mesure pouvant être liée à la somnolence au volant qui sera présentée ici est la conductance électrodermale. Souvent nommée activité électrodermale (EDA), cette mesure témoigne de la conductivité électrique enregistrée à la surface de la peau. Cette conductivité reflète l'activité des glandes sudoripares sous le contrôle du système nerveux autonome. L'EDA est quelquefois enregistrée comme mesure indirecte de l'attention, de l'effort cognitif et surtout des émotions. On distingue généralement deux composantes, tonique et phasique. La composante phasique (on parle alors de réponse électrodermale ou RED) correspond à une réponse à un stimulus, ce sont des événements discrets sur le signal qui se cumulent avec le signal tonique. Ce dernier concerne des variations plus lentes, plutôt liées au niveau d'activation du participant, et il est également sensible à la modulation sympathique.

2.3.3.4.1 Variables pertinentes

Afin d'étudier l'activité phasique de l'individu, la réponse électrodermale (RED) est analysée. Elle prend la forme d'une soudaine variation dans le signal enregistré. Il est de coutume de différencier les réponses électrodermales causées à la suite d'un stimulus (RED) des réponses électrodermales non spécifiques (RED-NS). Ces dernières peuvent se manifester au repos à une fréquence variant de 0.1 Hz à 0.3 Hz (Dawson, Schell, & Fillion, 2007). La figure 2-5 illustre deux formes différentes que peuvent prendre les RED. Une RED commence tout d'abord par une chute de la courbe pour parvenir à un minimum local, plus ou moins durable, puis ce niveau augmente de nouveau, c'est alors la récupération. Différentes mesures sont utilisées dans la littérature afin d'analyser, de quantifier et de décrire les RED et RED-NS telles que des indicateurs d'amplitude, des indicateurs de durée ou de pente de la chute.

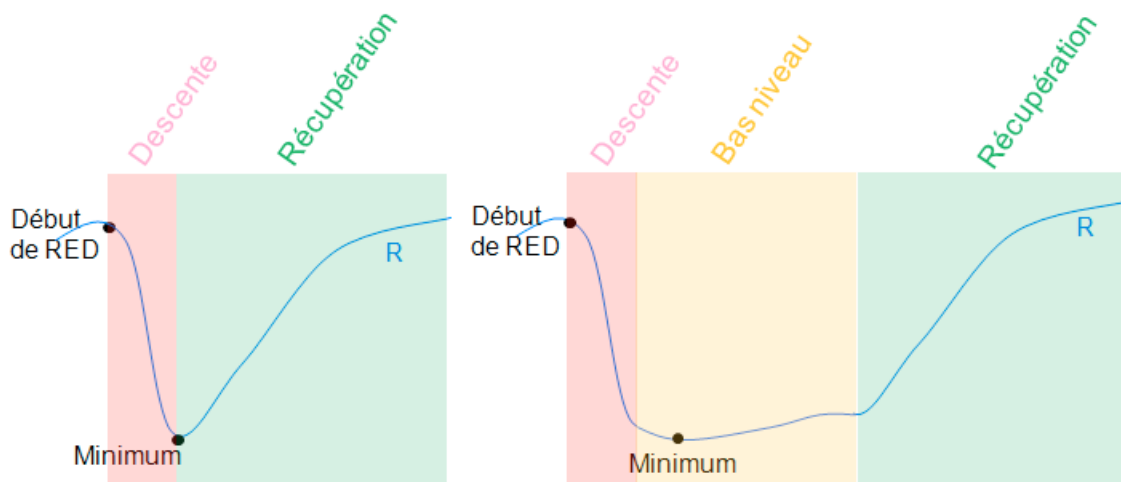


Figure 2-5 : Deux RED de formes différentes avec une phase de descente en rose, suivie ou non d'une phase de bas niveau (en jaune sur celle de droite), suivie d'une phase de récupération en vert (extrait de Clarion, 2009)

Quant à la composante tonique, elle est amenée à varier avec l'éveil physiologique de l'individu et elle est fortement liée à la puissance de la bande alpha de l'EEG (Vaez Mousavi, Barry, Rushby, & Clarke, 2007). En ce qui concerne cette mesure contenant des variations lentes, le niveau moyen ainsi que la dispersion sont généralement calculés.

2.3.3.4.2 Limites et avantages

Même si l'activité électrodermale a été utilisée dans la littérature pour détecter la somnolence au volant (Bundele & Banerjee, 2009; Larue, Rakotonirainy, & Pettitt, 2011) et la charge mentale d'un opérateur ou d'un conducteur (Clarion, 2009; Clarion et al., 2009), la principale contrainte de cette mesure est son caractère intrusif et gênant pour la conduite. En effet, le meilleur signal est présent

lorsque les électrodes sont placées sur les doigts de la main non dominante (Fowles et al., 1981). Le contact des électrodes avec le volant peut, donc, causer des mauvais contacts, ce qui rend l'enregistrement de données de qualité très délicat. De plus, des variations de l'activité électrodermale peuvent être également dues aux émotions (Rebolledo-Mendez, Reyes, Paszkowicz, Domingo, & Skrypchuk, 2014) et au stress (Healey & Picard, 2005).

Le lien entre les indicateurs physiologiques et les changements de l'état cognitif, comme la somnolence, nécessite d'extraire des indicateurs pertinents par des méthodes de traitement. Les principaux avantages de l'utilisation des mesures physiologiques reposent sur l'objectivité, sur la possibilité d'utilisation en temps réel et le fait qu'elles soient quantitatives. Cependant, comme il est difficile de contrôler l'état cognitif réel, les mesures peuvent être influencées par d'autres états et/ou également par le contexte. Malgré la miniaturisation des technologies, elles peuvent rester tout de même intrusives et le prétraitement, pour augmenter le rapport signal sur bruit, reste problématique.

2.3.4 Mesures sensorimotrices

La recherche de mesures objectives et pertinentes pour détecter la somnolence au volant s'est beaucoup orientée ces dernières années sur l'analyse du regard, des mouvements de tête et de paupières (Li et al., 2012; Liu et al., 2009).

La mesure des mouvements oculaires peut également être catégorisée dans la liste des mesures physiologiques car elle peut être enregistrée avec un électro-oculogramme (EOG) ou actuellement un vidéo-oculogramme (VOG), apte à détecter une activité sensorimotrice. Toutefois, nous avons décidé de séparer la mesure sensorimotrice des mesures physiologiques car, de nos jours, celle-ci se mesure prioritairement à l'aide de caméras (VOG) plutôt que par la mesure d'un signal physiologique (EOG).

2.3.4.1 Variables pertinentes

Le comportement des conducteurs change avec l'arrivée de la somnolence. Ils font moins de mouvements de tête et font plus de maniérismes (se toucher la tête, se gratter etc.). Ils penchent la tête d'un côté, ils baillent plus mais, surtout, l'activité des paupières se modifie (Eskandarian, Sayed, Delaigue, Mortazavi, & Blum, 2007). Plusieurs indicateurs, extraits des mouvements des paupières, tels que la fréquence, la durée des clignements et le PERCLOS (*PERcentage of eye CLOSure*) sont

amenés à varier avec le niveau de somnolence de la personne. Le PERCLOS est défini comme la proportion de temps passé pendant laquelle l'œil est fermé à plus d'un certain pourcentage pendant une durée spécifique (généralement 80 % pendant une minute) (Wierwille & Ellsworth, 1994). Il est reconnu comme un standard dans la communauté internationale, même s'il est notoirement insuffisant.

Listés comme des indicateurs de la fatigue, les clignements oculaires se produisent de façon spontanée et représentent par conséquent un phénomène normal, simple et facilement accessible reflétant l'activation du système nerveux central (Caffier, Erdmann, & Ullsperger, 2003). Dans la littérature, il est constaté que le temps de fermeture des paupières et la fréquence des clignements augmentent avec la somnolence (Caffier et al., 2003).

Une étude clé de Friedrichs et Yang (2010) a testé 18 indicateurs différents pour tenter d'estimer la somnolence en situation de conduite réelle : un réseau de neurones artificiels (ANN) a été entraîné à évaluer le niveau de fatigue (tel que mesuré par l'échelle KSS) à partir de combinaisons de ces différents indicateurs. Les indicateurs les plus prometteurs étaient sélectionnés au préalable, via l'algorithme de sélection séquentielle "*Sequential Floating Forward Selection*" (SFFS) qui s'est montré plus performant que les techniques de réduction de dimensions linéaires (PCA et LDA). Ils obtiennent ainsi un taux de reconnaissance moyen de 82,5 %. Cependant, le détecteur était bien plus performant pour estimer les états d'éveil que liés à un niveau de fatigue avancé. Cet état de fait peut en partie s'expliquer par la faible quantité de données disponibles pour un niveau de somnolence avancé (KSS>9). De plus, l'estimation ne peut être meilleure que celle permise par l'échelle KSS, elle-même imprécise, car subjective, et de faible résolution temporelle. Les indicateurs les plus pertinents se révèlent être, dans cette étude, ceux relatifs à la vitesse d'ouverture des yeux, au PERCLOS et aux mouvements de la tête. La durée du clignement est également adaptée pour détecter des états de fatigue avancés, mais la difficulté de l'algorithme de détection à différencier clignements et mouvements de regards verticaux en direction du tableau de bord limite en pratique l'utilisation de cette ouverture oculaire, s'atténuant avec la fatigue. De plus, les auteurs pointent la nécessité de prendre en compte les spécificités individuelles. D'une part, il est nécessaire de se référer à une ligne de base pour chacun des participants car de forts écarts de valeurs existent entre eux, notamment quant à la durée et à la fréquence des clignements. D'autre part, une détection de la somnolence, basée uniquement sur l'utilisation d'informations recueillies via une caméra (mouvement des yeux et de la tête), est inadaptée pour certains conducteurs (à cause du port de lunettes par exemple).

Au-delà des mouvements oculaires, de nombreux travaux portent également sur l'analyse de la position et du mouvement du regard du conducteur comme mesure d'états cognitifs (Li et al., 2012).

L'analyse combinée de la mobilité oculaire, céphalique et labiale (en vue de détecter d'éventuels bâillements), peut être exploitée pour évaluer l'état de vigilance du participant (Smith, Shah, & da Vitoria Lobo, 2000). L'analyse des mouvements de la bouche est également utilisée pour détecter la distraction du conducteur (conversations avec des passagers ou au téléphone avec le Bluetooth par exemple) (Young, Lee, & Regan, 2008).

2.3.4.2 Limites et avantages

Le principal avantage de ces mesures, qui justifie leurs utilisations importantes dans les applications industrielles de ces recherches, est le fait que ces mesures soient facilement accessibles de façon non intrusive car il n'y a pas besoin de capteurs en contact avec le participant. Par conséquent, elles peuvent être facilement utilisables en situation réelle.

Néanmoins, ces mesures reposent sur le fait que les acquisitions sont basées sur des algorithmes de traitement d'images afin de détecter la tête, la bouche, les yeux et le regard. De ce fait, il en découle que la qualité de l'estimation de l'état du conducteur est fortement dépendante de la qualité du traitement d'images. Ce dernier reste encore difficile dans le cas de la conduite automobile en particulier pour les personnes portant des lunettes, supportant des changements de luminosité ou des situations de faible luminosité (Friedrichs & Yang, 2010). A cela, s'ajoute un certain nombre de contraintes liées au placement de la caméra dans le cockpit, et du champ couvert par l'objectif. Différentes méthodes et algorithmes sont décrits dans la littérature pour détecter la somnolence en se basant uniquement sur des mesures oculaires comme SVM (*support vector machine*) (Sabet, Zoroofi, Sadeghniaat-Haghighi, & Sabbaghian, 2012), *adaboost* avec Haar (Lin, Lin, & Chung, 2013) ou divers techniques de traitement d'image (Fuletra & Bosamiya, 2013).

Les mesures comportementales ont connu un grand intérêt du fait de leurs sensibilités relatives à la somnolence au volant et de leurs facilités d'enregistrement. Néanmoins, la qualité des données dans le contexte automobile peut se révéler amoindrie.

2.3.5 Mesures véhicules

Les mesures véhicules, reflétant l'activité de conduite, peuvent être utilisées pour estimer le niveau de somnolence des conducteurs.

2.3.5.1 Variables pertinentes

Les différentes mesures utilisées peuvent être catégorisées selon leur action (contrôle longitudinal/latéral) et leur finalité (activité ou performance de conduite). De nombreux travaux tels que ceux menés par Arnedt, Wilde, Munt, & MacLean (2001) et De Valck, De Groot, & Cluydts (2003) ont ainsi utilisé la déviation standard de position latérale (*Standard Deviation of Lane Position* : SDLP) et une mesure liée à la vitesse comme indicateur de la somnolence au volant.

Tableau 2-3 : Les différentes mesures véhicules

	Performance de conduite	Activité de conduite
Contrôle longitudinal	Contrôle de la distance inter-véhiculaire et de la vitesse	Activité sur les pédales
Contrôle latéral	Performance de tenue de voie	Activité sur volant

Les mesures les plus utilisées concernent principalement le contrôle latéral avec l'écart-type de la position dans la voie (*Standard Deviation of Lane Position* : SDLP) et les mouvements du volant (*Steering Wheel Movement* : SWM) (Liu et al., 2009). Ainsi, il a été démontré, notamment par Ingre, Åkerstedt, Peters, Anund, & Kecklund (2006) que les conducteurs font en moyenne plus de déviations par rapport au centre de la voie (SDLP) lorsque leurs mesures subjectives de somnolence (échelle KSS) augmentent principalement à partir de niveaux KSS supérieurs à 7. Cette mesure semble donc pertinente pour détecter des états de somnolence déjà avancés comme explicité précédemment (Horne & Reyner, 1996). Cependant, il s'agit de la valeur moyenne pour l'ensemble des 10 participants de cette étude. Cette observation n'est pas vérifiée au cas par cas pour chacun des participants (certains participants présentant des SDLP toujours inférieurs à 0.3, même pour des niveaux très élevés sur l'échelle KSS). Quant au SWM, une diminution de petits SWM (de l'ordre de 1° à 5°) et une augmentation des grands SWM (de l'ordre de 6° à 10°) sont observées lorsque la durée de conduite augmente et s'accompagne implicitement, d'une augmentation de la somnolence (Fairclough & Graham, 1999; Thiffault & Bergeron, 2003b). Ainsi, un conducteur somnolent ne va plus forcément réaliser des petits écarts de voie mais plutôt des grands écarts de voie. Ces derniers seront donc

corrigés par des grands changements d'angle volant. De plus, une plus grande variabilité de l'angle au volant a également été observée avec une augmentation de la somnolence au volant observable par exemple grâce à un calcul d'entropie (Nakayama et al., 1999).

De ce fait, la plupart des études utilisent des sources multiples. Alhazmi (2013) utilise la vitesse angulaire au volant, la pression sur la pédale d'accélérateur et de frein et la force de préhension sur le volant. Larue, Rakotonirainy, & Pettitt (2011) en plus d'utiliser des mesures physiologiques, utilisent la SDLP, la variabilité de la vitesse du volant et également le temps pour franchir la ligne médiane. Zhang, Owechko, & Zhang (2004) ont quant à eux enregistré la vitesse moyenne du véhicule, l'écart-type de vitesse, sa position moyenne, l'écart-type de l'angle de braquage, l'accélération moyenne et l'écart-type de l'accélération du véhicule en plus des informations sur le diamètre de la pupille. Stein (1995) quant à lui a mis en exergue que la SDLP, l'écart sur la voie et la position latérale sont des indicateurs de somnolence.

2.3.5.2 Limites et avantages

Le principal avantage de ces mesures liés au véhicule, qui justifie d'ailleurs leur forte utilisation, est leur caractère non invasif et non intrusif. En effet, certaines informations directement liées à l'activité de conduite sont déjà disponibles au sein du véhicule grâce au réseau de contrôle du véhicule (Controller Area Network ou CAN) comme la pression sur la pédale d'accélérateur, de frein, l'angle au volant, etc. et ainsi elles ne nécessitent pas de capteurs spécifiques additionnels. D'autre part, des systèmes d'aide à la conduite, comme une aide au maintien de la trajectoire, s'appuient déjà sur des informations liées à la position du véhicule dans la voie. Cependant, les lignes sur les voies ne sont pas toujours visibles. De ce fait, la SDLP ou le temps pour franchir la ligne ne sont pas toujours disponibles en temps réel. En effet, la SDLP nécessite une détection robuste des lignes de bord de voie, ce qui selon l'état de la route, n'est pas toujours possible. Il est également difficile de comparer les valeurs numériques et les résultats présentés par différentes études car toutes ne font pas de différences entre biais et précision, tel que mentionné dans la revue de littérature de Liu et al. (2009). En effet, les conducteurs peuvent ou non, maintenir leur véhicule au centre de la voie d'une façon générale. Or, biais et précision sont vraisemblablement deux indicateurs différents des performances de conduite. Cette difficulté se rencontre d'une façon générale pour l'ensemble des mesures et indicateurs discutés : l'absence de standards méthodologiques rend extrêmement difficile la comparaison des études entre elles et l'évaluation de la pertinence des indicateurs utilisés. En outre, ces mesures peuvent être altérées par d'autres états que la somnolence, comme la distraction (Tango et al., 2009) ou un déclin des processus attentionnels (Marin-Lamellet et al., 2003). Finalement, dans le cas d'un véhicule autonome ou partiellement autonome, ces informations ne pourront pas, ou partiellement, être utilisées pour connaître l'état du conducteur et savoir s'il est dans un état apte à reprendre la main.

2.4 Véritable état du conducteur

Précédemment, nous avons vu différentes mesures qui permettent de manière indirecte d'estimer dans quel état se trouve le conducteur. La principale difficulté réside dans le fait que le véritable état de somnolence du conducteur n'est quant à lui pas accessible directement. On peut parler d'état caché. Cependant, l'étalonnage et la validation de ces mesures nécessitent de connaître le véritable état de celui-ci, communément appelé « ground truth » ou en français « vérité de terrain ». C'est pourquoi certaines d'entre elles sont également estimées comme plus fiables et plus directement liées à l'état du conducteur. Toutefois, il n'existe pas de consensus sur la référence absolue (gold standard) de la somnolence. Par rapport à toutes les mesures citées précédemment, celles utilisées comme vérité de terrain sont principalement l'échelle de somnolence de Karolinska (KSS), l'évaluation subjective par des personnes tierces, les tests de performance tels que les tests de temps de réaction simples, les mesures EEG, les franchissements de lignes ou également une combinaison de l'EEG et des clignements avec l'échelle de la somnolence objective (ou OSS pour *Objective Sleepiness Scale* en anglais) (Muzet et al., 2004). Notre choix s'est orienté sur l'évaluation subjective par des personnes tierces, nous justifierons notre choix dans la méthodologie.

2.5 Nécessité d'avoir une approche hybride

L'examen de la littérature sur le monitoring de la vigilance (forcément partiel car centré sur l'application au véhicule routier) nous conforte dans notre position initiale, à savoir qu'il n'existe pas de méthode idéale dans l'absolu, et que de toute façon aucune mesure n'est à même de refléter, à elle seule, les variations significatives de l'état de vigilance du conducteur. Afin de minimiser le nombre de fausses alarmes en maintenant un taux de reconnaissance élevé, les mesures ci-dessus sont bien souvent utilisées de façon conjointe (Dong et al., 2011; Samiee, Azadi, Kazemi, Nahvi, & Eichberger, 2014). En effet, une seule mesure peut être source d'imprécision, être incomplète ou incertaine. La relation entre une mesure et un état n'est jamais univoque. Ainsi, aucune mesure fiable, peu invasive et sûre n'émerge comme mesure de référence dans le domaine pouvant être utilisée en temps réel. C'est pourquoi les études se sont tournées vers l'utilisation conjointe de différentes mesures afin de multiplier les sources d'informations. De plus, le fait qu'il n'existe pas de relation univoque entre le changement d'une mesure donnée et la somnolence justifie également l'utilisation des mesures hybrides dans la littérature. En effet, un des principaux problèmes est la variabilité interindividuelle des mesures et de la dynamique de la somnolence dues à un manque de sommeil par exemple. En

effet, face à un manque de sommeil important, un individu peut souffrir d'une somnolence importante et avoir des performances de conduite hautement dégradées alors qu'un autre individu compensera, lui, cette dette de sommeil. Ces différentes réponses à la privation de sommeil semblent être systématiques et relativement stables entre les individus et dans le temps (Van Dongen, Baynard, Maislin, & Dinges, 2004). Dans la littérature, nous pouvons trouver différentes combinaisons de ces mesures par exemple :

- EEG, ECG, mouvement oculaire et contexte (Yang et al., 2010),
- clignement des paupières et activité cardiaque (Sun & Yu, 2014),
- activité cardiaque, cérébrale, hormones, mouvement oculaire, respiration, position sur le siège, activité du volant, position et pression sur le volant pour détecter la charge de travail (Reimer et al., 2009),
- activité cérébrale, mouvements oculaires et EMG (Kurt, Sezgin, Akin, Kirbas, & Bayram, 2009),
- activité cardiaque, électrodermale, oculaire et activité et performance de conduite avec un réseau de neurones (Larue et al., 2011),
- clignements des paupières détectés à partir de vidéos et EEG (Picot, 2009),

Une étude a également révélé la pertinence de l'utilisation de la fusion de données en montrant que la combinaison de 3 réseaux de neurones entraînés sur des données de sources différentes (position latérale, variation de l'angle au volant et comportement oculaire lié aux clignements) permettait d'améliorer les performances de détection (Samiee et al., 2014)

Au final, une limite à ces approches multimodales provient de l'augmentation du nombre de mesures. Même si celle-ci améliore l'évaluation de l'état du conducteur, la méthodologie pour détecter sa somnolence sera plus intrusive du fait de la multiplication du nombre de capteurs. Les algorithmes de traitement à mettre en place peuvent être également plus complexes et plus lourds en termes de temps et de traitement de calculs. Comme vu succinctement précédemment, un moyen de fusionner différentes mesures est l'utilisation de modèles d'apprentissage automatique.

2.6 Modélisation à l'aide d'apprentissage automatique

Dans le cadre de ce travail de recherche, afin de traiter le problème de l'intégration (fusion) de données hétérogènes, un modèle est conçu à l'aide d'approches du domaine de l'intelligence artificielle et plus précisément de l'**apprentissage automatique** (ou en anglais *machine learning*). Cette branche de l'intelligence artificielle consiste à développer, à analyser et à implémenter des méthodes permettant à une machine d'apprendre à effectuer une tâche à partir d'exemples, comme reconnaître des images.

Dès son plus jeune âge, l'être humain apprend à lire, à écrire, à reconnaître des objets... À partir d'observations, son cerveau est capable de transférer des informations en pleine connaissance. L'objectif de l'apprentissage automatique est d'imiter ces capacités humaines extraordinaires, ou tout du moins de s'en inspirer. Lorsqu'un enfant apprend à reconnaître des chats et des chiens, cet apprentissage se réalise grâce à plusieurs observations d'images et en situation réelle. Aucune description analytique ni topologique n'est généralement donnée à l'enfant. L'apprentissage de l'enfant a été réalisé grâce à « des exemples ». Le type de modélisation qui sera sélectionné dans le cadre de cette thèse utilisera des algorithmes d'apprentissage automatique en s'appuyant sur des « exemples » de personnes somnolentes et non somnolentes.

Par ailleurs, si un nouveau chat de taille différente, auquel il manque une oreille, d'une couleur différente, est présenté à l'enfant, on souhaite que l'enfant reconnaisse, malgré ces différences, le concept de chat. On appelle ce phénomène la « généralisation ». Cette notion est très importante en apprentissage automatique. En effet, on attend de l'algorithme qu'il soit capable d'apprendre, non seulement à reconnaître des exemples qu'il a « déjà vus », mais surtout de généraliser sur des nouveaux exemples, jamais vus jusqu'alors.

Cette partie va tout d'abord décrire brièvement en quoi consiste l'apprentissage automatique, puis, les étapes de développement d'un modèle seront détaillées. La notion de généralisation sera explicitée en abordant deux notions clés : le sous-apprentissage et le sur-apprentissage. Puis une étape clé de cette thèse sera décrite : la division des jeux de données. Ensuite, seront présentés quelques modèles utilisés pour détecter et prédire la somnolence dans la littérature. Enfin, une présentation d'un réseau de neurones artificiels (principale modélisation de cette thèse) conclura cette partie sur la modélisation.

2.6.1 Modèles de classification et de régression : définitions

L'apprentissage automatique se divise généralement en différents types d'apprentissage : apprentissage par renforcement, non supervisé et supervisé. L'apprentissage par renforcement consiste à apprendre à partir d'expériences et de situations avec comme finalité l'optimisation d'une récompense du temps. Cet apprentissage est souvent utilisé dans la robotique pour, par exemple, apprendre aux robots à prendre des décisions. L'apprentissage non supervisé consiste à regrouper des données entre elles en fonction de leurs similarités, sans que l'on connaisse a priori la réponse que l'algorithme doit trouver. Cette branche peut servir par exemple en bio-informatique pour découvrir des patrons d'ADN. Enfin, l'**apprentissage supervisé** cherche quant à lui la création de règles de décision à partir de données d'apprentissage contenant des exemples dit « étiquetés » (ce qui signifie qu'on connaît à priori la réalité) comme la reconnaissance de caractères manuscrits malgré la présence de différents styles d'écriture. Cette thèse se base particulièrement sur cette dernière approche.

En apprentissage supervisé, l'« **étiquette** » peut être une variable discrète ou continue. Dans le premier cas, lorsque l'étiquette est discrète, on parle de classification. L'objectif sera par exemple de déterminer simplement si le conducteur est éveillé ou somnolent. Dans le deuxième cas, dit de la régression, lorsque l'étiquette est continue, l'objectif cette fois sera plutôt de déterminer un niveau de somnolence ou d'éveil du conducteur. Nous nous plaçons dans le cadre de ce second cas. En effet, dans le cadre de la somnolence, la régression permettra d'obtenir une évaluation plus précise du niveau de somnolence.

Pour faire simple, l'objectif est d'apprendre une inférence d'une valeur de sortie y à partir d'observations x . y représente l'étiquette (la sortie) et x représente les **indicateurs** (ou *features* en anglais). La modélisation consiste à partir d'indicateurs mesurés de décrire y , mais pas d'expliquer y . Autrement dit, la fonction, ou ensemble de règles complexes construites, ne reproduit ou n'explique pas la « réelle » fonction au sens mathématique, derrière la relation entre x et y . La démarche de l'apprentissage automatique repose sur le fait qu'il n'existe pas de connaissance a priori sur cette relation entre x et y , elle ne combine pas d'une manière simple et directe toutes les entrées.

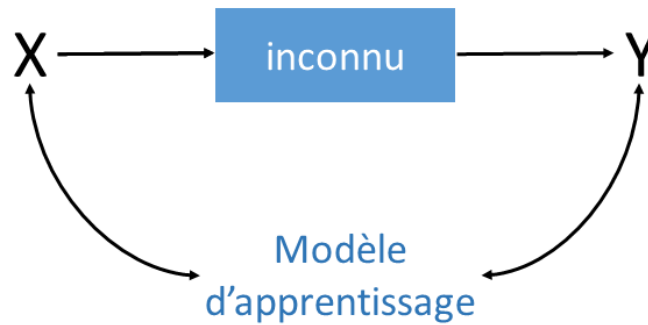


Figure 2-6 : Schématisation de la démarche suivie en apprentissage automatique pour inférer la valeur de y à partir de x . D'après Breiman, (2001)

Afin de déterminer cette méthode d'inférence, il existe plusieurs possibilités de catégorisation. Nous avons choisi de les exposer brièvement en fonction de leur finalité :

- **par analogie** avec l'algorithme le plus simple des k plus proches voisins dont l'objectif est de classer en fonction du ou des exemples le ou les plus « proches » ;
- **par approche probabiliste** en faisant des hypothèses sur les probabilités de distribution des classes en fonction des données enregistrées, avec l'exemple le plus connu de la méthode bayésienne dont la finalité est de déduire la probabilité d'un fait à partir de celles d'autres faits déjà évalués ;
- **par minimisation de l'erreur** avec comme exemple les réseaux de neurones artificiels (ANN : *Artificial Neural Network* en anglais) ;
- **par maximisation de la marge** avec les machines à vecteur de support en cherchant à augmenter la distance entre la frontière de séparation et les exemples les plus proches ;
- **par combinaison de tests élémentaires** en utilisant des arbres de décision ou des méthodes d'ensemble des arbres de décision, ou les forêts d'arbres décisionnels ou forêts aléatoires.

2.6.2 Étapes de développement d'un modèle

Le développement d'un modèle d'apprentissage automatique supervisé s'effectue en plusieurs étapes (adaptées selon Besson (2007) et Webb (2003)).

- 1) **La formulation du problème** : comprendre les objectifs.
- 2) **L'acquisition des données** : recueillir les mesures appropriées.

- 3) **L'examen préliminaire des données et nettoyage des données** : analyses descriptives pour comprendre la relation que peuvent avoir les sources de données avec le phénomène à prédire.
- 4) **L'extraction ou sélection d'indicateurs (*features*)** : sélectionner et créer les variables/indicateurs nécessaires à l'aide des mesures enregistrées à l'étape 2.
- 5) **La division du jeu de données** : souvent négligée, cette étape joue pourtant un rôle clé dans l'élaboration d'un modèle. Négligée, elle peut biaiser fortement l'analyse, la construction, l'interprétation d'un modèle, et sa capacité de généralisation. Elle a en effet joué un rôle très important dans le cadre de cette thèse. Elle consiste à diviser la base de données créée en différents jeux de données : un **jeu d'entraînement** qui servira à la classification ou régression de l'étape 6 et un autre jeu de données, **le jeu de test**, qui servira à l'étape 7. Ces points seront développés par la suite.
- 6) **La classification ou régression** : utiliser l'algorithme de classification (lorsque la cible y est catégorielle : éveillé ou somnolent) ou de régression (lorsque y est continu : état continu allant de « éveillé » à « extrêmement somnolent » par exemple). Le modèle est conçu à partir d'exemples issus de l'ensemble des données d'entraînement.
- 7) **L'évaluation des performances** : tester le modèle entraîné à l'étape 6 sur un autre ensemble de données non vu par le modèle : le jeu de test. L'évaluation se réalise à l'aide de métriques adaptées au problème.
- 8) **L'interprétation.**

Les étapes 5 et 7 vont être plus détaillées par la suite mais afin de comprendre l'importance de ces étapes le risque de sous-apprentissage et sur-apprentissage sera, dans un premier temps, développé.

2.6.3 Risque de sous et sur-apprentissages

Comme présentée succinctement précédemment, la généralisation d'un modèle consiste à connaître la capacité du modèle à prédire/estimer sur de nouvelles données. Le modèle est ainsi testé sur des données jamais vues précédemment. La division des jeux de données sera détaillée par la suite. Lors de la phase d'apprentissage du modèle et de ses paramètres, il est important d'éviter le **sur-apprentissage** et le **sous-apprentissage**.

Le sur-apprentissage (*overfitting* en anglais) se caractérise par des bonnes performances sur l'ensemble du jeu de données d'entraînement (c'est-à-dire une erreur faible) mais une capacité de

généralisation faible (c'est-à-dire une erreur élevée sur le jeu de test). Le modèle sera donc sensible à des variations sur les données. On explique le sur-apprentissage par le fait que le modèle est trop complexe par rapport aux données et/ou également par une quantité trop réduite de données. Quant au sous-apprentissage, les performances sont faibles avec les deux jeux de données, celui d'entraînement et de test (c'est-à-dire une erreur importante). Par rapport au modèle idéal, le sur-apprentissage fournit un modèle plus complexe que le modèle idéal alors que le sous-apprentissage fournit, lui, un modèle moins complexe. La difficulté de l'apprentissage automatique consiste donc à chercher ce modèle idéal, qui n'existe pas forcément dans l'absolu, et donc à rechercher un modèle se rapprochant de celui-ci. La figure 2-7 illustre ces phénomènes. Supposons que les points verts représentent les données d'entraînement et les carrés orange les données de test. L'objectif est d'apprendre la fonction $y = f(x)$. Les courbes en noirs représentent 3 polynômes du 1^{er} ordre, du 3^{ème} ordre et du 7^{ème} ordre. Celui du 7^{ème} ordre est « parfait », il capture tous les points. Cependant, il généralise mal, il est trop complexe, il représente donc le risque de sur-apprentissage. Nous pouvons également observer que le troisième ordre capture mieux les données que le premier. Le polynôme du troisième ordre se rapproche du modèle idéal alors que le polynôme du premier ordre n'est pas assez complexe, il représente le risque de sous-apprentissage. Afin de limiter le risque de sur-apprentissage et de sous-apprentissage, comme nous l'avons vu, la méthode repose dans la division du jeu de données en jeu d'entraînement et en jeu de test. On parle, alors, de compromis biais-variance (illustré dans la figure 2-7).

En apprentissage supervisé, le biais représente l'erreur provenant de données erronées dans le modèle ; la variance correspond à l'erreur due à des faibles variations sur les indicateurs du jeu de données. La figure 2-7 illustre ce compromis biais-variance, problématique centrale en apprentissage supervisé. Un modèle simple et peu complexe (risque de sous-apprentissage) aura une variance faible et un biais élevé sur les données d'entraînement et de test alors qu'un modèle complexe (risque de sur-apprentissage) aura une variance élevée mais un biais faible sur les données d'entraînement mais élevée sur les données de test. L'erreur sur le jeu d'entraînement décroît en fonction de la complexité du modèle alors que l'erreur sur le jeu de test, lui, décroît puis croît lorsque le modèle devient trop complexe. Tout l'enjeu de la modélisation réside par conséquent dans la recherche de ce modèle intermédiaire, à condition qu'il existe. La notion de complexité d'un modèle est généralement liée au nombre d'hyperparamètres utilisés par le modèle. Par exemple la complexité d'un polynôme de degré n est $n+1$ car il faut déterminer $n+1$ coefficients dont un lié à l'ordre 0.

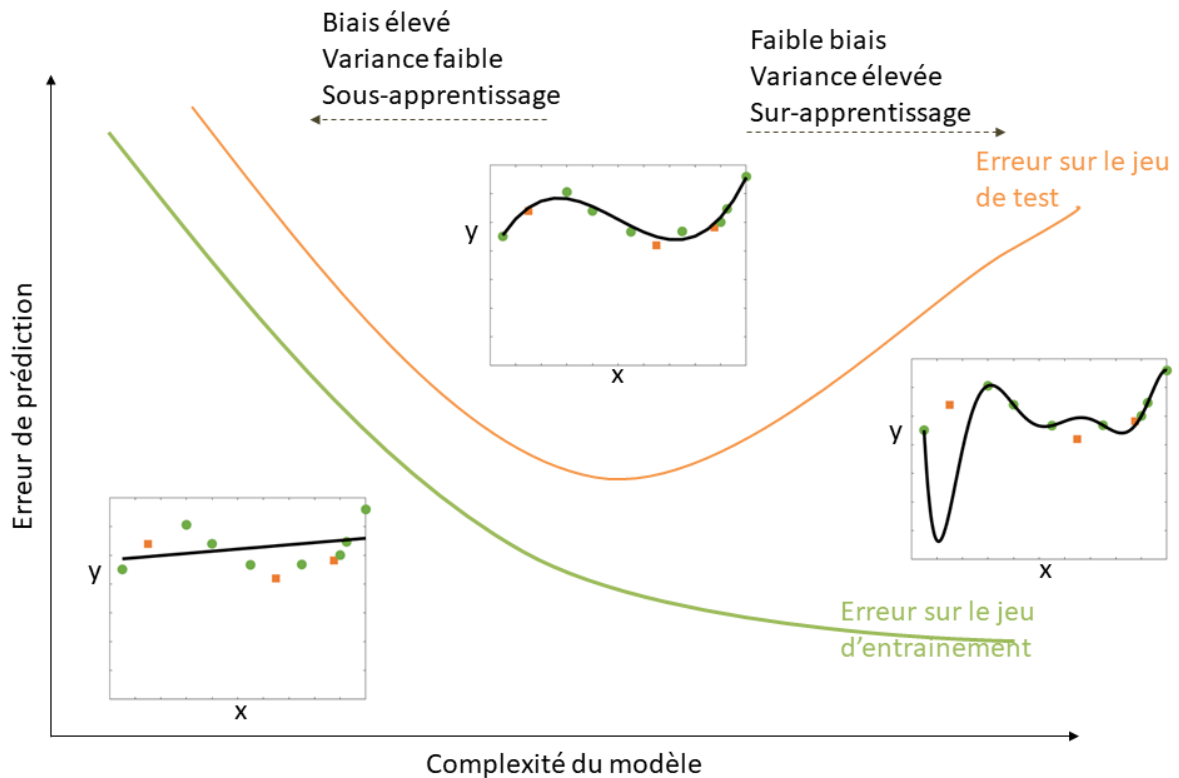


Figure 2-7 : Pour un problème de régression, illustration du compromis biais-variance : modèle trop simple à gauche (polynôme d'ordre 1), modèle trop complexe à droite (polynôme d'ordre 7) et un modèle intermédiaire entre les deux (polynôme d'ordre 3).

2.6.4 Division des jeux de données

Comme vu précédemment, dans l'objectif de trouver un modèle intermédiaire permettant une meilleure généralisation, il est nécessaire de diviser l'ensemble des données acquises en différents jeux de données : jeu d'entraînement et jeu de test. Il existe, néanmoins, différentes possibilités afin de diviser la base de données en jeu d'entraînement et jeu de test : la division simple entraînement-test, la validation croisée à k plis, la stratégie d'exclusion d'une donnée et la validation croisée par participant. Ces différentes méthodes de division des jeux de données vont être détaillées ci-dessous.

2.6.4.1 Division entraînement-test

La première méthode, la plus simple, consiste à diviser l'ensemble des données en deux ou trois jeux de données de manière aléatoire :

- le **jeu d'entraînement**, aussi appelé jeu d'apprentissage. Celui-ci sert à déterminer le ou les modèles. Il peut également se diviser en deux selon les cas : le **jeu d'entraînement**, et le **jeu de validation** sert quant à lui à définir et choisir les hyperparamètres clés.
- le **jeu de test**, aussi appelé jeu de généralisation. Cet ensemble de données n'a jamais été vu précédemment par le modèle. Il permet de connaître les performances de généralisation du modèle, savoir comment se comporte le modèle avec des données jamais vues précédemment.

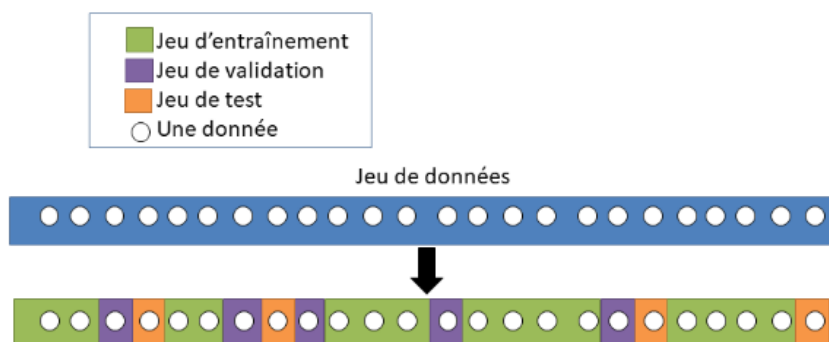


Figure 2-8 : Division du jeu de données en jeu de données d'entraînement, de validation et de test.

Dans la suite de cette thèse, le code suivant sera toujours utilisé : jeu de données global en bleu, jeu d'entraînement en vert, jeu de validation en violet, et jeu de test en orange.

2.6.4.2 Validation croisée à k plis

Lorsque la base de données n'a pas un volume très important, se présente le risque qu'avec une division simple, il ne soit pas possible de savoir si le modèle possède de bonnes performances de généralisation. Il existe donc la validation croisée à k plis afin de contrer la faible ou moyenne dimension du jeu de données disponibles. Cette méthode consiste à diviser le jeu de données en k sous-ensembles appelé k plis de même taille. Un modèle est entraîné sur k-1 sous-ensembles et sa performance est mesurée sur le k^{ème} sous-ensemble. L'approche est répétée k fois en faisant varier la partie servant à l'évaluation, le jeu de test.

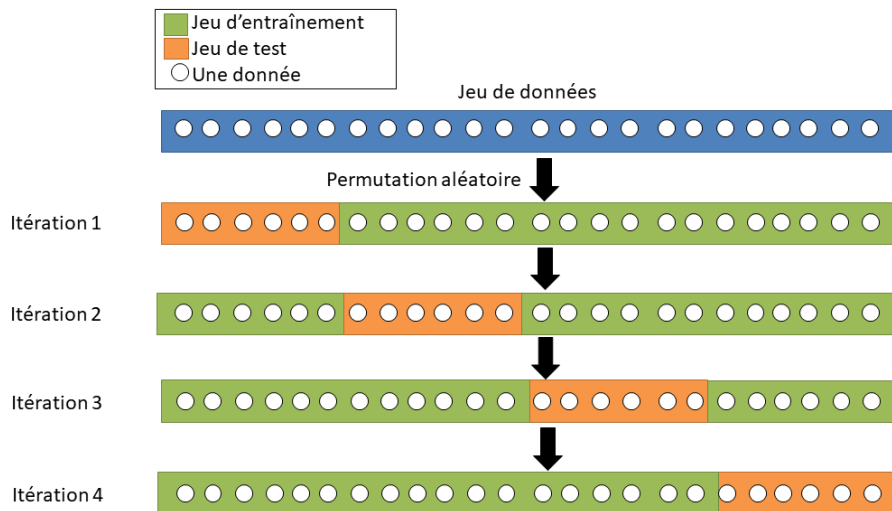


Figure 2-9 : Une cross-validation à 4 plis : Chaque point appartient à 1 des 4 jeux de test et aux 4 autres jeux d'entraînement

2.6.4.3 La stratégie d'exclusion d'une donnée

La stratégie d'exclusion d'une donnée (*Leave-One-Out strategy*) est une variante de la stratégie précédente où le nombre de sous-ensembles est égal au nombre d'échantillons présents dans le jeu de données. Chaque modèle est donc entraîné sur $n-1$ exemples et testé sur un exemple. Finalement, n modèles sont donc entraînés (avec n le nombre de données).

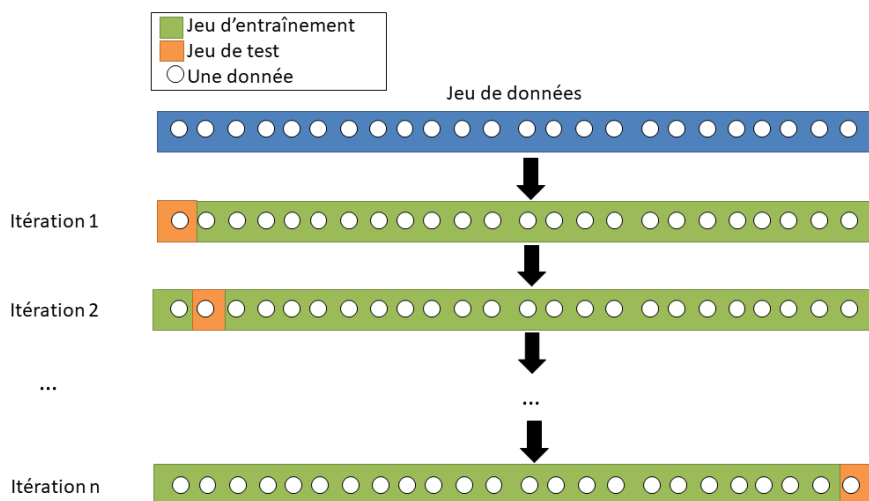


Figure 2-10 : La stratégie d'exclusion d'une donnée

2.6.4.4 La validation croisée par sujet

Dans notre cas d'usage particulier, qui est la conduite automobile et la modélisation de la somnolence au volant, nous pouvons nous demander si le modèle sera performant sur un nouveau conducteur, qui n'aura jamais été vu précédemment par le modèle. Les participants utilisés pour l'entraînement du modèle et pour le test vont alors dans ce cas varier. Chaque participant va, à son tour, être retiré du jeu d'entraînement et servira de jeu de test. Cette stratégie peut être appelée validation croisée inter-individuelle ou par sujet ou en anglais : *one-subject-left-out-strategy* ou *cross-subject-validation*. Elle est similaire à la validation croisée à k plis en considérant que chaque pli représente un participant/conducteur.

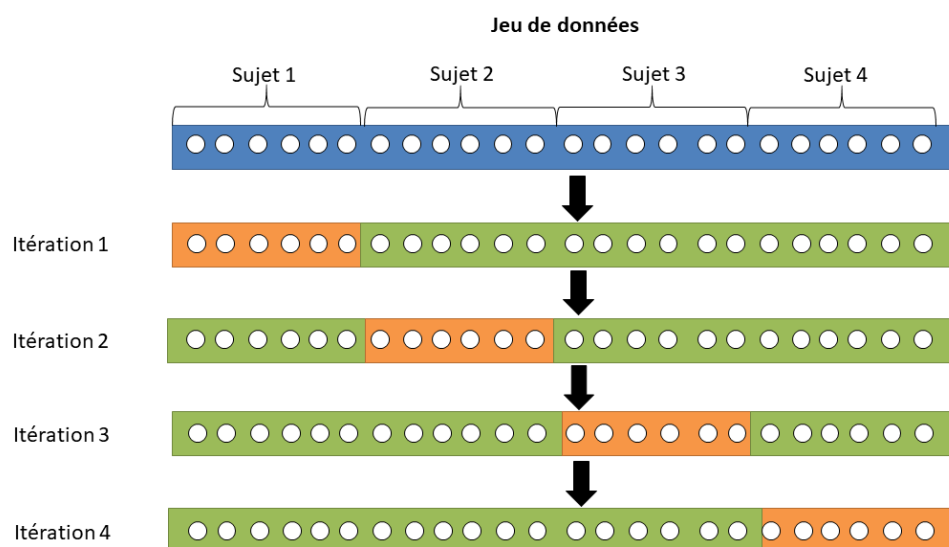


Figure 2-11 : Validation croisée inter- individuelle

En apprentissage automatique, la division des jeux de données entre le jeu d'entraînement et de test joue un rôle important aussi bien pour restreindre le risque de sous et sur-apprentissages que pour estimer correctement les performances de généralisation du modèle.

2.6.5 Application à la détection et la prédiction de la somnolence

Comme il en ressort de la littérature (Tango et al., 2009), il existe une tendance à choisir des techniques d'apprentissage automatique pour l'étude de la modélisation des comportements humains, qui sont, eux, non déterministes et fortement non linéaires. Les modèles non linéaires sont généralement utilisés pour leur pouvoir prédictif supérieur à celui de modèles linéaires, et leur capacité plus grande à extraire des relations plus complexes au sein de x pour prédire y . Un des avantages des modèles linéaires, malgré leur plus faible performance prédictive, est en l'occurrence leur interprétation, plus aisée. La détection et la prédiction de la somnolence ont souvent été réalisées avec des approches d'apprentissage automatique, et plus particulièrement avec des algorithmes supervisés. Afin de modéliser l'état d'un conducteur, l'apprentissage automatique est souvent employé. Cette technique présente, en effet, de nombreux avantages. Le modèle peut extraire des informations à partir de données bruitées et il est généralement plus robuste (Dong et al., 2011). Dans la littérature, de nombreuses méthodes d'apprentissage supervisé ont été utilisées afin d'estimer l'état dégradé de différents indicateurs.

2.6.5.1 Par analogie

La méthode la plus classique dans l'approche par analogie consiste, à l'aide d'une mesure de distance, à chercher la ou les données plus proche(s) dans la base de données d'entraînement, d'où le nom de k -plus-voisins. La méthode consiste à chercher le nombre de données à considérer ainsi que la mesure de distance. Par exemple, cet algorithme a été utilisé avec des données acquises sur un simulateur de conduite avec des mesures d'EEG, EOG et ECG. Les auteurs obtiennent des performances entre 82-84 % (Khushaba, Kodagoda, Lal, & Dissanayake, 2013). Une autre étude s'est focalisée sur l'utilisation de mesures extraites du pouls pour classer l'état de fatigue post et pré-conduite des conducteurs de véhicules (Chauhan, Saroliya, & Sharma, 2015).

2.6.5.2 Par approche probabiliste

En approche probabiliste, l'exemple le plus connu est la méthode bayésienne, une méthode d'inférence dont l'objectif est de déduire la probabilité d'un événement à partir d'autres événements. Elle s'appuie donc sur les probabilités conditionnelles. Les deux avantages de cette méthode sont qu'il existe des modèles dynamiques qui permettent de saisir la dynamique de la somnolence et qu'il est facilement possible d'insérer des probabilités concernant des éléments de contexte comme le bruit en plus de mesures physiologiques (Ji, Lan, & Looney, 2006; Yang et al., 2010). Cependant la plupart des études ont uniquement utilisé des mesures physiologiques (Lee & Chung, 2012). Cette méthode est

étudiée pour la fusion de données, néanmoins, la principale limite réside dans le fait qu'il est nécessaire de calculer ou de connaître des probabilités liées à des variations de mesures physiologiques ou à divers événements.

2.6.5.3 Par minimisation de l'erreur

Même si l'objectif poursuivi par toute méthode d'apprentissage automatique est de minimiser l'erreur, les moyens pour construire les modèles ne reposent pas dans tous les cas sur ce principe. La modélisation la plus utilisée et la plus connue, s'appuyant sur la minimisation de l'erreur, est basée sur les réseaux de neurones artificiels (ANN : *Artificial Neural Network* en anglais). L'avantage de cette modélisation est la possibilité de fusionner les sources de données et de se révéler robuste à la perte de signal (Samiee et al., 2014). Cette dernière étude a construit 3 réseaux en se basant sur trois sources de données différentes (position latérale, angle volant et patrons de clignements du conducteur) qui sont par la suite combinées. Le meilleur des cas permet une bonne classification à près de 94% de réussite. Une autre étude de Daza et al. (2014), comprenant des mesures provenant du véhicule (position latérale, angle au volant, erreur de cap, vitesse, freinage et accélération) et du conducteur (PERCLOS, fréquence et durée de clignements) a testé différentes combinaisons. Ces auteurs ont démontré que la fusion du PERCLOS avec un indicateur de la conduite permet d'obtenir les meilleures performances, soit près de 96% de bonne classification. Plus récemment, des méthodes utilisant de l'apprentissage profond sont utilisées (Hajinoroozi et al., 2015).

2.6.5.4 Par maximisation de la marge

Une autre méthode existe pour optimiser le modèle est de maximiser la marge, c'est-à-dire la distance entre la frontière de séparation/décision et les données les plus proches. C'est une généralisation de la régression linéaire logistique en cherchant la fonction qui optimisera la frontière de décision. Des exemples clé peuvent être les machines à vecteur de support ou les séparateurs à vaste marge (SVM, *support vector machine* en anglais). Afin d'aider la recherche de cette frontière de décision sous forme d'une séparation linéaire, les SVM utilisent également une fonction noyau pour projeter les données non séparables linéairement dans un autre espace dans lequel ce sera possible. En utilisant trois classes (très somnolent, somnolent et alerte), un SVM avec comme données d'entrées des mesures provenant de l'EOG, EMG et EEG, permet 100% de détection (30 cas) pour cet état critique, 86% pour somnolent (26 bien classés et 4 comme très somnolent) mais un taux bien plus faible pour la dernière condition 83% avec 5 cas sur 30 de faux négatifs. Les SVM n'ont également été utilisés qu'avec des mesures provenant de l'EEG (Shen, Li, Ong, Shao, & Wilder-Smith, 2008; Yeo et al.,

2009) ou qu'avec des mesures provenant de l'ECG (Li & Chung, 2013). Une étude plus originale a utilisé par ailleurs la parole comme données d'entrée et des SVM (Krajewski et al., 2009), mais est utilisable seulement quand la personne parle.

2.6.5.5 Par combinaison de tests élémentaires

Les méthodes proposées reposent sur l'utilisation d'une fonction complexe. Une autre branche a été développée en apprentissage automatique, elle consiste à utiliser un ensemble de tests élémentaires ou de fonctions plus simples et de les combiner pour fournir un algorithme présentant de bonnes performances. Par exemple, McDonald et al. (2013) ont utilisé des forêts aléatoires (ensemble d'arbres de décision) avec comme données d'entrée des mesures provenant de l'angle au volant. Un ensemble de ridge régressions multinomiales entraînées grâce à l'algorithme Adaboost (Vural et al., 2007) a été utilisé avec des mesures provenant des mesures oculaires et faciales et a été testé avec un apprentissage, avec un test intra-participant (96% de bonnes estimations avec apprentissage et test sur le même participant) et un test inter-participants (validation croisée inter-individuelle, 90% de bonnes estimations avec apprentissage sur un groupe de participants et test sur un autre). Néanmoins, la condition pour déterminer un état de somnolence est l'enregistrement de la minute précédant un crash, ce qui est déjà extrêmement critique pour la conduite et trop tardif pour présenter une utilité préventive. Un système expert a également été utilisé pour fusionner des informations provenant de l'EEG et d'une caméra (Picot, 2009). Par ailleurs, la méthode de Viola et Jones (Viola & Jones, 2001) s'appuyant sur plusieurs algorithmes en cascade a été de nombreuses fois utilisée pour détecter le visage et/ou les yeux d'une personne.

2.6.5.6 Modèles pour prédire la somnolence

La plupart des recherches a porté sur la détection de l'état dégradé, c'est-à-dire sur l'estimation à un certain instant t , même si elles utilisaient le terme « prédiction » (Chen, 2013; Hargutt & Kruger, 2001; Ji, Zhu, & Lan, 2004; Verwey & Zaidel, 2000). En apprentissage automatique, en effet, le terme est également utilisé dans le sens de déduire l'étiquette d'un objet non encore vu pendant la phase d'apprentissage. Certaines études ont, néanmoins, tenté de prédire la vérité quelques minutes avant en décalant d'une époque, lors de l'apprentissage, les données et la vérité (Kaida, Åkerstedt, Kecklund, Nilsson, & Axelsson, 2007) voire plus, allant jusqu'à 10 minutes (Larue et al., 2011). Une étude s'est intéressée à la quantité de données nécessaires pour prédire avec 20 secondes d'avance un épisode de somnolence (Murata, Ohta, & Moriwaka, 2016). Ces chercheurs ont montré qu'en utilisant les données entre 120 et 20 secondes avant l'évènement somnolent, les performances du modèle, avec

comme données d'entrée des données comportementales provenant du conducteur et du véhicule, étaient les plus importantes. Une autre étude s'est, quant à elle, intéressée à la prédiction du prochain micro-sommeil lorsqu'au moins un événement de ce type a déjà eu lieu (Watson & Zhou, 2016). Toutefois, la principale limite de ce travail est que le premier micro-sommeil ne peut pas être détecté. Par ailleurs, une étude récente a également modélisé a posteriori le temps restant avant un niveau de somnolence élevée à partir de mesures pouvant l'influencer telles que les habitudes de sieste, le moment de la journée, la température ambiante, la vitesse moyenne du véhicule, le pourcentage de temps passé à une faible vitesse à l'aide de modèle à durée de vie accélérée (approche cherchant à prendre en compte des facteurs pouvant accélérer le temps restant avant une « défaillance ») (Wang, Sun, Fang, Fu, & Stipancic, 2017).

La plupart des études considère le problème de l'estimation de l'état dégradé comme un problème de classification. Le conducteur est-il dans un état dégradé ou non ? Cependant, l'évolution de l'état du conducteur peut également être considérée comme un problème de régression, le conducteur passant par divers états continus. Or, le modèle de régression est rarement utilisé dans la littérature. Nous pouvons donc estimer que l'utilisation d'un modèle de régression permettra d'obtenir une dynamique plus fine de l'état du conducteur et de pouvoir définir différents seuils selon, par exemple, le niveau d'automatisation du véhicule ou la contre-mesure à proposer aux conducteurs.

2.6.6 Une méthode d'apprentissage automatique : les réseaux de neurones artificiels

Avant de décrire brièvement les réseaux de neurones artificiels (« *Artificial Neural network* », ou « ANN », en anglais), nous allons justifier pourquoi nous avons utilisé principalement cette méthode.

2.6.6.1 Choix d'un réseau de neurones artificiels

Pour la modélisation de la somnolence au volant, nous nous sommes orientés vers des réseaux de neurones artificiels. En effet, cette approche est connue pour sa capacité à résoudre des problèmes grâce à des exemples. Une comparaison de différentes méthodes a montré que les ANN permettent d'obtenir de meilleurs résultats pour la détection et prédiction de la somnolence (Larue et al., 2011). D'autre part, comme nous le verrons plus tard, nous avons considéré l'état du conducteur comme un état continu, les ANN sont très pratiques et étudiés pour la régression.

2.6.6.2 Présentation des réseaux de neurones artificiels

Un ANN est un algorithme d'apprentissage automatique se basant à l'origine sur le fonctionnement des neurones biologiques. Un neurone biologique est la cellule de base du système nerveux central. Son rôle est la réception, le traitement et la transmission d'influx nerveux. Un neurone reçoit des informations d'autres neurones par les synapses au niveau de ses dendrites ou son corps cellulaire et s'active ou non pour transmettre un signal via son axone et une nouvelle synapse vers d'autres neurones. Quant au neurone artificiel, celui-ci est en fait une fonction (appelée fonction d'activation ou de transfert) qui va avoir plusieurs entrées, et une sortie. La figure 2-12 établit la mise en correspondance entre un neurone biologique et un neurone artificiel.

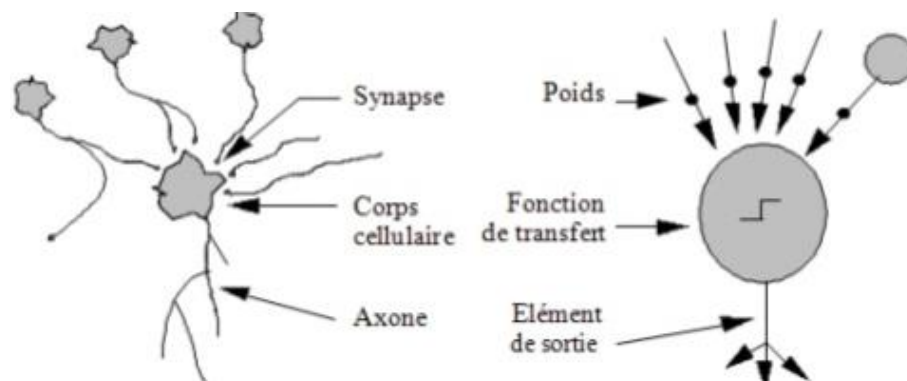


Figure 2-12 : Mise en correspondance neurone biologique / neurone artificiel

D'un point de vue plus formel, un neurone artificiel est une fonction non linéaire, paramétrée à valeur bornée. La sortie d'un neurone artificiel est une fonction d'intégration des entrées pondérées par leur poids (données d'entrée ou sorties d'autres neurones). Généralement, l'utilisation successive d'une somme pondérée et d'une fonction d'activation/transfert est utilisée. La fonction d'activation est généralement de forme sigmoïdale. La sortie d'un neurone peut être modélisée par la fonction suivante :

$$y = h\left(\sum_{i=0}^n x_i w_i\right) \text{ avec :}$$

- x_i les n entrées du neurone,
- w_i les poids correspondants aux entrées du neurone,
- h la fonction de transfert ou d'activation,
- w_0 représentant le biais, relié à un vecteur $x_0=1$. w_0 , le biais est un terme constant, il n'a aucun lien avec le biais d'un estimateur.

Un réseau de neurones artificiels est un ensemble de neurones rangés en couches : couche d'entrée, couches cachées et couche de sortie. La complexité et l'intérêt des réseaux de neurones résultent dans l'association en réseaux de plusieurs neurones en plusieurs couches. Ces couches ont pour conséquence qu'un réseau de neurone artificiel est perçu comme une « boîte noire » du fait de la multiplication des poids et de l'activation via des fonctions non linéaires.

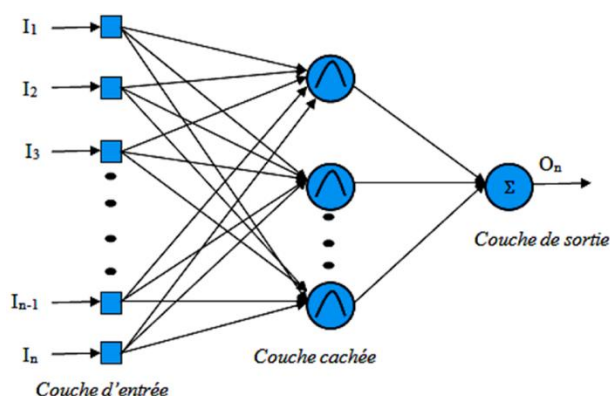


Figure 2-13 : Exemple d'un réseau de neurones artificiels à une couche cachée (I_1 à I_n correspond aux entrées du neurone)

Barron a montré en 1993 que la raison pour laquelle les ANN sont très souvent utilisés est que l'approximation dépend des paramètres ajustables de manière non linéaire. Elle est toutefois plus parcimonieuse que si elle dépend linéairement des paramètres. Cette propriété découle du fait que toute fonction bornée, suffisamment régulière, peut être approchée uniformément, avec une précision arbitraire, dans un domaine fini de l'espace de ces variables, par un réseau de neurones comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire (Hornik, Stinchcombe, & White, 1989), il en découle que la dimension de Vapnik-Chervonenkis³ des réseaux de neurones varie plus lentement en fonction du nombre de variable que les modèles linéaires, donc cela est plus avantageux.

La première étape consiste à définir la structure du réseau de neurones, c'est-à-dire le nombre de neurones par couche, le nombre de couches cachées et le type de réseaux de neurones (bouclé ou non bouclé par exemple). Il est difficile de savoir quelle sera celle optimale a priori. L'ordre de grandeur du nombre de neurones est lié aux nombres d'échantillons. Plus il y a de données, plus il y a de possibilités

³ Une mesure quantitative relative à la complexité du modèle. En classification, elle représente le nombre maximal de points pouvant être séparés sans faire d'erreur par une fonction indicatrice.

d'avoir des réseaux de neurones complexes et profonds. En général, une recherche en grille sur ces paramètres est réalisée en faisant varier ces paramètres. Le jeu de données de validation est alors utilisé afin de choisir la structure du réseau de neurones avec l'erreur la plus faible.

Après le choix d'une structure (nombre de couches et de neurones par couche), il est nécessaire d'estimer les valeurs des poids du réseau de neurones. Pour cela, la méthode consiste à optimiser une fonction de coût, généralement définie comme la fonction des moindres carrés, calculée sur l'ensemble d'apprentissage :

$$J(\mathbf{w}) = \sum_{i=1}^N (y_i - g(\mathbf{x}_i, \mathbf{w}))^2$$

Avec

- \mathbf{N} est le nombre d'exemples de l'ensemble d'apprentissage ;
- y_i la valeur de la grandeur à modéliser pour l'exemple i ;
- \mathbf{x}_i le vecteur des valeurs des variables d'entrée pour l'exemple i ;
- g la fonction représentant le réseau de neurones artificiels dépendant de \mathbf{w} les poids des différents neurones (pour simplifier, nous avons supposé que le réseau possède une seule sortie) ;
- \mathbf{w} les poids synaptiques.

Généralement, des méthodes de descente du gradient sont utilisées dans l'objectif de modifier les paramètres (les poids) du réseau jusqu'à ce qu'un minimum de la fonction de coût empirique soit atteint, ou qu'un critère d'arrêt soit satisfait.

Phénomène complexe, la somnolence au volant a été souvent modélisée grâce à l'utilisation de modèles d'apprentissage automatique. Ce travail s'appuiera sur les réseaux de neurones artificiels connus pour leurs performances.

2.7 Synthèse du chapitre

La somnolence, état intermédiaire entre le sommeil et la veille, représente un état dégradé pour la conduite et affecte donc potentiellement les performances de conduite. Cet état (qu'on considère généralement comme une variable cachée) ne peut pas être mesurée facilement et sans contraintes pour le conducteur. Ainsi, de nombreuses recherches se focalisent sur la caractérisation de cet état en étudiant les facteurs et les variations de diverses mesures. Ces facteurs et mesures sont multiples. L'utilisation combinée de plusieurs sources de données semble prometteuse afin de répondre à la variabilité intra et inter-individuelle. Pour prendre en compte ces informations, l'utilisation de modèles d'apprentissage automatique est couramment utilisée.

Trois limites ont été mises en valeur. Une grande majorité des études cible l'estimation à un instant t de l'état du conducteur avec différentes sources de données (détection) mais peu se focalisent sur une estimation temporelle du temps avant un état dégradé (prédiction). Par ailleurs, si nous considérons la somnolence comme un état intermédiaire, il est également plus pertinent de s'intéresser à des modèles de régression afin d'avoir une étude plus fine du niveau de somnolence. Finalement, les performances de généralisation d'un nouveau modèle inter-individuel, c'est-à-dire sur un nouveau conducteur ont également été peu étudiées. Cette thèse cherche des éléments de réponses sur ces trois limites.

Pour répondre à ces différentes limites, nous allons modéliser la somnolence à partir de mesures physiologiques, comportementales et à partir de son activité et performance de conduite. Ainsi, la méthodologie pour le recueil des données sera tout d'abord décrite (chapitre 3). Ensuite, nous étudierons la dynamique de la somnolence au volant (chapitre 4). Par la suite, nous procéderons à une modélisation du niveau de somnolence (détection) et du temps restant avant l'occurrence d'un niveau critique (prédiction) avec des réseaux de neurones artificiels (méthodologie : chapitre 5). Dans un premier temps, notre analyse s'orientera vers l'intérêt de mesures contextuelles telles que le temps de conduite et des informations personnelles sur le conducteur (chapitre 6), puis nous analyserons les performances de généralisation de ces modèles sur des nouveaux participants (chapitre 7) et proposerons une méthode pour augmenter ces performances consistant à adapter un ANN grâce à une faible quantité de données provenant d'un nouveau conducteur (chapitre 8).

Nous supposons qu'avec les mêmes données généralement utilisées pour détecter la somnolence au volant il est possible de prédire de manière temporelle quand un état dégradé sera présent. Afin d'améliorer les performances du modèle, nous supposons également que le temps de conduite et des informations personnelles sur le conducteur seront bénéfiques. Finalement, nous supposons qu'entraîner un ANN sur un groupe d'individus puis ensuite personnaliser cet ANN sur un nouveau conducteur jamais vu par le modèle grâce à une faible quantité de données permettra d'améliorer les performances sur ce conducteur.

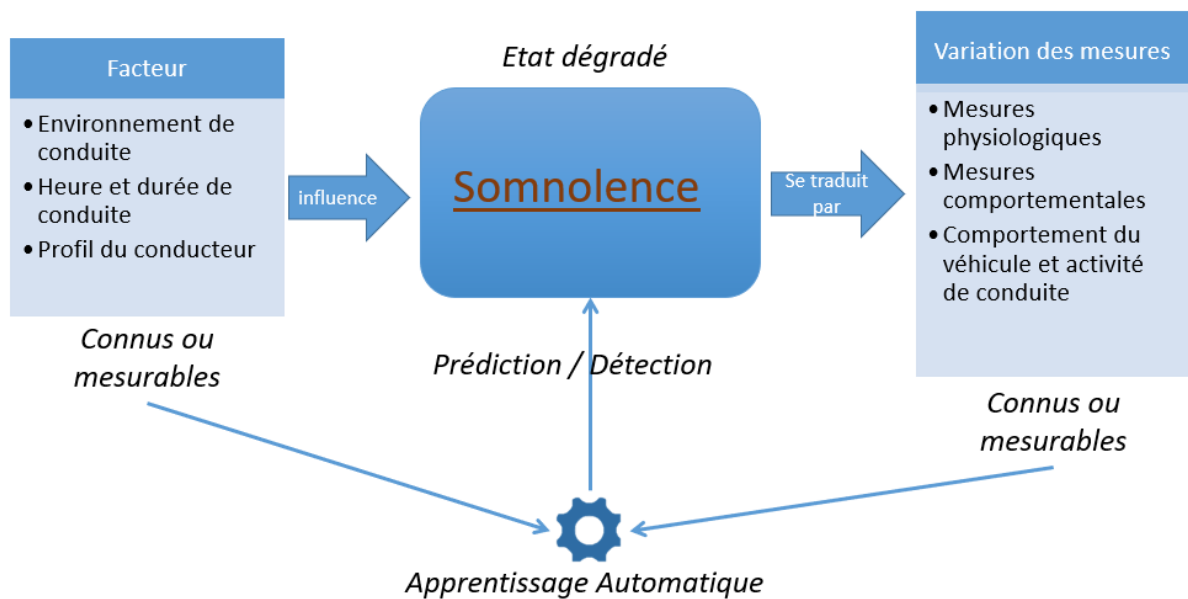


Figure 2-14 : Conclusion de la revue de littérature et méthodologie mise en place

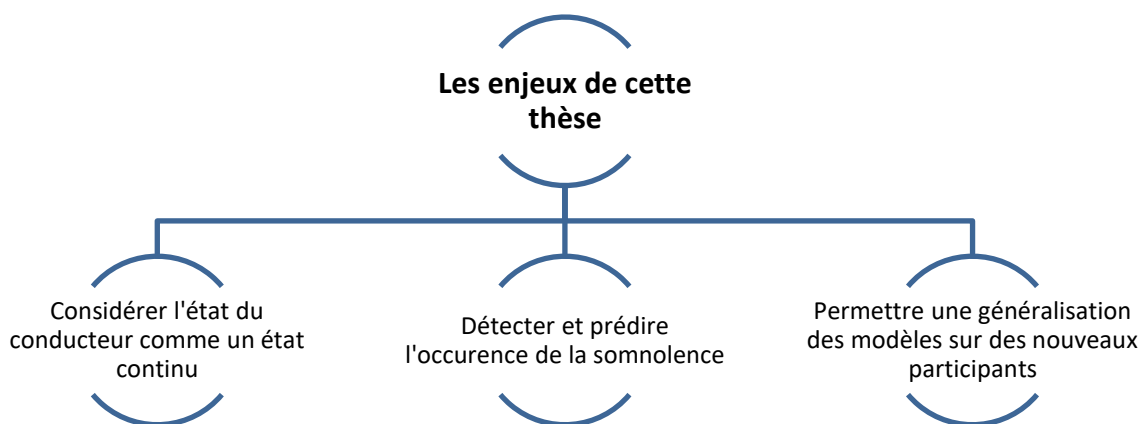


Figure 2-15 : Les enjeux de ce travail de thèse

Chapitre 3 : Méthodologie – Recueil de données

Les études qui composent cette thèse se basent essentiellement sur de la modélisation à l'aide de réseaux de neurones artificiels. Il s'agit donc d'un travail à dominante informatique. Cependant, comme toujours lorsqu'on traite d'apprentissage automatique, il est nécessaire d'apporter un soin méticuleux à l'acquisition des données, à leur mise en forme et leur « nettoyage », phase de pré-traitement et de transformation des données appelée en anglais « *features engineering* ». Comme nous l'avons vu dans la première partie de ce mémoire, le choix des « bonnes » variables est crucial, le résultat de la modélisation dépendant au premier chef des choix qui seront faits en amont. Cette partie du mémoire détaille donc quels signaux et comment ils ont été enregistrés dans l'objectif de les utiliser pour modéliser par la suite la somnolence. Tout d'abord, les informations et critères d'inclusion des participants seront décrits, puis seront précisés le protocole d'acquisition des données pendant la conduite sur simulateur, ainsi que la méthode choisie afin de déterminer le niveau supposé réel de somnolence (*ground truth*). La méthode concernant une première analyse sur la dynamique de la somnolence (chapitre suivant) sera également explicitée. La méthodologie relative à la modélisation sera, quant à elle, décrite dans le chapitre 5.

3.1 Participants

Un groupe de 21 participants a été inclus dans l'étude (âge moyen et écart-type : 24.09 ± 3.41 ; 11 hommes et 10 femmes). Une population jeune a été choisie dans l'objectif de limiter la variabilité inter-individuelle. Par ailleurs, comme souligné dans la partie 2.2.7 « Créneau d'âge », une population jeune est une population à risques concernant la somnolence au volant. Les critères d'inclusion sont : être détenteur d'un permis de conduire valable depuis au moins 6 mois ;

- ne pas être porteur de corrections optiques telles que des lunettes (afin de ne pas gêner l'oculométrie);
- ne pas être sensible au mal de simulateur, évalué grâce à un questionnaire (Golding, 1998) ;
- ne pas être sujet à une somnolence diurne, évaluée grâce à un score sur l'échelle d'Epworth inférieur à 14 (Johns, 1991). Selon cette échelle, un score en dessous de 8 signifie que la personne n'a pas de dette de sommeil, un score entre 9 et 14, la personne a quelques signes de somnolence diurne, un score au-dessus de 15, la personne présente des signes importants de somnolence diurne.

Par ailleurs, le jour de l'expérience, les participants n'étaient pas autorisés à consommer alcool, café ou thé. Il leur était demandé également s'ils avaient pris un quelconque médicament. De plus, avant l'expérience, les participants ont rempli un questionnaire permettant d'obtenir des informations

personnelles. Ces dernières sont détaillées dans le tableau 3-1. Ces informations permettent d'obtenir aussi bien des renseignements sur les habitudes de conduite que des informations relatives à leurs sommeil et leur typologie circadienne pour savoir si ils sont plus du matin ou du soir avec le questionnaire de Horne et Östberg (Horne & Östberg, 1975)

Tableau 3-1 : Informations recueillies sur les participants

Informations recueillies
Âge
Score sur l'échelle de somnolence d'Epworth (Johns, 1991)
Fréquence de consommation de café (jamais, rarement, 1 à 2 tasses par jour, plus de 2 par jour)
Fréquence de conduite (occasionnellement, plusieurs fois par mois, semaine, jour)
Nombre de kilomètres par an
Score sur le questionnaire de typologie circadienne de Horn et Östberg (Horne & Östberg, 1975)
Nombre d'heures de sommeil la nuit précédente
Qualité habituelle du sommeil (évaluée sur une échelle de 1 à 10)
Avez-vous déjà eu l'impression de vous endormir au volant ? (OUI/NON)
Genre

3.2 Protocole mis en place

Les participants ont conduit après le déjeuner entre 105 et 120 minutes sur un simulateur de conduite statique dans une pièce où l'air conditionné maintenait une température fixée à 24° Celsius (voir Fig. 3.1 pour une image du simulateur). En effet, comme vu précédemment (2.2.2), il existe deux créneaux horaires critiques de 2 h à 6 h et de 14 h à 16 h. C'est pourquoi nous avons choisi ce second créneau intermédiaire.

La route et le trafic ont été créés et générés avec SCANeR Studio®. Pendant la conduite, les signaux ont été enregistrés avec des logiciels et matériels spécifiques :

- SCANeR Studio® pour l'activité et les performances de conduite, logiciel de simulation de conduite automobile permettant de créer l'environnement de conduite, la simulation de conduite et de récupérer les données relatives à la conduite,

- FaceLAB® pour les mesures sensorimotrices, dispositif de suivi oculaire composé de deux caméras séparées et d'un émetteur,
- Système Biopac® MP150 et Acqknowledge® pour les mesures physiologiques : enregistrement de l'ECG, de la respiration et de l'EDA (à cause de pertes importantes du signal EDA, ces données ne seront pas exploitées dans la suite de cette étude). Le Biopac® MP150 permet l'acquisition des signaux et AcqKnowledge® représente le logiciel permettant de visualiser les signaux acquis.

Les différents signaux enregistrés ont été par la suite transformés pour créer les variables d'entrées des modèles. Ces transformations seront détaillées par la suite dans le chapitre 5 – Méthodologie de la modélisation de la somnolence.

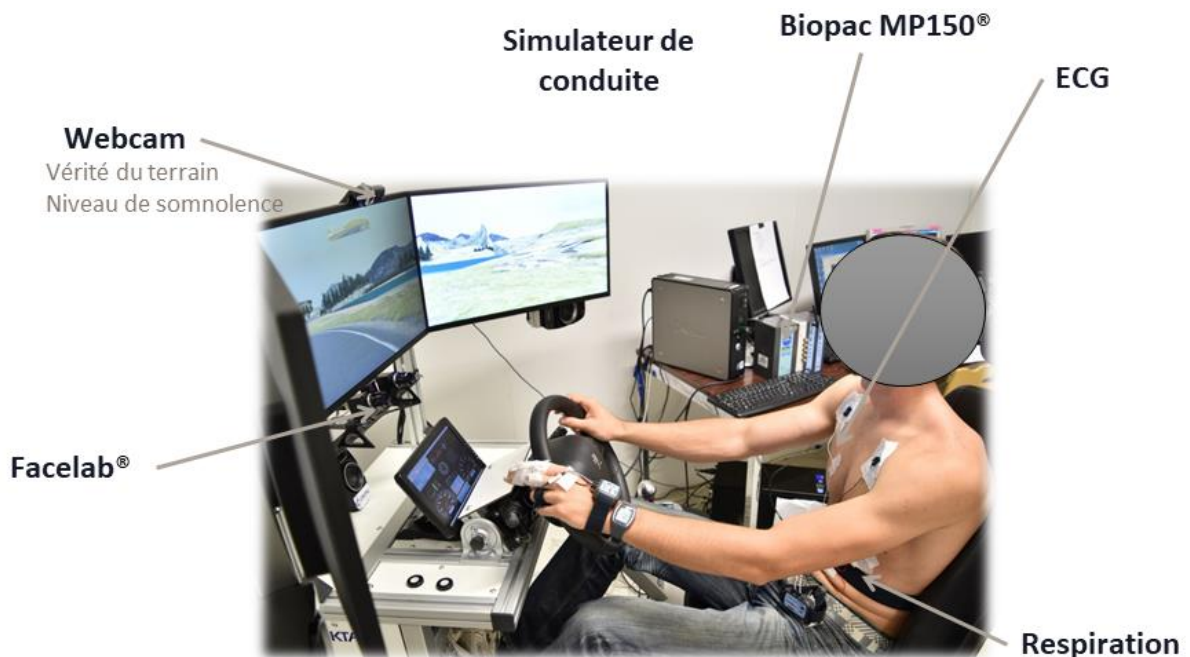


Figure 3-1 : Dispositif et différentes données enregistrées

Les participants ont tout d'abord conduit sur une autoroute pendant environ 95 minutes puis ils ont été amenés à quitter l'autoroute, à conduire 5 minutes sur une route de campagne avant d'arriver en ville pour conduire encore 5 minutes. Pendant la partie sur autoroute, à environ 2/3 du parcours, des voitures apparaissaient par la droite sur une voie d'insertion et disparaissaient quelques kilomètres plus loin. Le trafic a été ajouté dans l'objectif de changer le niveau de somnolence du conducteur. Des caractéristiques plus précises sont présentes dans le tableau 3-2. L'autoroute fictive a été modélisée

en s'appuyant sur une portion de l'autoroute A86. Cette portion a ensuite été juxtaposée 6 fois pour allonger le temps total de trajet. Une visualisation des écrans du simulateur est présente en figure 3-2. Cet environnement routier monotone a été choisi dans l'objectif d'amener les conducteurs à être somnolents.

Tableau 3-2 : caractéristiques du parcours

Type de route	Caractéristiques	Limitation de vitesse	Trafic
Autoroute	Peu de virages	130 km/h	22 voitures apparaissent et rentrent sur l'autoroute au 2/3 pour disparaître quelques kilomètres plus loin
Campagne	Intersections + courbes	70 km/h	Nul
Ville	Intersections + courbes	50 km/h	Faible variant selon la portion de route



Figure 3-2 : Vue des trois écrans à l'entrée de la ville et le tableau de bord

3.3 Détermination du niveau de somnolence

Comme nous l'avons vu précédemment, l'apprentissage supervisé nécessite de connaître, pour les données utilisées pour l'apprentissage lui-même, l'état (dans notre cas le niveau de somnolence) que l'on souhaite voir associé aux données. Dans notre cas, la détermination du niveau de somnolence est basée sur une évaluation subjective, effectuée par deux évaluateurs indépendants, à partir d'une vidéo des participants pendant l'enregistrement de la conduite. Cette évaluation s'appuie sur l'échelle TOR (Belz et al., 2001) (voir tableau 3-3 pour plus de détails). Deux évaluateurs ont noté le niveau de somnolence à l'aide de cette échelle, et ce pour chaque minute de chaque vidéo. Chaque segment a été évalué sur une échelle de 0 (alerte) à 4 (extrêmement somnolent). La moyenne des deux évaluations a été utilisée comme vérité de terrain (*ground truth*), c'est-à-dire comme véritable niveau de somnolence du conducteur à l'instant t. Nous avons défini l'état du conducteur comme étant dégradé lorsque celui-ci était supérieur ou égal à 1,5. Cette valeur intermédiaire a été choisie parce que si la moyenne est 1,5, cela signifie qu'un des évaluateurs a estimé ce niveau à 2 (modérément somnolent), déjà critique pour la conduite pendant que l'autre l'a estimé à 1 (légèrement somnolent). Cette limite, fixée à 1,5 a été choisie à partir d'une étude utilisant cette même échelle, mais qui a, elle, déterminé la limite entre « non somnolent » et « somnolent » entre les niveaux 1 et 2 (0 et 1, qui correspond à alerte et légèrement somnolent, correspond pour eux à non somnolent ; 2, 3 et 4, qui correspond de modérément à extrêmement somnolent, correspond pour eux à somnolent) (McDonald et al., 2013).

Tableau 3-3 : Observer Rating of Drowsiness

Niveau	Description
0 : Alert	Un tonus facial normal, clignements rapides des yeux, coup d'œil court
1 : Peu somnolent	Un coup d'œil un peu plus long, clignements des yeux plus aussi rapides
2 : Moyennement somnolent	Maniérismes (se frotter les yeux ou le visage, se gratter, contorsion faciale, bouger dans son siège), fermeture plus lente des yeux, baisse du tonus facial, yeux vitreux, regard fixe
3 : Très somnolent	Fermeture des paupières de 2 à 3 secondes accompagnée souvent de roulement des yeux vers le haut ou sur les côtés, focus des yeux pas effectué correctement, yeux qui louchent, tonus facial bas, peu d'activité apparente, mouvements larges et isolés (ou ponctués), réorientation de la tête depuis une position penchée
4 : Extrêmement somnolent	La personne tombe de sommeil, fermeture des paupières prolongée de 4 secondes, ou plus, et période prolongée de manque d'activité, large mouvement ponctué

La Kappa Statistique Kappa est souvent utilisée afin de valider la fiabilité inter-évaluateur. Néanmoins, comme le niveau de somnolence est considéré comme un état continu, il est plus judicieux dans ce cas d'utiliser le test de corrélation linéaire de Pearson (Wiegand, McClafferty, McDonald, Hanowski, 2009).

3.4 Analyse de la dynamique de la somnolence

Avant de mettre en place la modélisation de la somnolence, nous nous sommes intéressés à sa dynamique, c'est-à-dire à la variation du niveau de somnolence au cours du temps de conduite. Pour cela, nous avons choisi deux variables reflétant la dynamique de la somnolence dans l'objectif de la caractériser :

- le niveau maximum de somnolence atteint sur toute l'expérience, permettant de représenter l'état le plus critique obtenu par le participant ;
- le temps d'atteinte du niveau critique, défini à 1.5, permettant d'estimer la vitesse individuelle de dégradation de la vigilance pour chaque participant.

Ces deux variables ont été adoptées car elles permettent de représenter deux éléments importants relatifs à la somnolence : la criticité à la somnolence (première variable) et la temporalité (deuxième variable).

Afin de tenir compte de la variabilité des comportements des participants, l'analyse portera alors sur la classification ou groupement des participants selon ces deux variables qui ont été auparavant centrées et réduites. Des méthodes de groupement (« *clustering* » en anglais) telles que l'algorithme des k-moyennes et une arborescente hiérarchique afin de fixer k, le nombre de groupes, ont été mises en place. La classification par arborescente hiérarchique est une technique permettant d'organiser des observations, dans notre cas les participants, en les regroupant de façon hiérarchique selon leurs distances. Ces liens hiérarchiques sont ensuite représentés dans une figure appelée dendrogramme. Cette figure permet de représenter la liaison entre les observations et le niveau de proximité/dissimilarité avec la hauteur des branches. La configuration du dendrogramme nous permettra de définir le nombre de groupes appelé k. Ensuite afin de valider les k groupes, nous procéderons à l'algorithme des k-moyennes. C'est une méthode de partitionnement de données qui permet de diviser les observations en k groupes en cherchant à minimiser la distance entre les observations d'un même groupe.

Pour donner suite à la formation des groupes, les variables : âge, score sur l'échelle de somnolence d'Epworth, fréquence de consommation de café, fréquence de conduite, consommation de café, nombre de kilomètre par an, score sur le questionnaire de Horn et Östberg, nombre d'heures de sommeil et qualité de sommeil ont été soumises à une analyse de la variance (ANOVA) avec comme groupe les k groupes précédemment construits. Lorsque cela sera significatif, les effets et les interactions seront analysés avec des tests post-hoc avec la correction de Bonferroni. Quant aux variables catégorielles (le genre et si la personne s'est déjà endormie au volant), elles seront analysées à l'aide d'un test de Chi-deux.

Chapitre 4 : Dynamique de la somnolence

Ce chapitre présente des résultats relatifs au niveau de somnolence des participants au cours de l'expérimentation. La dynamique de la somnolence sera analysée selon les deux variables décrites précédemment afin de regrouper des participants entre eux. Nous essayerons par la suite de voir s'il est possible d'expliquer ces différents groupes à l'aide des informations personnelles. Une discussion suivra ces résultats.

4.1 Résultats

4.1.1 Evolution du niveau de somnolence au cours de l'expérimentation

Cette partie s'intéresse donc au niveau de somnolence au cours du temps. La fiabilité de l'évaluation des deux évaluateurs a été estimée grâce à une corrélation linéaire de Pearson ($R = 0,71$ et $p=0,00$). Quant à la comparaison des deux évaluateurs (Fig. 4.1), nous pouvons observer que la moyenne des notations est globalement très proche. Le niveau moyen, noté par l'évaluateur 1, est très légèrement plus élevé que le niveau moyen noté par l'évaluateur 2.

Le graphique 4-1 représente, minute après minute, la moyenne du niveau de somnolence des 21 participants, noté par les deux évaluateurs. Nous observons que le niveau de somnolence croît, comme attendu en environnement monotone, puis diminue avec l'apparition du trafic, puis augmente de nouveau avant de diminuer avec la sortie de l'autoroute. Il est important de noter que cette moyenne représente la moyenne des profils des conducteurs et non le profil moyen. En effet, nous pouvons observer également une importante variabilité du niveau de somnolence au cours du temps. Cette variabilité est faible au début mais augmente ensuite car tous les conducteurs ont commencé à conduire en étant parfaitement alerte.

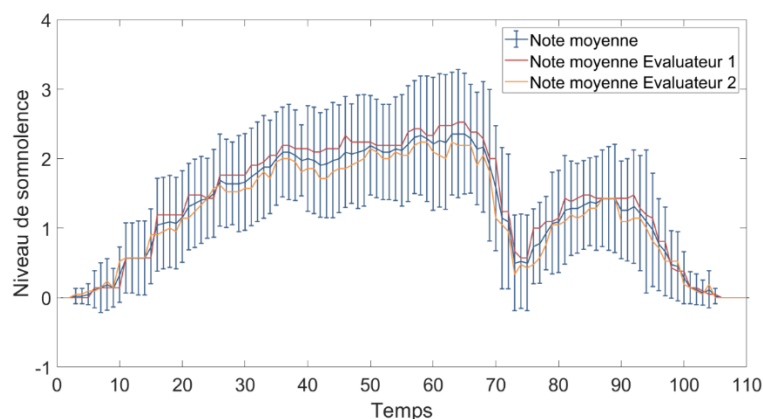


Figure 4-1 : Moyenne et écart-type du niveau de somnolence au cours du temps

4.1.2 Variables reflétant la dynamique de la somnolence

Cette partie cherche à regrouper les conducteurs selon les deux variables décrites précédemment et à expliquer ces groupes à l'aide d'informations personnelles sur les participants. Les deux variables sont, pour rappel, les suivantes :

- le niveau maximum de somnolence atteint sur toute l'expérience permettant de représenter l'état le plus critique obtenu par le participant ;
- le temps d'atteinte du niveau critique, défini à 1,5, permettant d'estimer la vitesse personnelle individuelle de dégradation de la vigilance pour chaque participant.

La figure 4.2 représente le dendrogramme. Sur la figure 4.3, chaque point représente un conducteur selon les deux variables décrites précédemment, coloré en fonction de l'algorithme des k-moyennes.

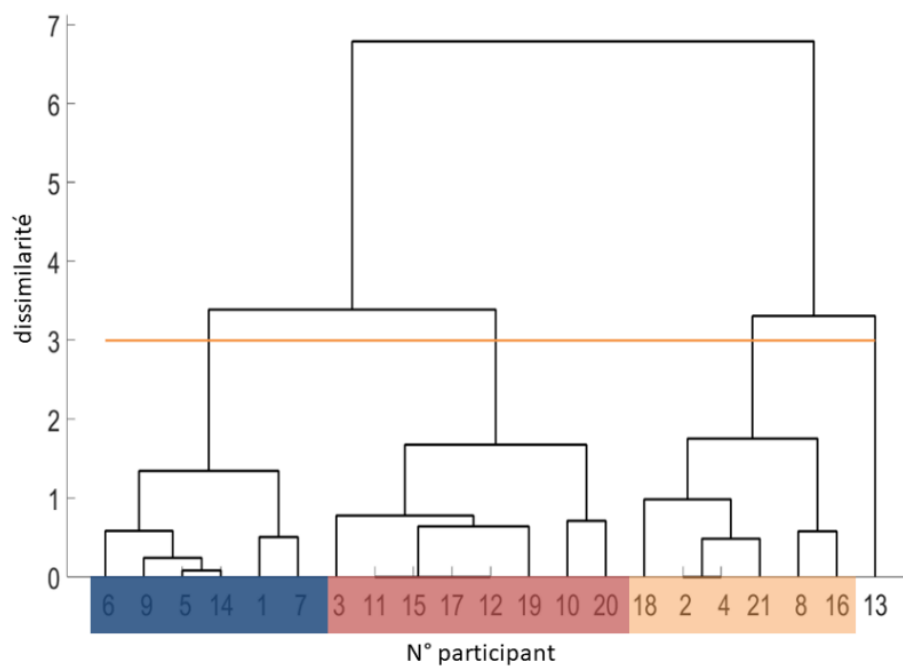


Figure 4-2 : Dendrogramme - Arbre de classification hiérarchique ascendant

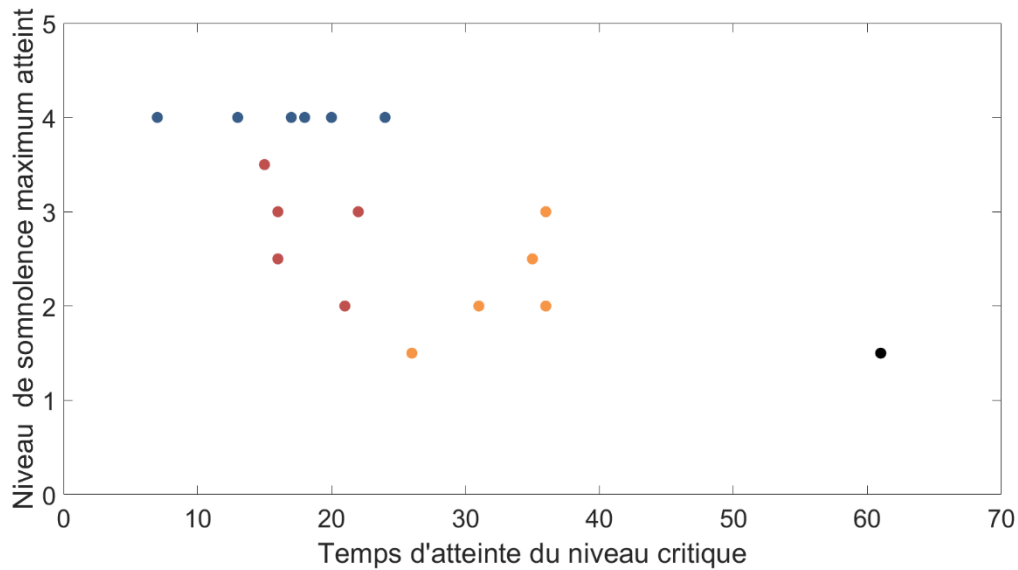


Figure 4-3 : Répartition des participants selon le niveau de somnolence maximum atteint et le temps de conduite restant avant le niveau critique (en minute). Colorisation des participants selon la classification résultant du dendrogramme et k-moyenne (Fig. 4-2).

Un premier examen de la cartographie des participants selon cette classification révèle une répartition des participant en trois groupes distincts. Cet examen révèle également que nous sommes en présence d'un participant atypique (le participant 13). La principale différence (comportementale) avec les autres participants est que celui-ci a mis beaucoup de temps à atteindre le niveau critique de somnolence (défini à 1,5), soit 60 minutes. Les algorithmes de « *clustering* » ou regroupement, révèlent donc l'existence de 3 groupes de participants, discriminés en fonction de leur niveau de criticité et leur vitesse d'endormissement. Il est important de noter que les deux méthodes de *clustering* donnent le même résultat. Le premier groupe en haut à droite de la Fig. 4-3 (points de couleur bleue) représente un groupe amené à somnoler très vite et à atteindre un niveau de somnolence très important. Le troisième groupe (points orange), moins critique, atteint un niveau plus faible de somnolence et met également plus de temps avant d'atteindre le seuil critique. Quant au deuxième groupe intermédiaire (points en rouge), celui-ci se situe entre les deux. Les numéros attribués à ces groupes (1, 2 et 3) peuvent donc faire référence à un niveau de criticité, de risque, de susceptibilité ou de dangerosité. En effet, le premier groupe est très critique car il va rencontrer davantage de périodes de somnolence que les autres groupes. Quant au participant isolé, le questionnaire initial révèle qu'il jouit d'un sommeil de très bonne qualité et de plus, le debriefing

suyvant la conduite lui a fait avouer qu'il avait tout fait pour ne pas somnoler durant l'expérience, faisant des efforts importants pour rester éveillé malgré la monotonie du contexte de conduite.

Une ANOVA a été réalisée afin de comparer les moyennes des différents groupes pour les variables quantitatives des informations sur les conducteurs (âge, score sur l'échelle de somnolence d'Epworth, fréquence de consommation de café, fréquence de conduite, consommation de café, nombre de kilomètres par an, score sur le questionnaire de Horne et Östberg, nombre d'heures de sommeil et qualité de sommeil). Un test de Chi-deux a été appliqué aux variables catégorielles (le genre et si la personne s'est déjà endormie au volant). L'ANOVA révèle qu'aucune des informations personnelles continues ne permet statistiquement de différencier les groupes résultant des classifications pratiquées. Néanmoins, nous constatons que la fréquence de conduite semble avoir en moyenne une tendance différente pour les divers groupes ($F(2,18) = 3,375$; $p = 0,082$). Une analyse post-hoc, avec la correction de Bonferroni, amène à constater que les groupes 2 et 3 manifestent une fréquence de conduite différente ($p=0,033$; voir Figure 4-4).

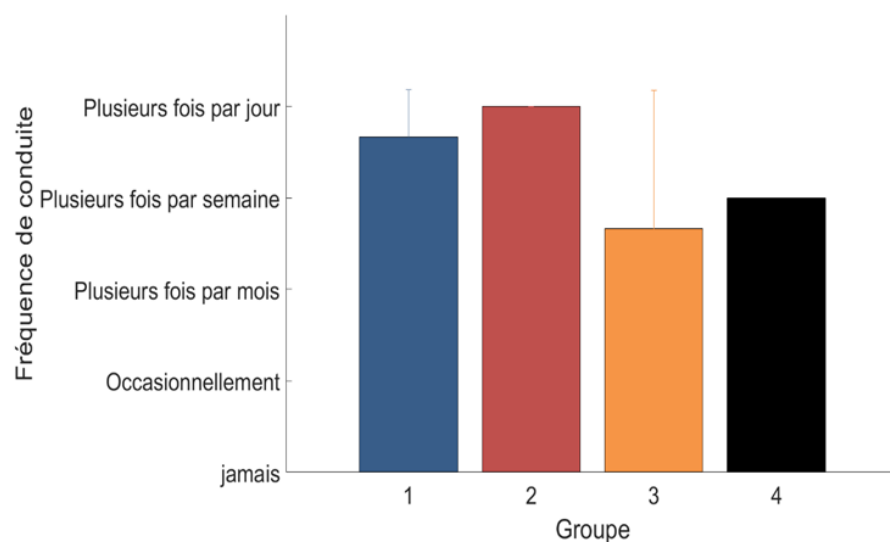


Figure 4-4 : Moyenne et écart-type de la fréquence de conduite regroupés par groupes.

Concernant les variables catégorielles, un test de Chi-deux a montré que, selon les groupes, la répartition de la variable « Vous êtes-vous déjà endormi(e) au volant ? » est différente selon les groupes ($X^2 = 6,50$; $p = 0,038$; voir tableau 4-1). Le tableau de contingence est présenté dans le tableau 3-2.

Tableau 4-1 : Tableau de contingence des groupes et de la variable catégorielle « Déjà dormi au volant »

	1	2	3	4	
Non	0	2	4	0	6
Oui	6	6	2	1	15
Effectif total	6	8	6	1	21

4.2 Discussion

Sans contrôler la durée du sommeil ou sans provoquer un manque de sommeil chez les conducteurs participants, nous avons adopté le choix d'être plus proches des conditions observables dans le quotidien en étudiant la conduite après une nuit de sommeil normale et habituelle pour les participants. Cependant, il est important de noter que certains conducteurs ont atteint un niveau de somnolence très élevé et même parfois critique (*extrêmement somnolent*) avec seulement des facteurs potentiellement à risques tels que l'heure de conduite ou la monotonie. Comme attendu, nous avons observé une très large variabilité inter-individuelle concernant le niveau de somnolence atteint au cours du temps de conduite et du contexte. La variabilité inter-individuelle et l'effet du contexte vont donc être un élément de discussion.

4.2.1 Variabilité inter-individuelle

Les performances neurocomportementales varient grandement d'un individu à un autre (Van Dongen, Baynard, et al., 2004). Toutefois, avec une privation de sommeil importante, seulement quelques participants voient leurs performances de conduite hautement dégradées (Philip et al., 2004). Dans cette étude, nous constatons également d'importantes différences dans le niveau de somnolence des conducteurs. À l'aide d'une ANOVA ou d'un test Chi-deux, nous avons voulu étudier si des informations personnelles, facilement accessibles, pourraient permettre de comprendre et/ou d'expliquer ces groupes. Tous les participants du groupe 1 ont déjà eu l'impression de s'endormir au volant. Quant à la différence des moyennes des groupes concernant la fréquence de conduite, celle-ci semble trop faible pour être en mesure d'expliquer les écarts de profils de somnolence entre les groupes. D'autres informations auraient pu être recueillies *a priori* telles que les traits de personnalité,

la dépendance au champ visuel (Thiffault & Bergeron, 2003b) ou en l'occurrence, la sensibilité à la caféine (Rétey et al., 2006). Ces informations auraient pu permettre d'expliquer les différences entre les groupes. Toutefois, dans le souci de proposer une méthode adaptable par la suite à des groupes plus importants, il n'a pas été choisi ici d'acquiescer ces informations car, en contexte de grand nombre, elles auraient été alors contraignantes.

4.2.2 Effet lié au contexte

En moyenne, pour l'ensemble des participants, est observé un profil similaire concernant la dynamique de la somnolence. En effet, au début de la conduite, les participants sont alertes et totalement éveillés. Puis, en raison de la monotonie de la conduite sur autoroute (Larue et al., 2011; Pierre Thiffault & Bergeron, 2003b) et du créneau horaire choisi (début d'après-midi) (Rossi, Gastaldi, & Gecchele, 2011), les participants deviennent somnolents à différents niveaux allant de légèrement somnolent (niveau 1) à extrêmement somnolent (niveau 4). Par la suite, avec l'arrivée de nombreux véhicules, amenant variabilité et stimulation, les conducteurs voient leur niveau de somnolence baisser. En ville et avec le trafic, la charge visuelle et les actions possibles du conducteur augmentent, cela pourrait expliquer la variation du niveau de somnolence (Ahlström, Anund, Fors, & Åkerstedt, 2018). Par ailleurs, à la fin de la conduite, l'expérimentateur demande au conducteur s'il a l'impression que le trafic l'a réveillé. 20 participants sur 21 ont répondu « oui ». Finalement, il est constaté qu'avec l'arrivée de la conduite en ville, les conducteurs retrouvent tous un niveau « alerte ». La dynamique de la somnolence moyenne décrit donc seulement une moyenne des profils, mais si nous regardons, au cas par cas, il est observé des variations nettes entre les niveaux et entre les phases.

4.3 Conclusion

Comme attendu les conditions expérimentales ont amené un niveau de somnolence très critique chez certains participants. Les variations du type de route (route de campagne ou de ville) et du trafic ont provoqué des différences dans la dynamique de la somnolence. Ainsi, nous pourrions penser que dans le cas où les véhicules seraient autonomes sur autoroute, mais que le conducteur devrait reprendre la main pour la sortie d'autoroute pour conduire en ville, il y a alors des chances pour que celui-ci redevienne naturellement alerte. Cependant, il est à noter que des accidents causés par la somnolence ont lieu également en ville (Philip et al., 2010).

Cette expérience a mis en a valeur une forte variabilité inter-individuelle, difficilement explicable avec des informations personnelles basiques. Toutefois, différentes études ont pu montrer qu'il existe de nombreux facteurs de variation de la sensibilité à la somnolence (Thiffault & Bergeron, 2003a ; Rabat et al., 2016 ; Van Dongen et al, 2004), qui n'ont pas été mesurés ici. Par ailleurs, cette expérience ne s'est pas intéressée à la variabilité intra-individuelle pour laquelle un participant devrait reconduire de nombreuses fois avec les mêmes conditions expérimentales. De plus, cette importante variabilité dans le niveau de somnolence atteint par chaque participant, à chaque minute, amène à penser qu'un modèle basé seulement sur le temps écoulé et les facteurs prédisposant à la somnolence au volant n'est pas assez fiable. Il est donc intéressant d'utiliser des mesures indirectes pour modéliser la somnolence au volant en temps réel.

Chapitre 5 : Méthodologie de la modélisation de la sommolence

Ce chapitre présente la méthodologie relative à la modélisation de la somnolence à l'aide de réseaux de neurones artificiels. Tout d'abord le pré-traitement et la transformation des signaux bruts afin de créer les variables d'entrée des modèles seront détaillés. En effet, dans le domaine de l'apprentissage automatique, une importance particulière et un soin méticuleux sont donnés à la création des variables qui serviront d'entrées aux modèles. Cette partie sera suivie d'une description de la méthode d'entraînement utilisant des réseaux de neurones artificiels. Pour cela, la méthode d'optimisation et le critère d'arrêt de l'entraînement seront détaillés. Etant donné le fait qu'un des enjeux de la thèse est de détecter et prédire la somnolence à partir de variables indirectes, les sorties des modèles attendus seront donc présentées.

5.1 Entrées des modèles : création des variables

Les données brutes enregistrées avec les différents systèmes d'acquisition, SCANer[®], faceLAB[®] et Biopac[®], ont été filtrées, nettoyées et agrégées afin de servir d'entrées des modèles.

5.1.1 Filtrage et nettoyage des signaux

Les données de faceLAB[®] ou SCANer[®] contiennent une perte de données caractérisée par un changement soudain des données à une valeur fixe aberrante. Pour les données enregistrées par faceLAB[®], ces pertes de données sont principalement causées par des mouvements extrêmes de la tête qui perturbent complètement l'acquisition de la modélisation 3D. Pour éliminer le bruit lié à ces pertes de données une interpolation linéaire entre le début et la fin de chaque perte de données a été effectuée. Quant aux pertes de données issues de SCANer[®], celles-ci sont principalement causées par une impossibilité de calculer une mesure.

Les artefacts concernant l'ECG ont été enlevés de manière non automatique. Pour ce faire, nous avons créé le tachogramme pour chaque participant. Sur celui-ci, nous avons identifié ensuite la perte ou la modification du signal dans le tachogramme et avons corrigé les données ECG de manière manuelle en nous basant sur le document suivant ("233 - Variabilité du taux cardiaque - Préparation des données pour l'analyse | BIOPAC" nd). Le PPG (Pulse-Plethysmography) a également servi dans ce cas spécifique.

5.1.2 Création des variables agrégées (transformation)

La fréquence d'échantillonnage était différente pour chaque source d'information et pour la vérité de terrain (le véritable niveau de somnolence du conducteur). La fréquence d'échantillonnage du niveau de somnolence est déterminée par la littérature. La fréquence de FaceLAB® est fixée par le logiciel. Quant à SCANeR®, la contrainte a été fixée par le logiciel, celle-ci devait être inférieure à 30Hz afin que toutes les données soient enregistrées. Nous avons choisi 1000 Hz pour l'enregistrement des mesures physiologiques afin d'être le plus précis sur le pic de l'ECG. Ces différentes fréquences sont répertoriées dans le tableau 5-1.

Tableau 5-1 : Sources de données et fréquences d'échantillonnage

Sources	Niveau de somnolence	Biopac®	SCANeR®	FaceLAB®
Fréquences d'échantillonnage en Hz	1/60	1000	10	60
Choix libre	✗	✓	✗ Inférieur à 30 Hz	✗

Après l'enregistrement de tous les signaux, il est nécessaire de ramener les données enregistrées à la même fréquence d'échantillonnage. Dans notre cas, nous avons choisi la plus petite valeur : 1/60Hz correspondant à 1 échantillon toutes les minutes. Par conséquent, nous avons utilisé des fonctions d'agrégation comme l'écart-type et la moyenne ou autres mesures spécifiques (voir tableau 5-2). Dans un objectif d'homogénéisation des différences individuelles, la moyenne des 5 premières minutes (état normal) a été soustraite pour chaque variable. Ainsi, la nouvelle variable représente la variation comparée à un état normal. Finalement, pour optimiser l'apprentissage, chaque variable a été normalisée pour obtenir toutes valeurs dans l'intervalle [-1 ; 1]. Le choix de ces variables s'est fait selon le cadre théorique présenté précédemment.

Tableau 5-2 : Variables extraites par systèmes d'enregistrement

Mesures physiologiques extraites du Biopac®	Mesures comportementales extraites de FaceLAB®	Mesures physiologiques extraites de SCANeR®
Rythme cardiaque : moyenne et écart-type (battement par minute)	Durée des clignements en secondes (moyenne et écart-type)	Distance latérale de la proche ligne par rapport au centre en mètres (moyenne et écart-type)
Svlf : Signal à très basse fréquence (0.0-0.04 Hz)	Fréquence des clignements (moyenne et écart-type) (par minute)	Temps pour franchir la ligne la plus proche (moyenne et écart-type)
Slf : Signal à basse fréquence (0.04-0.15 Hz)	PERCLOS (moyenne et écart-type) (% de fermeture de l'œil)	Angle du volant (moyenne et écart-type)
Shf : Signal à haute fréquence (0.15-0.4 Hz)	Position de la tête en x (moyenne et écart-type)	Vitesse de l'angle au volant (moyenne et écart-type)
Svhf : Signal à très haute fréquence (0.4-3.0 Hz)	Position de la tête en y (moyenne et écart-type)	Entropie de l'angle au volant
Le rapport sympathique (Slf / (Svlf + Slf + Shf))	Position de la tête en z (moyenne et écart-type)	Nombre de changements de direction du volant par minute (calculé par rapport à l'angle du volant)
Le rapport vagal (Shf / (Svlf + Slf + Shf))	Rotation de la tête en x (moyenne et écart-type)	Angle de la pédale d'accélération (moyenne et écart-type)
Le rapport sympathique-vagal (Slf / Shf)	Rotation de la tête en y (moyenne et écart-type)	Déplacement latéral du centre du véhicule par rapport au centre de la voie (moyenne et écart-type)
Rythme respiratoire (moyenne et écart-type par minute)	Rotation de la tête en z (moyenne et écart-type)	Vitesse du véhicule (km/h) (moyenne et écart-type)
	Fréquence des saccades (moyenne et écart-type) (par minute)	Nombre de fois où le conducteur franchit une ligne par minute

5.2 Modélisation : exploitation d'un réseau de neurones artificiels

Dans le cadre de ce travail, nous nous sommes intéressés à des réseaux de neurones non bouclés, dits statiques. L'information arrive par les entrées dans le réseau de neurones pour se diriger vers les sorties sans « retour en arrière ».

La programmation a été effectuée à l'aide de la neural network toolbox sur MatlabR2013a (Beale, Hagan, & Demuth, 1992). Nous nous sommes orientés vers un modèle de régression plutôt que vers des modèles de classification car les sorties sont continues (voir partie suivante). Après le choix d'une structure, il est nécessaire d'estimer les valeurs des poids du réseau de neurones. Pour cela, la méthode consiste à optimiser une fonction de coût, généralement définie comme la fonction des moindres carrés, calculée sur l'ensemble d'apprentissage.

$$J(\mathbf{w}) = \sum_{i=1}^N (y_i^p - g(\mathbf{x}_i, \mathbf{w}))^2$$

Avec :

- N = nombre d'exemples de l'ensemble d'apprentissage
- y_i^p = valeur de la grandeur à modéliser pour l'exemple i
- x_i = vecteur des valeurs des variables pour l'exemple i
- g = fonction représentant le réseau de neurones artificiels dépendant de \mathbf{w}
- w = poids des différents neurones (pour simplifier, nous avons supposé que le réseau possède une seule sortie).

Généralement, des méthodes de gradient sont utilisées dans l'objectif d'optimiser le ANN, c'est-à-dire modifier les paramètres (les poids) du réseau jusqu'à ce qu'un minimum de la fonction de coût empirique soit atteint ou qu'un critère d'arrêt soit satisfait. La méthode d'optimisation puis les critères d'arrêt vont être présentés ci-dessous.

5.2.1 Méthode d'optimisation choisie : algorithme de Levenberg-Marquardt

Nous avons utilisé l'algorithme de Levenberg-Marquardt (Levenberg, 1944). En effet, selon la documentation de la neural network toolbox que nous avons utilisée pour la programmation, cet algorithme est recommandé pour les problèmes de régression et dans le cas où sont utilisés des réseaux de neurones de petites ou moyennes tailles. En effet, il est rapide et précis pour converger. Il permet de converger vers un minimum local ou éventuellement global d'une fonction continue dérivable deux fois. Dans Hagan & Menhaj (1994), la théorie et les applications sont détaillées. Ces auteurs ont montré également que cet algorithme est de 10 à 100 fois plus rapide que la méthode traditionnelle de rétropropagation de la descente du gradient.

L'objectif de l'algorithme de Levenberg-Marquardt consiste à déterminer les poids des neurones \mathbf{w} en cherchant à minimiser la fonction suivante :

$$\sum_{i=1}^N (y_i^p - g(\mathbf{x}_i, \mathbf{w}))^2 = F(\mathbf{w})$$

La procédure de l'algorithme est itérative et consiste à mettre à jour les poids \mathbf{w} grâce à une nouvelle estimation à chaque itération avec $\mathbf{w} = \mathbf{w} + \mathbf{q}$. Afin de déterminer \mathbf{q} , g est approchée en étant linéarisée grâce au développement de Taylor du premier ordre :

$$g(\mathbf{w} + \mathbf{p}) \approx g(\mathbf{w}) + J \cdot \mathbf{q},$$

avec J la matrice Jacobienne

A un minimum, nous avons la dérivée qui s'annule, c'est-à-dire $\nabla_{\mathbf{q}} S = 0$, en dérivant et en posant $\mathbf{y} = g(\mathbf{w} + \mathbf{p})$, nous avons :

$$(J^T J) \mathbf{q} \approx J^T (\mathbf{y} - g(\mathbf{p}))$$

Il est ainsi possible d'avoir aisément \mathbf{p} en inversant $(J^T J)$. L'algorithme de Levenberg consiste à se rapprocher de cette équation, en l'« amortissant » un peu.

$$(J^T J + \mu \text{diag}(J^T J)) \mathbf{q} \approx J^T (\mathbf{y} - g(\mathbf{p}))$$

Le facteur d'amortissement positif μ est ajusté à chaque nouvelle itération.

Ce processus afin de mettre à jour les poids est effectué jusqu'au critère d'arrêt suivant.

5.2.2 Critères d'arrêt de l'entraînement des réseaux de neurones artificiels

Les critères d'arrêt utilisés pour interrompre l'optimisation de la fonction du coût, et donc l'apprentissage, sont les suivants :

- la magnitude du gradient a atteint un minimum (10-7)
- les performances ont atteint un minimum (0)
- le nombre d'époques a atteint un maximum (103)
- l'erreur du jeu de validation croît plus de 6 fois.

Plus concrètement, ce dernier critère d'arrêt cité est celui utilisé le plus souvent. Il permet d'éviter le sur-apprentissage et d'améliorer les performances de généralisation du modèle grâce à un arrêt précoce de la mise à jour des poids des neurones. Le principe de ce critère consiste à diviser l'ensemble du jeu de données en trois jeux distincts : le jeu d'apprentissage, le jeu de validation et le jeu de test.

- Le jeu d'entraînement est utilisé pour mettre à jour les poids des neurones.
- L'erreur sur le jeu de validation diminue lors de l'entraînement du réseau de neurones. Cependant, l'erreur sur ce jeu augmente lorsqu'augmente le risque de sur-apprentissage sur l'ensemble de données d'entraînement. Ce jeu permet également de fixer la structure du réseau, c'est-à-dire le nombre de neurones dans les couches cachées.
- Quant au jeu de test, celui-ci n'est pas utilisé pendant la phase d'entraînement car il sert à estimer l'erreur de généralisation.

La division concernant ces jeux de données variera en fonction du chapitre de résultat. Elle sera présentée avant chaque session de résultat.

5.3 Sortie des modèles à estimer : détection et prédiction de la somnolence

La finalité de ces travaux, comme dit précédemment, est non seulement de détecter la somnolence, mais également de la prédire, c'est-à-dire estimer dans combien de temps un certain niveau de somnolence risque de survenir. La création de deux modèles sera donc réalisée : un modèle dit de « détection » et un modèle de « prédiction ».

5.3.1 Niveau de la somnolence au cours de l'expérimentation

Comme détaillé précédemment (3.3 Détermination du niveau de somnolence), le niveau de somnolence utilisé est basé sur une évaluation subjective effectuée sur une vidéo des conducteurs par 2 évaluateurs indépendants. Cette évaluation se base sur l'échelle TOR (Belz et al., 2001). Les 2 évaluateurs ont noté chaque minute à partir de la vidéo de chaque participant. Chaque segment d'une minute a été évalué sur une échelle de 0 à 4 (alerte à extrêmement somnolent). La moyenne des 2 évaluations (effectuées par deux évaluateurs indépendants) a été utilisée comme vérité de terrain. Lorsque le modèle sera dédié à de la détection, sa sortie sera le niveau de somnolence allant de 0 à 4 avec un pas de 0.5 pour chaque minute (correspondant aux cas où les avis des évaluateurs diffèrent d'une unité).

5.3.2 Prédiction du temps restant avant le niveau de somnolence critique

Quant au modèle de prédiction, sa sortie sera le temps restant avant qu'un certain niveau critique survienne. Par exemple, si le niveau critique arrive à la 30ème minute, avec les données de la 30ème minute, la sortie du modèle à estimer sera 0, avec les données de la 20ème minute, la sortie du modèle sera alors 10. Ce niveau critique a été défini à 1,5 car il signifie que l'un des 2 évaluateurs a noté le segment en tant que 2 (modérément somnolent), degré déjà critique et dangereux pour la conduite. En effet, une autre étude utilisant la même échelle de notation a discriminé les classes alertes et somnolentes de cette manière (McDonald et al., 2013) :

- alerte : alerte + légèrement somnolent
- somnolent : modérément + très + extrêmement somnolent.

Ces entrées, modèles et sorties seront ceux utilisés dans les chapitres 6, 7 et 8.

Chapitre 6 : Détection et prédiction de la somnolence au volant en utilisant des réseaux de neurones artificiels

Ce chapitre présente l'article « *Detection and prediction of driver drowsiness using artificial neural network models* » publié dans la revue « *Accident analysis and prevention* » dans une édition spéciale « *Managing fatigue* ». Son objectif est double :

- Peut-on prédire l'arrivée de la somnolence avec les mêmes données d'entrées que celles utilisées pour la détection de la somnolence dans la littérature ?
- L'ajout de variables tels que le temps de conduite depuis le début de l'expérience et les informations personnelles recueillies à propos des participants permettent-ils d'améliorer les performances des modèles ?

La méthodologie pour recueillir les données est décrite dans le chapitre 3. La méthodologie servant à modéliser la somnolence est quant à elle explicitée dans le chapitre 5 « Méthodologie de la modélisation de la somnolence ». Avant l'insertion de l'article de la revue « *Accident analysis and prevention* », nous allons détailler la division des jeux de données et les variables utilisées.

6.1 Division des jeux de données

La base de données, formée par les données enregistrées sur les 21 participants pendant l'expérience, est divisée en 3 jeux de données : jeux d'entraînement (70%), jeux de validation (15%) et jeu de test (15%). Ce processus de division sera répété 10 fois aléatoirement. Chaque jeu a une fonction différente.

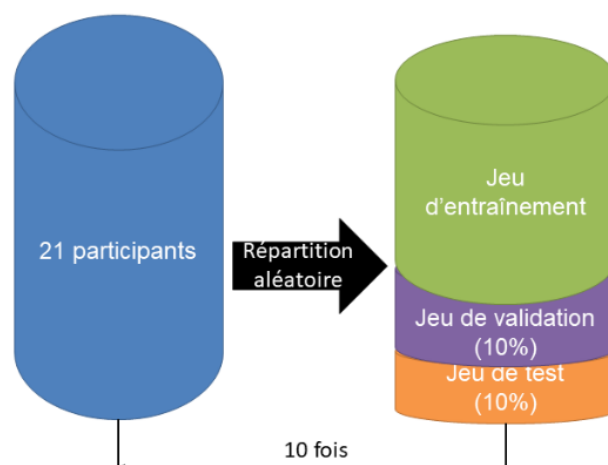


Figure 6-1 : Division des jeux de données pour la détection et prédiction de la somnolence au volant en utilisant des réseaux de neurones artificiels

6.2 Sources d'information du jeu de données

Par ailleurs, les sources d'information présentes dans les jeux de données ont également varié. Nous avons testé plusieurs combinaisons de variables possibles avec :

- ajout et suppression du « temps de conduite » (1 et 0 dans les tableaux de résultats ; le temps écoulé depuis le début de la conduite) ;
- ajout et suppression des informations sur les participants (1 et 0 dans les tableaux de résultats).

Pour ces 4 combinaisons, plusieurs cas de figure ont été envisagés :

- seulement les variables comportementales ;
- seulement les variables physiologiques ;
- seulement celles du véhicule ;
- combinaison complète des variables.

Soit un ensemble de 16 jeux de données différents, détaillés dans la figure suivante.

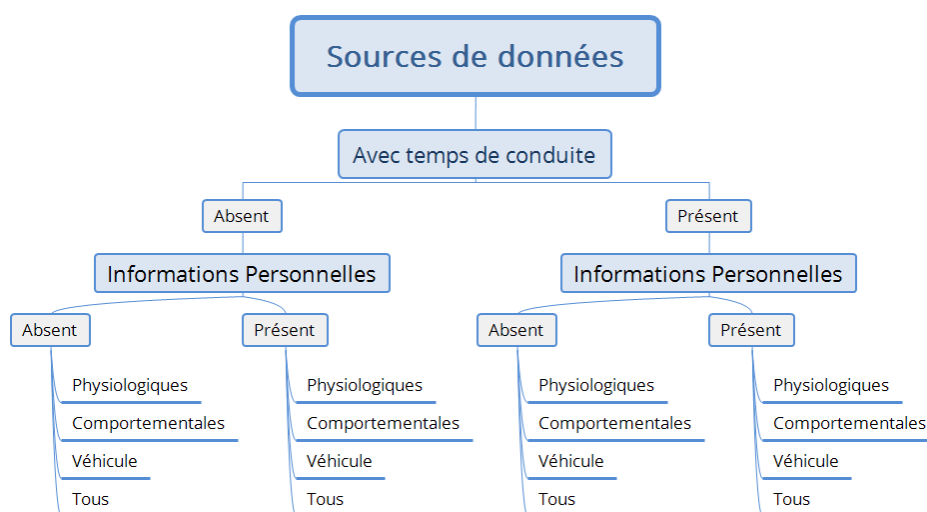


Figure 6-2 : Sources de données pour la détection et prédiction de la somnolence au volant en utilisant des réseaux de neurones artificiels

6.3 Problématique du premier article

Dans le cadre des travaux de cette étude générale sur l'état du conducteur, nous avons recueilli des informations provenant de différentes sources : données physiologiques, comportementales, l'activité et les performances de conduite ainsi que des informations psychologiques sur le conducteur.

L'objectif de cette étude est de développer et d'évaluer des modèles utilisant des réseaux de neurones artificiels (ANN) afin d'estimer le niveau de somnolence (modèle de détection) et de prédire le temps d'occurrence avant un état dégradé (modèle de prédiction). Différents ensembles de données utilisant des sources variées d'informations seront testés afin de déterminer la combinaison optimale. Notre première hypothèse est de montrer qu'il est possible de prédire quand un état dégradé se produira en utilisant des indicateurs généralement utilisés pour la détection de la somnolence. Nous supposons également que l'ajout d'informations personnelles telles que le temps de conduite et les informations sur les participants améliorera les résultats.

Ces hypothèses ont été validées grâce aux résultats de l'article suivant.

6.4 Detection and prediction of driver drowsiness using artificial neural network models



Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap



Detection and prediction of driver drowsiness using artificial neural network models

Charlotte Jacobé de Naurois^{a,b,*}, Christophe Bourdin^a, Anca Stratulat^b, Emmanuelle Diaz^b, Jean-Louis Vercher^a

^a Aix Marseille Univ, CNRS, ISM, Marseille, France

^b Groupe PSA, Centre Technique de Vélizy, Vélizy-Villacoublay, Cedex, France

6.4.1 Abstract

Not just detecting but also predicting impairment of a car driver's operational state is a challenge. This study aims to determine whether the standard sources of information used to detect drowsiness can also be used to predict when a given drowsiness level will be reached. Moreover, we explore whether adding data such as driving time and participant information improves the accuracy of detection and prediction of drowsiness. Twenty-one participants drove a car simulator for 110 minutes under conditions optimized to induce drowsiness. We measured physiological and behavioral indicators such as heart rate and variability, respiration rate, head and eyelid movements (blink duration, frequency and PERCLOS) and recorded driving behavior such as time-to-lane-crossing, speed, steering wheel angle, position on the lane. Different combinations of this information were tested against the real state of the driver, namely the ground truth, as defined from video recordings via the Trained Observer Rating. Two models using artificial neural networks were developed, one to detect the degree of drowsiness every minute, and the other to predict every minute the time required to reach a particular drowsiness level (moderately drowsy). The best performance in both detection and prediction is obtained with behavioral indicators and additional information. The model can detect the drowsiness level with a mean square error of 0.22 and can predict when a given drowsiness level will be reached with a mean square error of 4.18 minutes. This study shows that, on a controlled and very monotonous environment conducive to drowsiness in a driving simulator, the dynamics of driver impairment can be predicted.

Key words: Drowsiness, Prediction, Artificial Neural Network, physiological measurement, behavioral measurement, driving performance and activity

6.4.2 Introduction

Driving a car is a complex, multifaceted and potentially risky activity requiring full mobilization of physiological and cognitive resources to maintain performance over time. Any loss of these resources can have dramatic consequences, including accidents. Moreover, the promise of autonomous vehicles makes it even more important to determine the driver's operational state. This has recently generated a large number of studies, both from the fundamental perspective and with a view to potential applications. The challenge is ambitious: not only detecting, but also predicting, degradation in the driver's operational state.

A driver's operational state while driving a car involves a complex set of psychological, physiological and physical parameters. During driving activities, several factors can be critical: in particular, fatigue and monotony may cause a loss of attention, drowsiness and even sleepiness (Dong, Hu, Uchimura, & Murayama, 2011). The present study focuses on a specific type of impaired operational state: drowsiness. Drowsiness is an intermediate state between alertness and sleep. In this article, we will consider drowsiness as a continuum, or scalar state. Unfortunately, drowsiness cannot be recorded directly but has to be estimated, and several estimation techniques have been proposed in the literature. These methods can be classified in different categories according to source of information: subjective assessment, sensorimotor indicators, physiological features and driving behavior and performance (Dong et al., 2011).

In the last few years, the Karolinska Sleepiness Scale (KSS), a 9-graded Lickert scale (Shahid, Wilkinson, Marcu, & Shapiro, 2011), has become the most commonly employed instrument for the subjective self-assessment of drowsiness (Alhazmi, 2013; Daza et al., 2014; Friedrichs & Yang, 2010; Krajewski, Batliner, & Golz, 2009; Lee, Lee, & Chung, 2016; Li et al., 2014; Murata & Naitoh, 2015). Nonetheless, although often used, this method raises three principal issues. Firstly, the driver's state can only be assessed every 15 minutes, since greater frequency would probably keep the driver awake. Secondly, according to Friedrichs & Yang (2010), when the experiment involves more than three hours of monotonous driving, the KSS becomes inadequate because drivers have difficulty judging their alertness. Lastly, subjective assessment clearly does not constitute an objective measure of drowsiness, and when the task is very monotonous, individual ratings on drowsiness differ from the person's physiological alertness level (Brown, 1997).

Features extracted from eye and head movements, classified as sensorimotor indicators, are also promising parameters to detect the operational state and are now included in many research approaches (Chen & Ji, 2012; Liu, Hosking, & Lenné, 2009). Video-oculo-graphy (VOG) is commonly used to study the following features: blink frequency, blink duration and PERCLOS (PERcentage of eye

CLOSure). Changes in these features are considered under low-level control, offering an easy way to monitor the activity of the neurovegetative system (Caffier, Erdmann, & Ullsperger, 2003; Wierwille & Ellsworth, 1994). These features are generally extracted with image processing algorithms based on eye, head and gaze movement tracking. Thus, the quality of the estimation is highly dependent on this first signal-processing step.

Physiological features are also frequently used to assess drowsiness because they are continuously available and could be considered as an objective, more direct, measure of the functional state. The main recordings of signals related to drowsiness are the electroencephalogram (EEG), the electrocardiogram (EKG) and electro-dermal activity (EDA) (Borghini, Astolfi, Vecchiato, Mattia, & Babiloni, 2014; Dong et al., 2011). The gold standard appears to be the EEG, the most direct indicator of central nervous system activity (De Gennaro, Ferrara, Curcio, & Cristiani, 2001). However, the EEG is quite intrusive, and proper installation of an extensive set of electrodes on the participant's scalp requires expertise and time. It has been established that when a change in vigilance is observed, changes on psychophysiological arousal can be also observed, and these changes can be monitored by measures of the central and autonomic nervous system activity (Haarmann & Boucsein, 2008). Concerning EKG, since heart rate variability (HRV) is linked to the autonomic nervous system this feature is often used as an indicator of drowsiness because change on HRV can provide information about the autonomic nervous system (Elsenbruch, Harnish, & Orr, 1999; Lal & Craig, 2001; Riemersma, Sanders, Wildervanck, & Gaillard, 1977; Stein & Pu, 2012). Moreover, some studies on drowsiness, vigilance or workload also record and analyze respiration rate and amplitude (Besson et al., 2013; Ju et al., 2015; Reimer, Coughlin, & Mehler, 2009; Rodriguez Ibañez, García González, Ramos Castro, & Fernández Chimeno, 2011).

Yet a direct relationship between physiological features and cognitive state is hard to define, because these physiological features vary with other states (including, but not limited to, emotion, workload, physical fatigue) or with the context. These variations according to state also differ from one person to another. Thus, each physiological indicator has its own limits. Heart rate usually decreases during driving and when the driver is tired (Lal & Craig, 2001), but the opposite may also occur (Apparies, Riniolo, & Porges, 1998). Peiris and collaborators (2005) showed that two independent experts analyzing EEGs to detect drowsiness may not make the same assessment for the same participant at the same time. On the other hand, EDA can be influenced by stress (Healey & Picard, 2005) and emotions (Rebolledo-Mendez, Reyes, Paszkowicz, Domingo, & Skrypchuk, 2014). Taken alone, therefore, these indicators in themselves cannot be considered as adequate and exclusive indicators of drowsiness or fatigue.

Driving behavior and performance analyses have the main advantage of being non-intrusive. Some signals such as pressure on pedals or car movements are easily available. The standard deviation of car position relative to lane midline (also named standard deviation of lane position (SDLP)), and steering wheel movements, are the most common features used to detect drowsiness (Arnedt, Wilde, Munt, & MacLean, 2001; De Valck, De Groot, & Cluydts, 2003; Liu et al., 2009; Philip et al., 2004). However, here again, driving performance and activity are not specific indicators of drowsiness. For example, driving performance can decrease with other factors such as distraction (Tango, Calefato, Minin, & Canovi, 2009), or with a decline in attention (Marin-Lamellet et al., 2003)

Since none of these feature families is consensually considered as a specific indicator of drowsiness, various measures are often used jointly. Such a hybrid approach minimizes the number of false alarms while maintaining a high rate of recognition (essential for good acceptance of the system by the human operator, Dong et al., 2011), mainly because no signal emerges as the reference marker allowing real time measurement that is both relatively non-invasive and reliable. Moreover, there is no direct link between all these features and the “operational state”, which is why methods such as machine learning or statistical models are used, combining the different measures.

The different algorithms used include k-nearest neighbors (Chauhan, Saroliya, & Sharma, 2015), decision trees (Lee et al., 2010; Sukanesh & Vijayprasath, 2013), Bayesian classifiers (Lee & Chung, 2012; Yang, Lin, & Bhattacharya, 2010), Support Vector Machines (Bhowmick & Chidanand Kumar, 2009; Krajewski, Batliner, et al., 2009; Liang, Reyes, & Lee, 2007; Yeo, Li, Shen, & Wilder-Smith, 2009), artificial neural networks (ANN) (Bundele & Banerjee, 2009; Eskandarian, Sayed, Delaigue, Mortazavi, & Blum, 2007; Sayed & AEskandarian, 2001; Samiee, Azadi, Kazemi, Nahvi, & Eichberger, 2014), ensemble methods like random forest (Krajewski, Sommer, Trutschel, Edwards, & Golz, 2009; McDonald, Lee, Schwarz, & Brown, 2013; Torkkola et al., 2008; Zhang, Owechko, & Zhang, 2004) and, more recently, deep learning (Hajinoroozi, Mao, & Huang, 2015). Most studies consider the problem of estimating the driver’s impaired operational state as a classification problem. Is the driver in an impaired state or not? Is the driver drowsy or not? However, the evolution of the state of the driver can also be considered as a regression problem, i.e. the driver goes through various continuous states, although regression models are rarely used in the literature (Murata & Naitoh, 2015). Nonlinear modeling machine learning (such as with ANNs) is also often used. With these techniques, the model can extract information from noisy data, and can avoid over-fitting, making it generally more robust (Dong et al., 2011). Since in the context of driving we expected over-fitting and noisy data, the present study uses machine-learning techniques based on artificial neural networks.

Most researches focus on the detection/estimation of an impaired state, rather than on its prediction, even though they adopt the term “prediction” (Chen, 2013; Hargutt & Kruger, 2001; Ji, Zhu, & Lan, 2004; Verwey & Zaidel, 2000). This is because in machine learning, the term “prediction” is used to infer the label of an object not seen during the learning phase. However, some studies try to predict what the ground truth will be in the subsequent few minutes: the ground truth was shifted for one epoch (Kaida et al., 2007), while different lags (+1, +2, +3, +4, +5, +7, + 10 minutes) were tested by Larue, (2010). Murata, Ohta, & Moriwaka (2016) obtained the highest prediction accuracy using the data between 20 and 120 s before the prediction. Watson & Zhou (2016) detect micro-sleep with 96% accuracy and are able to predict, between 15 seconds and 5 minutes in advance, the time when the next micro-sleep will occur. However, the time when the first micro-sleep occurs obviously cannot be predicted by such methods.

As explained above, using a single source of information does not seem to be an efficient way to accurately assess the state of the driver. Different sources of information and different models are used in the literature, and results are hard to generalize away from well-controlled laboratory conditions. In the present study, we collected information originating from different sources: physiological, behavioral, and psychological data from the driver, as well as performance information from the vehicle. The goal of this study is to develop and evaluate a model with an artificial neural network (ANN), so as to predict when a given impaired state will be reached in addition to detecting this impaired state. We deliberately chose unobtrusive recording techniques easily applicable in a car. Different datasets using different sources of information were tested, to determine which kind of information yields the most powerful model. We put forward two hypotheses. First, we hypothesized that it is possible to predict when the impaired state will arise by using the sensorimotor, physiological and performance indicators used to detect drowsiness. Second, we hypothesized that adding information such as driving time and participant information will improve the accuracy of the model.

6.4.3 Materials and methods

6.4.3.1 Participants

A total of 21 participants were included in the study (mean age \pm SD: 24.09 \pm 3.41 years; 11 men and 10 women). On the day of the experiment, the participants were not allowed to drink alcohol, coffee or tea. Inclusion criteria were: valid driver’s licence for at least 6 months, no visual correction needed to drive, not susceptible to simulator sickness (as assessed by the Motion Sickness Susceptibility Questionnaire, Short-form (MSSQ-Short, Golding, 1998) and an Epworth scale score (assessing susceptibility to drowsiness) below 14 (Johns, 1991). A score of below 8 on this scale means the person has no sleep debt. A score of from 9 to 14 means the person shows signs of sleepiness, and if the score

is above 15, the person shows signs of excessive sleepiness. Before the experiment, participants were questioned on their age, their quality of sleep (on a scale of 1 to 10), their caffeine consumption (never, rarely, one or two cups per day, more than two cups per day), driving frequency (occasionally, several times a month, a week or a day), number of kilometers per year. To assess their circadian typology, their score on the Horne and Östberg morning/evening questionnaire (Horne & Östberg, 1975) was also noted. All these indicators concerning the participants were later considered as participant information, and used with a view to improving the performance of the model.

6.4.3.2 Protocol

The participants drove during between 100 and 110 minutes in a static driving simulator in an air-conditioned room with temperature control set at 24° Celsius, after lunchtime. According to the literature about circadian rhythms, the probability of falling asleep between 02:00 to 06:00 and 14:00 to 16:00 is 3 times higher than at 10:00 or at 19:00, respectively (Horne J, Reyner 1999). We chose a period corresponding to an intermediate level between a low risk of drowsiness (in the morning) and the highest risk (end of the night). The road and traffic were generated with SCANer Studio®. While driving, data on driving performance, eyelid and head movements, and physiological data were recorded using the following hardware and software: SCANer Studio® for driving performance at 10 Hz, faceLAB® for sensorimotor signals at 60 Hz, and EKG, pulse plethysmography (PPG), EDA and Respiration with the Biopac® MP150 system and Acqknowledge® software at 1000 Hz. In this study, EDA was also recorded but not used due to extensive signal loss. A webcam was placed on top of the central screen of the simulator to video-record the participants during the session.

At the beginning of the session, the participants drove along a highway for roughly 90 minutes, then turned off the highway and drove for around 5 minutes to reach a city. Finally, they drove in an urban environment for roughly 5 minutes. During most of the highway stretch, there was no traffic. Some 2/3 of the way along, 22 cars appeared from the right of the highway, disappearing a few kilometers later. This sudden addition of traffic was intended to change the driver's level of drowsiness. Rossi and collaborators (2011) demonstrated that a driver is more susceptible to sleepiness in a simulator with a monotonous scenario, and during the afternoon.

6.4.3.3 Data analysis and modeling

The level of drowsiness, the so-called ground truth (indeed, the real state of the driver is not directly accessible and must be evaluated), determined as a reference in this study is based on subjective assessment by video analysis and independently coded by two raters. Their evaluation was based on

a method proposed by Weirwille et al. (1994), which used a scale between 0 and 100. For practical reasons in relation with the ANN, we decided to use a smaller scale (from 0 to 4 with a step of 0.5) as proposed by Belz, Robinson, & Casali (2001). This ground-truth determination method was chosen because the assessment by video coding is reliable and allows a comprehensive assessment of the driver state. Other methods, such as questionnaires (e.g. KSS), reaction time to a double task or even EEG are quite invasive and may disturb the driver and thus influence his/her state. However, video analysis is long and requires several observers with a certain level of training. In order to be more reliable, this method can use criteria and rating scale as a basis for different observers. The ORD relies on a continuous scale from “alert” to “extremely drowsy” with a list of criteria which can be observable in the driver, characteristics of a drowsy driver (Wierwille & Ellsworth, 1994). The two trained raters evaluated each minute of video and rated each segment on a scale ranging from 0 (alert) to 4 (extremely drowsy). The mean of the two raters was taken as the drowsiness level. Inter-rater reliability was computed with the Pearson's linear correlation ($R = 0.71$ and $p=0.00$).

In order to synchronize data obtained at various sampling frequencies, we averaged data over periods of 1 minute. Thus the final sampling rate is 1/min for each feature, including ground truth.

The modeling process can be divided into two phases. First, one Artificial Neural Network (ANN) detects the level of drowsiness from a predetermined set of features (detection model). This ANN is used to detect the impaired state (level of drowsiness). Second, if drowsiness is under 1.5, a second ANN predicts (in minutes) when it will reach 1.5 and gives this time as its output (for instance when the level is reached), otherwise its output is 0 (prediction model). The threshold was set at 1.5 for the following reason. McDonald et al. (2013) defined the limit between “not drowsy” and “drowsy” at a level between 1 and 2 (0 or 1, not drowsy; 2, 3, 4: moderately, very or extremely drowsy). We chose the level of 1.5 as a threshold for defining the impaired state because this level means that at a given time, one of the two raters has evaluated the state of the participant as moderately drowsy (level 2) while the other evaluated the state as 1. These two ANNs were trained independently.

The neural network toolbox (Beale, Hagan, & Demuth, 1992) of Matlab R2013a was used to create the ANNs. Two feedforward neural networks were used with 2 hidden layers, and a back propagation training method was applied using the Levenberg-Marquardt algorithm (Levenberg, 1944). The error was validated by ten-fold cross-validation and a search grid. The performance function used for learning was the mean squared error (the average squared error between the network outputs and the target output). To avoid overfitting, the total dataset was distributed in a training sub-dataset (70% of the total set, to learn the network's node weights), a validation sub-set (15%: to stop learning and

avoid over-training) and a testing sub-set (15%: to evaluate the model's ability to work on previously unseen data. This property is also called 'generalization').

In addition, three other metrics were used to evaluate the model: first, the percentage of numbers of absolute errors below a threshold (0.5 for detection of degree of impairment and 5 minutes for predictions and for the testing dataset: the higher this metric, the better the model performs); second, the range of errors containing 95 % of the values; and third, the coefficient R of the correlation between outputs and targets.

Driving performance and driving behavior indicators (car dataset) used in the model were: lateral distance relative to the midline, time-to-line-crossing (Bergasa, Nuevo, Sotelo, Barea, & Lopez, 2006), steering wheel angle, accelerator pedal angle, shift relative to the lateral line, speed, and number of line crossings. Physiological features used in the model (physiological dataset) were the heart rate and its variability, and the respiration rate and its variability. Sensorimotor features (behavioral dataset) extracted from FaceLab data were blink duration and its frequency, PERCLOS, head movement in translation and rotation, and saccade frequency. Participant information recorded consisted of score on circadian typology, score on Epworth scale, sleep quality, driving frequency, number of cups of coffee a day, age. Driving time (the time elapsed since the beginning of the driving session, in minutes) was also used as an input feature for the model (see Table 1). In an attempt to rebase individual differences, we subtracted from each signal the mean of the first five minutes of this signal, so that the signal represents variation from an initial state. To optimize learning, each feature was normalized such that minimum and maximum values lie within [-1;1].

Tableau 6-0-1 : All the variables (grouped by source of information, in column) computed for each participant for each minute, used as input for ANNs

Physiological measurements	Behavioral measurements	Car measurements
HR: Heart Rate (average and standard deviation) (beat/min)	Blink duration (average and standard deviation)	Lateral distance from the closest lane and the center of the car in m (average and standard deviation)
S_{vlf} : HR signal Very Low Frequency Power (0.0-0.04 Hz)	Blink frequency (average and standard deviation) (per minute)	Time to lane crossing (average and standard deviation)
S_{lf} : HR signal Low Frequency Power (0.04-0.15 Hz)	PERCLOS (average and standard deviation) (% of eye-closure time)	Steering angle (average and standard deviation)
S_{hf} : HR signal High Frequency Power (0.15-0.4 Hz)	Head position x (average and standard deviation)	Steering angle velocity (average and standard deviation)
S_{vhf} : HR signal Very High Frequency Power (0.4-3.0 Hz)	Head position y (mean and standard deviation)	Steering entropy (computed from steering angle)
Sympathetic ratio ($S_{lf} / (S_{vlf} + S_{lf} + S_{hf})$)	Head position z (average and standard deviation)	Number of direction change (0-crossings) per minute (computed from steering angle)
Vagal ratio ($S_{hf} / (S_{vlf} + S_{lf} + S_{hf})$)	Head rotation x (average and standard deviation)	Accelerator pedal angle (average and standard deviation)
Sympathetic-vagal ratio (S_{lf} / S_{hf})	Head rotation y (average and standard deviation)	Lateral shift of the vehicle center relative to the lane center (average and standard deviation)
Respiration Rate (average and standard deviation) (per minute)	Head rotation z (average and standard deviation)	Vehicle speed (km/h) (average and standard deviation)
	Saccade frequency (mean and standard deviation) (per minute)	Number of out-the road per minute

6.4.4 Results

The ANNs were trained 16 times (4x2x2) with different datasets. Each dataset results from the combination of the following: the three sources of information tested alone or all together (thus 4 combinations), with or without elapsed time (2 cases) and with or without information about the participants (2 cases). The tables 2 and 3 present the performances obtained with each of the 16 datasets. In this section, the results will be presented with the driving time (labeled with '1' in tables) and without (labeled with '0' in tables 2 and 3), with the information about the participant (labeled with '1' in tables) and without (labeled with '0' in tables 2 and 3). The grouping was decided according to how these variables were recorded in our experiment (and possibly in a real car), that is to say with which equipment. Indeed, the vehicle information can be recorded from the vehicle's Controller Area network (in our experiment with SCANeR® software), the behavioral measurements with a camera and a specific image processing system (in our experiment with faceLAB®) and physiological measurements with physiological sensors and an A/D system (in our experiment with Biopac®, in a real car it could be with a smart-watch).

6.4.4.1 Detection

In this section, we present model performance in detecting drowsiness level, as defined by the ORD scale (from 0 to 4, see Methods section). The error is the difference between the real state (as given by the subjective evaluation, the so-called ground truth) and the output, squared and averaged over epochs to provide the mean squared error of the trained model.

Tableau 6-0-2: Model performance in detecting drowsiness level for the testing dataset: mean square error (MSE), standard deviation, (STD), according to dataset used, with (1) or without (0) driving time, with (1) or without (0) participant information. The worst performance (highest MSE) is highlighted with a * while the best performance (lowest MSE) is highlighted with a #.

Driving Time	Participant information	Dataset	Source	MSE	STD	Error 95%	% Error <0.5
0	0	Testing	All	0.43	0.04	1.16	0.63
0	0	Testing	Behavioral	0.42	0.02	1.16	0.64
0	0	Testing	Car	0.69	0.04	1.48	0.50
0	0	Testing	physiological	0.81*	0.05	1.51	0.43
0	1	Testing	All	0.41	0.04	1.10	0.62
0	1	Testing	Behavioral	0.39	0.04	1.14	0.69
0	1	Testing	Car	0.62	0.03	1.34	0.54
0	1	Testing	Physiological	0.76	0.03	1.52	0.44
1	0	Testing	All	0.27	0.02	0.91	0.80
1	0	Testing	Behavioral	0.23	0.02	0.80	0.83
1	0	Testing	Car	0.40	0.05	1.20	0.66
1	0	Testing	physiological	0.38	0.05	1.06	0.70
1	1	Testing	All	0.24	0.02	0.84	0.81
1	1	Testing	Behavioral	0.22#	0.02	0.87	0.80
1	1	Testing	Car	0.23	0.06	0.73	0.86
1	1	Testing	physiological	0.29	0.07	0.75	0.82

From an absolute point of view, the dataset configuration providing the best performance (lowest mean square error) in training the model contains driving time, participant information and behavioral features (# in Table 2). With this dataset, the mean square error is 0.22 ± 0.02 and more than 80% of the absolute value of the error of the testing data is under 0.5 (less than one-half of a state level, as

defined by the ORD scale). Ninety-five percent of the absolute value of the error is under 0.87. In other words, the model is off by less than one drowsiness level on our scale, in 95% of cases. Performance is similar when car information is included. The mean square error is 0.23 ± 0.06 . More than 86.34% of the absolute value of the error of the testing data is under 0.5. Ninety-five percent of the absolute value of the error is under 0.73, i.e. in 95 % of cases the model is off by less than one drowsiness level on our scale.

When neither driving time nor participant information is used (line 0-0 in Table 2), or when only one of these is used (0-1 or 1-0), the model performs better with all datasets used together or with the behavioral dataset used alone; performance is slightly worse with the physiological or car datasets used alone. As stated above, the model performs best, for each dataset or for all three datasets used together, when both driving time and participant information are included (1-1).

Figures 1 and 2 present, respectively with (Fig. 1) and without (Fig. 2) driving time and participant information, the frequency histogram of distribution of error (left panel, A) and the correlation (right panel, B) between real state (target, horizontal axis) and estimated state, the output of the ANN (vertical axis). The model is trained with behavioral data in Fig. 1 and with all datasets in Fig. 2; thus, Fig. 1 illustrates the best, and Fig. 2 the worst, performance for the training, validation and testing datasets. Linear regressions were applied to the output of the model to correlate them with the ground truth. With a perfect model, all data points would be on the diagonal line of the correlation graph. Figure 1 shows that, for each of the three datasets, simulated values are well correlated with expected values (ground truth). The R-values are actually very close to unity (0.93, 0.91, 0.91 respectively for the training, validation and testing datasets). Moreover, the slopes of the regression lines are very close to unity (0.87, 0.88, 0.88 respectively for the training, validation and testing datasets) and the intercepts are close to zero (0.17 for all three datasets). Errors are calculated, at each 1-minute epoch, as the difference between the output of the model and the ground truth. The graph on the left of Fig. 1 shows a peak at 0.05, meaning that most of the errors are close to 0. Also, more than 95% of the instances had an error of between -1.16 and 1.16. In Fig. 2, the correlations between output and target are still good but there is greater variability ($R = 0.87, 0.74, 0.78$ respectively for the training, validation and testing datasets). The model used for the results presented in Fig. 2 (Behavior, physiology and car) is less accurate than the model which results are presented on Fig. 1 (behavior, elapsed time and participant information). As for errors, the graph on the left shows a single but broader peak at 0.2 and -0.02, also meaning that most of the errors are close to 0.

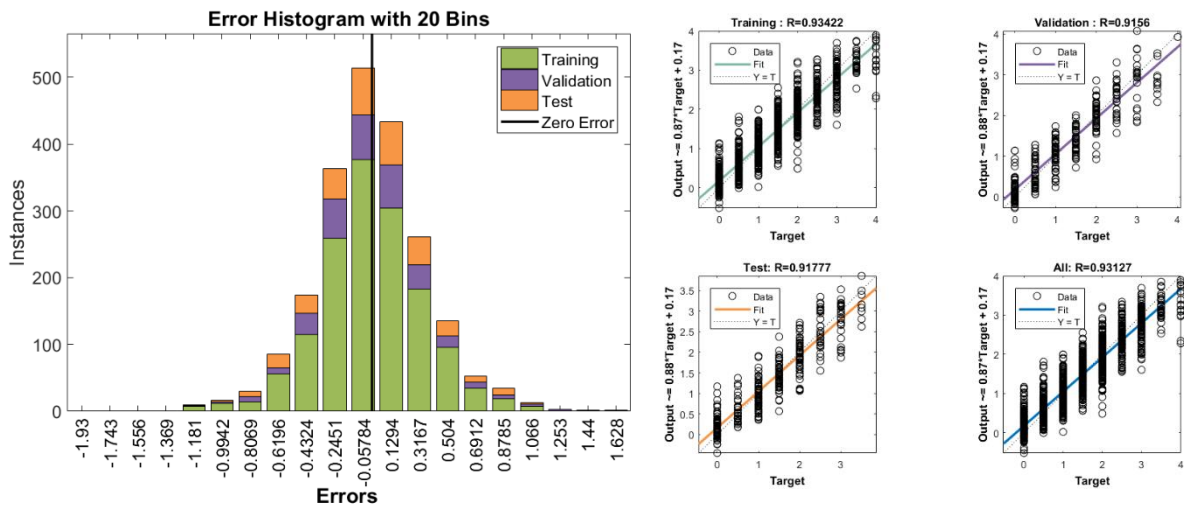


Figure 6-3⁴: frequency histogram of error distribution (left panel) and correlation (right panel) between real and estimated state, for a model trained with behavioral dataset, driving time and participant information.

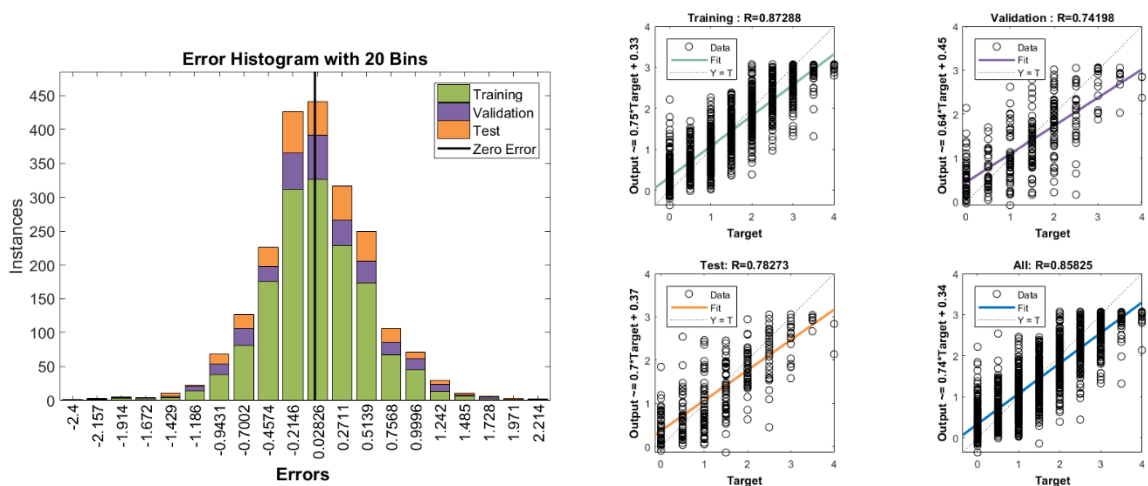


Figure 6-4: frequency histogram of error distribution (left panel) and correlation (right panel) between real and estimated state, for a model trained with behavioral, car and physiological datasets.

6.4.4.2 Prediction

This section presents the performance of the second model, aimed at predicting when a driver will reach a given drowsiness level (here 1.5). The error, for each epoch, is the difference between the time remaining from the current epoch before the target level is really reached (as per the subjective

⁴ Les couleurs des graphiques ont été changés afin de correspondre au code couleur utilisé dans ce document.

evaluation) and the time predicted by the trained model (squared and averaged over epochs to provide the mean squared error).

The best performance is achieved with a combination of driving time, participant information and the behavioral dataset. The mean square error is 4.18 ± 1.17 minutes. For 95% of the testing data, the absolute value of error is under 2 minutes and more than 99% of the absolute value of error is under 5 minutes. Similar, but not higher, accuracy is achieved with the car and physiological datasets (4.67 ± 1.33 and 5.51 ± 1.84). Ninety-five percent of the absolute value of error is under 2.43 and 2.62, respectively. For more than 97% of the testing data, the absolute value of error is under 5 minutes.

The worst model performance in predicting drowsiness is with the car dataset alone (60.09 ± 6.19 minutes). Performance improves with the addition of participant information (50.21 ± 8.84 minutes), or of driving time (31.14 ± 10.72 minutes). The model becomes very accurate when both driving time and participant information are included with the car dataset (4.67 ± 1.33 minutes).

Tableau 6-3: Performance of the model in predicting drowsiness level with the testing dataset: mean square error (MSE), standard deviation (STD), according to whether dataset is used with (1) or without driving time (0), participant information, and source of recorded information. The * symbol indicates the worst performance and the # symbol the best performance.

Driving Time	Participant information	Dataset	Source	MSE	STD	Error 95%	% Error <5
0	0	Testing	All	33.64	7.63	9.29	0.79
0	0	Testing	Behavioral	23.61	3.15	8.12	0.86
0	0	Testing	Car	60.09*	6.19	13.12	0.73
0	0	Testing	Physiological	43.77	6.24	11.47	0.74
0	1	Testing	All	28.26	2.82	8.79	0.82
0	1	Testing	Behavioral	22.83	4.03	7.98	0.89
0	1	Testing	Car	50.22	8.84	12.11	0.73
0	1	Testing	Physiological	41.82	4.11	11.83	0.74
1	0	Testing	All	10.64	3.39	4.26	0.97
1	0	Testing	Behavioral	5.46	1.50	2.92	0.99
1	0	Testing	Car	31.14	10.73	6.25	0.93
1	0	Testing	Physiological	15.97	1.70	7.01	0.89
1	1	Testing	All	7.69	2.17	3.12	0.98
1	1	Testing	Behavioral	4.18#	1.17	1.98	0.99
1	1	Testing	Car	4.67	1.33	2.43	0.99
1	1	Testing	Physiological	5.51	1.84	2.62	0.98

For each source of information (all, behavioral, car and physiological datasets), the model is more accurate when both driving time and participant information are included in the dataset than with either driving time or participant information alone, or with no additional information.

Figure 3 and Figure 4 present the frequency histogram of distribution of errors (left panel) and the correlation between real time (target, horizontal axis) and estimated time (vertical axis) of appearance of drowsiness, respectively with (Fig. 3) and without (Fig. 4) driving time and participant information, The model is trained with behavioral data in Fig. 3 and with all datasets in Fig. 4, so that Fig. 3 illustrates the best, and Fig. 4 the worst, performance. On Fig. 3, the graph on the right shows that the relation between target and output is very precise, data are close to the diagonal (very high R, better than 0.98

for the training, validation and testing datasets, the slopes are better than 0.99). On the left part of Fig. 3, the main peak is at 0.3, meaning that the model has an error inferior at 0.3.

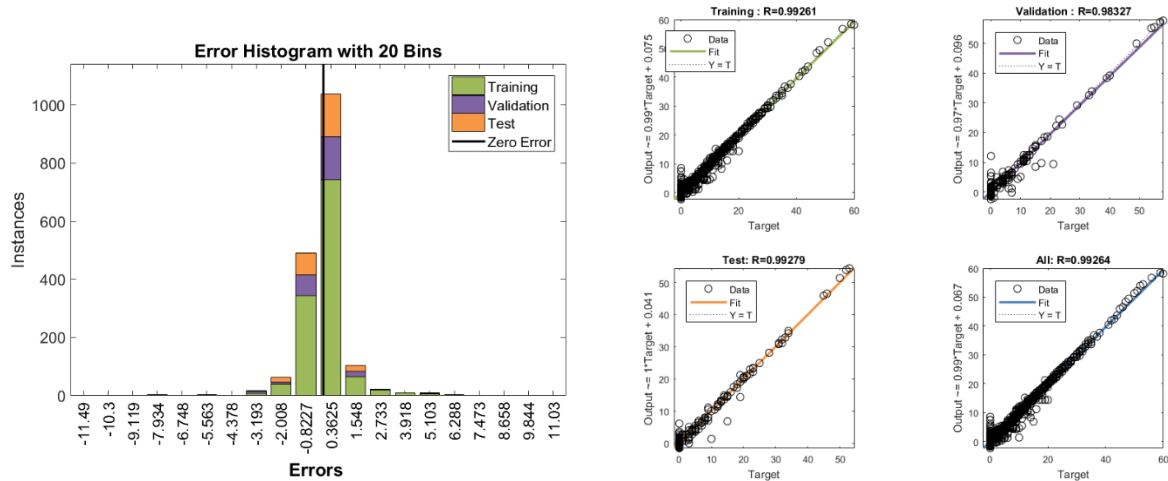


Figure 6-5: frequency histogram of error distribution (left panel) and correlation (right panel) between real and estimated times, for a model trained with behavioral dataset, driving time and participant information.

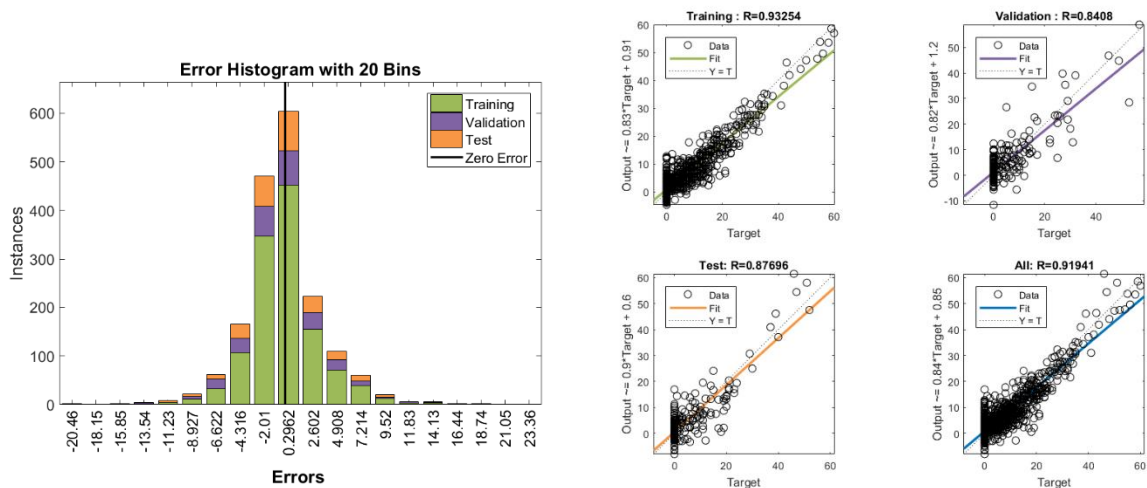


Figure 6-6: frequency histogram of error distribution (left panel) and correlation (right panel) between real and estimated times, for a model trained with behavioral, car, and physiological datasets.

6.4.5 Discussion

Detecting impairment of a driver's operational state is a major safety issue, addressed in numerous studies. While recent car models go some way towards providing this detection capacity, it is clear that recent technological developments are not sufficient to meet the challenge of safety in modern vehicles. Predicting the degree of driver impairment, and when it will occur, remain important research objectives requiring more complex treatment of heterogeneous information from diverse sources. The objective of this study was to assess whether the time of occurrence of a given state of drowsiness could be predicted by using ANN models (one to detect drowsiness and a second one to predict drowsiness).

Overall, our results demonstrate that, using an ANN trained with the same information used to detect drowsiness, it is possible to predict when a driver's impairment will appear to an accuracy of approximately 5 minutes. Moreover, to further improve accuracy, external information such as driving time or a driver profile can be added to the model. In his study, Larue (2010) accurately predicted a driver's decreased vigilance up to five minutes in advance, and up to 10 minutes in advance with 70% to 80% accuracy. Under quite different conditions, and with different types of information, our model seems to be more accurate. In our worst case, for 95% of the test dataset, the model can predict when the impairment will appear to within 13.11 minutes. In our best case, for 95% of the test dataset, the model can predict the impairment to within 1.97 minutes.

As explained in the results section, model performance, both on detecting a drowsiness level and on predicting when this level will be reached, varies considerably according to the datasets used to train the model. This raises the question of the relevance of using physiological signals, behavioral features and driving activity, and of the respective roles of these different datasets in model performance. An important point highlighted by our results is how temporal (driving time) and idiosyncratic (participant information) data impact model performance. The limitations of our model with regard to generalization (i.e. the ability of the model to accurately treat previously unseen data), and from a more general point of view, inter-individual variability, will also be discussed.

6.4.5.1 Dataset comparison: behavioral / physiological / car

Our objective was to use the same information both to detect drowsiness and to predict the time when a given drowsiness level would be reached. Interestingly, when trained with all datasets, either singly or in combination, the model gave satisfactory results. The dataset giving the best performance is the behavioral dataset (followed by the physiological dataset and finally the car dataset), both in detecting the degree of drowsiness and in predicting when a given drowsiness level will occur. Similar

results were previously reported. Samiee and collaborators (2014) showed that information about blinks leads to highly accurate detection (90.74% detection of a drowsy state), while lateral deviation of the car and steering wheel angle provide 85.37% and 87.22% accuracy, respectively. However, when all three sources of information (blinking, lateral position and steering angle) were used together, accuracy increased to 94.69%, although this was not borne out by our study. As in our study, Daza and collaborators (2014) obtained better results with features extracted from eyelid movement (such as PERCLOS) than with features extracted from driving behavior. In the literature, HRV data showed a correlation with drowsiness (Elsenbruch et al., 1999; Lal & Craig, 2001; Stein & Pu, 2012). Yet our model gave better results with ocular and head parameters than with physiological variables: the ORD scale showed a stronger correlation with the ocular parameters than with physiological variables such as EKG and Respiration (Rost et al., 2015). Wang & Xu (2016) consider eye features as the prime input for detection of drowsiness. However, since they are usually computed by image processing, these features cannot be considered fully reliable. Although techniques have progressed considerably in recent years, detecting face and gaze movements remains tricky in complex situations (for example, subjects with glasses, variable or low light conditions, Benoit & Caplier, 2005; Friedrichs & Yang, 2010).

Our behavioral, physiological and, to a lesser extent, car datasets led to the best model performance. With all sources of information in the same neural network, performance could be expected to improve because the neural network can better learn dependencies between different kinds of information. Unfortunately, our results do not bear this out. A single ANN-based model may not be the best way to take advantage of the dependencies between the different sources of information. An alternative, inspired by Samiee and collaborators (2014), might be to linearly combine the outputs of three ANNs, each trained with a different dataset: car, physiological or behavioral.

Surprisingly, other information often included in the literature appears less relevant here. For instance, car deviations relative to the road and line crossings are often considered as signs of drowsiness (Philip et al., 2005). Yet our results unexpectedly show that a model trained with the car dataset alone is less accurate than models trained with other datasets. This may be due to the fact that driving activity and performance are non-linearly correlated with degree of drowsiness. Thus, they may be more useful to detect a critical state (very or extremely drowsy) than to assess a monotonous evolution of the driver's state (alert, slightly, moderately, very, extremely drowsy). Ingre and collaborators (2006) showed that the SDLP score (Standard Deviation of the Lateral Position) dramatically increased with a subjective measure of drowsiness (KSS scale: from 1 to 9). Since our postulate was to consider drowsiness as a continuous variable, the car dataset was obviously not the most appropriate for training our model.

Finally, a potential bias in detection might be suspected from the fact that the ground truth is based on subjective evaluations from video recordings of the participant's motor behavior, which could be thought to explain the superior performance of the behavioral dataset. However, it is worth noting that these features are consensually described in the literature as the most objective and pertinent indicators of drowsiness. It is therefore difficult to conclude on whether the high performance of a model trained with behavioral data is due to the way ground truth is set or to the greater relevance of this particular set of data.

6.4.5.2 The role of driving time

Driving time (the time elapsed since the beginning of the driving session, in minutes) plays an important role here, greatly improving the performance of the model. Obviously, the longer a driver drives under monotonous conditions, the greater the probability of being drowsy (Philip, Taillard, Guilleminault, et al., 1999). This is why drivers travelling on highways are often reminded to take a rest break after two hours of driving (Philip, Taillard, Quera-Salva, Bioulac, & Åkerstedt, 1999). Thus, the model can be considered to have learned a linear relationship between elapsed time and the remaining time before the occurrence of the critical state (naturally until the critical level is reached, after that the predicted time will be 0). It could therefore be deduced that driving time is sufficient per se to predict impairment of the driver's state. However, our experiment showed that participants reached a critical level at different times after the session began. Some participants reached the critical state as early as 10 minutes after the beginning of the driving session, and others after around 30 minutes. Moreover, we observed that some participants could be drowsy at a particular time and subsequently become alert again. It can therefore be concluded that there is not a simple linear relationship between driving time and the time before a given drowsiness level is reached. To determine the real weight of driving time, we consecutively trained two models with this sole feature, and then tested their detection (model 1) and prediction (model 2) capabilities. For the detection of the drowsiness level, the mean square error was 0.47 ± 0.54 . For the prediction of the time before the drowsiness level is reached, the mean square error in the generalization phase was 17.77 ± 2.15 minutes. Interestingly, we find that the models trained with driving time alone perform better than models trained with car or physiological datasets alone, but worse than models trained with behavioral dataset alone or with behavioral, car and physiological datasets combined. This shows that, while driving time is a good predictor of drowsiness, it is not the best.

Secondly, a model based on driving time alone would be unable to account for waking events, such as a rest period or a traffic change. For instance, caffeine is reported to reverse time-on-task

degradation of performance on sleep-deprived participants (Wesensten, Belenky, Thorne, Kautz, & Balkin, 2004). A short nap or rest may counteract drowsiness (Anund, Fors, Kecklund, Leeuwen, & Åkerstedt, 2015). Thus, if the driver drinks a cup of coffee or takes a rest, a model based on driving time alone would need to be reinitialized. How and when this reset should be performed is an important question, requiring further experiments.

6.4.5.3 Generalization and inter-individual variability

Generalization is highly relevant in an industrial context. However, we cannot prove that our model can be generalized to new participants whose data have never been used to train the model. Inter-individual variability (sensitivity to drowsiness, behavioral, physiological or psychological idiosyncrasies) may be a limiting factor for generalization (how the model behaves with previously unseen data) and transfer (how knowledge acquired in a given domain can be adapted to another domain). In our study, the data subset used for the tests (e.g. to evaluate the model's ability to treat previously unseen data, also called the 'generalization process') was randomly chosen among the full set of data from all subjects. Thus, at this stage, it is not possible to determine whether the algorithm would perform well with the full dataset for a given subject whose data were not used to train the model. To do so would require multiple replications of the experiment under the same conditions over a longer period.

It is a major challenge to find a general model which can be trained with a limited number of drivers and then applied to other drivers (Karrer, Vöhringer-Kuhnt, Baumgarten, & Briest, 2004), due to inter-individual variability. Many studies (for a review see Liu et al., 2009) reported great variability in how drowsiness affects performance and physiological parameters in general. It is now recognized that neurobehavioral and cognitive performances vary considerably from one individual to another (Van Dongen, Baynard, Maislin, & Dinges, 2004; Van Dongen, Maislin, & Dinges, 2004). For instance, Philip and collaborators (2004) studied cognitive performance after sleep deprivation. They found that performance was highly impaired, but more so in elderly participants than in younger participants. In car driving, according to Ingre and collaborators (2006), there is extensive inter-individual variability in driving behavior and eye behavior: under similar conditions, individuals can present differing profiles of drowsiness evolution over time, and for a given self-declared drowsiness level, markers such as eye blink duration also vary considerably. In our study, participant information (like age or circadian activity) significantly improved accuracy both in detection and in prediction. These results point in the same direction as those of Wang & Xu (2016), who found that including individual factors improved accuracy. Sensitivity to drowsiness is an idiosyncratic factor which may also impact generalization.

According to Van Dongen and collaborators (2003), the high variability in individual performance following sleep deprivation can be explained by the cognitive performance observed when the individual is not sleep-deprived. Van Dongen and collaborators (2003) also showed that individuals probably differ in their vulnerability to sleep deprivation, and that this is partially predictable from individual cognitive performance without deprivation, i.e. from the individual cognitive profile. Indeed, in driving simulator studies, drowsiness is often observed to develop in differing ways (Thiffault & Bergeron, 2003). Situational and personality factors, sleeping habits and driving history can contribute to the understanding of why some people fall asleep at the wheel while others do not. This points to the need to take into account drivers' traits or profiles when calibrating systems for the detection and prediction of driver fatigue.

6.4.6 Conclusion

In this study, different ANNs were used either to detect a drowsiness level or to predict when a driver's state will become impaired. The best models (those whose rates of successful detection or prediction are the highest) used information about eyelid closure, gaze and head movements and driving time. Performance on prediction is very promising, since the model can predict to within 5 minutes when the driver's state will become impaired. Moreover, modeling drowsiness as a continuum can lead to more precise detection systems offering refined results beyond simply detecting whether the driver is alert or drowsy. Future performance improvements could be achieved by using recurrent neural networks or dynamic neural networks to add temporality to the model, or adding other features like context information (traffic, type of road, weather etc.). These factors can influence the driver's state. However, as eyelid and head movements are difficult to record in a real car, the focus should be on improving a model using only driving performance, driving behavior (based on data provided by sensors in the car) and physiological measurements. Finally, a larger and more realistic dataset (far more subjects (wider range for age for example), recorded in real, on-road, conditions (different times of the day for example) would be required to validate these models.

6.5 Synthèse des résultats

Non seulement détecter, mais aussi prédire la dégradation de l'état opérationnel d'un conducteur de voiture est un défi. Cette étude vise à déterminer si les sources d'information standard utilisées pour détecter la somnolence peuvent également être utilisées pour prédire quand un niveau donné de somnolence sera atteint. De plus, nous explorons si l'ajout de données telles que le temps de conduite et les informations sur les participants améliore la précision de la détection et de la prédiction de la somnolence.

Cet article a démontré qu'avec les mêmes données utilisées dans la littérature pour détecter l'état de somnolence de manière binaire (alerte ou somnolent), il est possible, grâce à des ANN d'une part, de détecter le niveau de somnolence. Il s'agit d'estimer de manière précise le niveau de somnolence du conducteur avec une erreur moyenne quadratique de 0.22 selon une échelle allant de 0 à 4. D'autre part, il est possible de prédire l'arrivée d'un état critique, c'est-à-dire estimer le temps restant avant un état critique avec une erreur moyenne quadratique inférieure à 5 minutes. Un autre résultat significatif de cet article est l'importance de l'ajout du temps de conduite et d'informations personnelles sur le conducteur, aptes à rendre le modèle encore plus performant. Tant pour la prédiction que pour la détection de la somnolence au volant, les meilleures performances sont obtenues grâce au jeu de données comportementales, au temps de conduite et informations personnelles.

Cependant, ces modèles ne nous permettent pas, en raison du choix de division des données, de déterminer les performances de généralisation des connaissances apprises se rapportant à un nouveau conducteur.

Chapitre 7 : Généralisation des modèles de détection et prédiction de la somnolence sur des nouveaux conducteurs

Ce chapitre analyse les performances de généralisation des modèles, entraînés à partir d'un groupe de participants, et testés sur un nouveau participant en se basant sur la même méthodologie que le chapitre précédent. En effet, dans le chapitre 6, nous avons mis en évidence qu'avec un jeu de test tiré aléatoirement dans la base de données, il était possible de détecter et de prédire la somnolence au volant. Cependant, tel que le jeu de test (de généralisation) est construit, nous ne pouvons pas savoir comment le modèle se comporte avec un nouvel individu. Ainsi, pour ce nouveau chapitre, le jeu de test sera un nouveau participant. Avant de présenter les résultats, la division et les sources des jeux de données seront présentées.

7.1 Division des jeux de données

Afin de connaître les performances de généralisation sur un nouvel individu, nous allons réaliser une validation croisée par participant (pour plus de détails, voir partie 2.6.4.4 La validation croisée par sujet). Un ANN sera entraîné puis choisi à cet effet grâce aux données fournies par 20 participants. Les modèles seront entraînés sur la base de 80% des données et seront ensuite validés et choisis avec les 20% de données restantes. L'ANN choisi sera alors testé sur le jeu de données créé par le 21^{ème} participant. D'une part, une validation croisée à 5 plis sera effectuée pour identifier l'ANN le plus performant. D'autre part, cette procédure sera répétée 21 fois, à chaque fois avec un participant différent pour la phase de test (21 étant le nombre de participants présents dans la base de données) : validation croisée inter-individuelle.

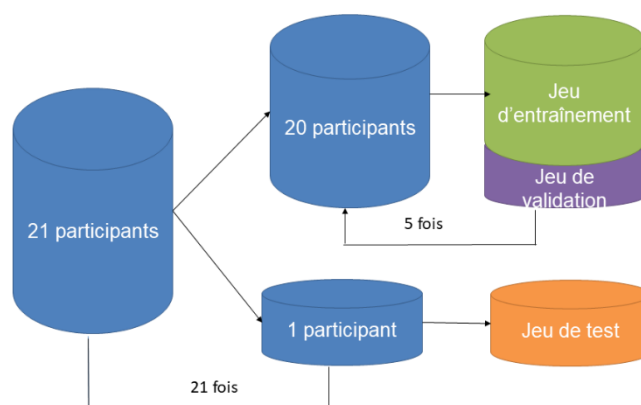


Figure 7-1 : Division des jeux de données (entraînement / Validation / test) pour la généralisation des modèles de détection et prédiction de la somnolence au volant sur un nouveau conducteur

7.2 Sources du jeu de données

Par ailleurs, comme vu dans le chapitre précédent, le temps de conduite (le temps écoulé depuis le début de la conduite) et les informations concernant les participants permettent d'améliorer les performances des modèles. Par la suite, ces variables seront toujours présentes dans la base de données.

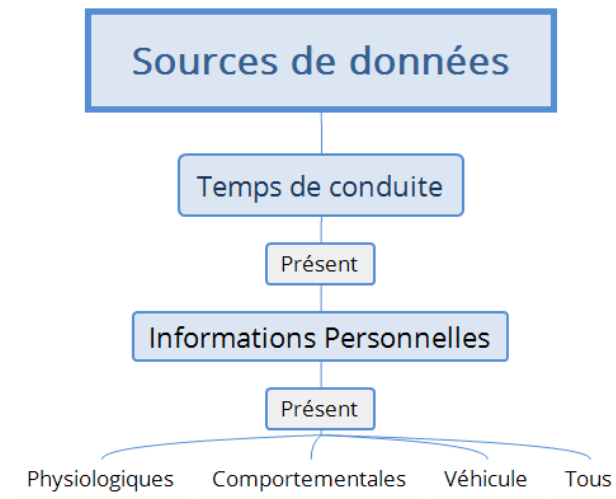


Figure 7-2 : Source de données pour la généralisation des modèles de détection et de prédiction de la somnolence au volant sur un nouveau conducteur

7.3 Problématique de ce chapitre

Le chapitre 7 cible les performances inter-sujets des modèles de détection et de prédiction afin d'estimer s'il est possible de généraliser les connaissances apprises à partir d'un ensemble de participants sur un nouveau conducteur (dont aucune donnée n'a jamais servi à l'entraînement). Toutefois, nous supposons qu'à cause de la grande variabilité inter-individuelle des données et du comportement, nos modèles ne seront pas en mesure de nous permettre de généraliser d'une manière très précise.

7.4 Résultats avec validation croisée inter-

Cette partie des résultats présente les performances des modèles pour le jeu de validation et le jeu de test (nouveau participant) pour les différentes sources d'information : données physiologiques, données liées au véhicule, aspects comportementaux et données avec toutes ces sources confondues. La métrique de performance utilisée est la racine de l'erreur quadratique moyenne (REQM). Comme précédemment, les deux sorties seront exposées : la détection du niveau de somnolence et la prédiction du temps restant avant un niveau critique. Un F-test, à deux échantillons, sera mis en place pour comparer la variance avant un t-test à deux échantillons. Ces tests statistiques seront effectués dans le but de comparer les moyennes et les variances du jeu de validation et du jeu de test (nouveaux participants).

7.4.1 Détection du niveau de somnolence

Avec la source de données « Tous », nous observons une différence significative des moyennes et des écarts-types entre le jeu de validation et de test ($t=-6,686$ $p=0,000$ et $F=0,001$; $p=0,000$ pour la moyenne et l'écart-type respectivement), soit une augmentation de 143,9% en moyenne. Avec la source de données « Physiologique », nous observons une différence significative des moyennes et des écarts-types entre le jeu de validation et de test ($t=-5,51$ $p=0,000$ et $F=0,001$; $p=0,000$ pour la moyenne et l'écart-type respectivement), soit une augmentation de 177,3% en moyenne. Avec la source de données « Véhicule », nous observons une différence également significative des moyennes et des écarts-types entre le jeu de validation et de test ($t=-6,313$ $p=0,000$ et $F=0,000$; $p=0,000$ pour la moyenne et l'écart-type respectivement), soit une augmentation de 282,42% en moyenne. Avec la source de données « Comportementale », nous observons de nouveau une différence significative des moyennes et des écarts-types entre le jeu de validation et de test ($t=-6,213$ $p=0,000$ et $F=0,001$; $p=0,000$ pour la moyenne et l'écart-type respectivement), soit une augmentation de 213,77% en moyenne. Ainsi, pour toutes les sources d'informations confondues, l'erreur de généralisation en situation de nouveau participant est significativement plus importante que l'erreur de validation. Nous observons également une plus importante variance. La figure 7-3 illustre la REQM (racine de l'erreur quadratique moyenne, mesure de performance) en fonction de la source de données concernant à la fois le jeu de validation et le jeu de test dans le cadre de la détection du niveau de somnolence.

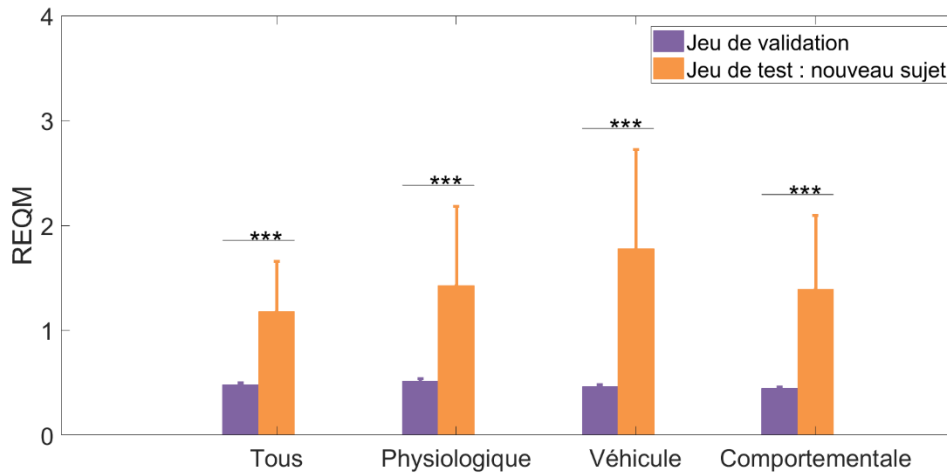


Figure 7-3 : REQM du jeu de validation et du jeu de test pour les différentes sources de données pour la détection du niveau de somnolence. Les étoiles représentent le degré de significativité (NS: $p > 0,05$; *: $p < 0,05$; **: $p < 0,01$; ***: $p < 0,001$).

7.4.2 Prédiction du temps restant avant un niveau dégradé

Avec la source de données « Tous », nous observons une différence significative des moyennes et des écarts-types entre le jeu de validation et de test ($t = -5,778$ $p = 0,000$ et $F = 0,000$; $p = 0,000$ pour la moyenne et l'écart-type respectivement), soit une augmentation de 920,88% en moyenne. Avec la source de données « Physiologique », nous observons une différence significative des moyennes et des écarts-types entre le jeu de validation et de test ($t = -7,548$ $p = 0$ et $F = 0,001$; $p = 0,000$ pour la moyenne et l'écart-type respectivement), soit une augmentation de 604,19% en moyenne. Avec la source de données « Véhicule », nous observons une différence significative des moyennes et des écarts-types entre le jeu de validation et de test ($t = -8,049$ $p = 0,000$ et $F = 0,001$; $p = 0,000$ pour la moyenne et l'écart-type respectivement), soit une augmentation de 756,28% en moyenne. Avec la source de données « Comportementales », nous observons une différence significative des moyennes et des écarts-types entre le jeu de validation et de test ($t = -5,424$ $p = 0,000$ et $F = 0,000$; $p = 0,000$ pour la moyenne et l'écart-type respectivement), soit une augmentation de 979,77% en moyenne. Ainsi, pour toutes les sources d'informations confondues, l'erreur de généralisation avec un nouveau participant est significativement plus importante que l'erreur de validation. Nous observons encore une importante variance. La figure 7-4 illustre la REQM (racine de l'erreur quadratique moyenne, mesure de performance) en fonction de la source de données pour le jeu de validation et le jeu de test dans le cadre de la prédiction du niveau de somnolence.

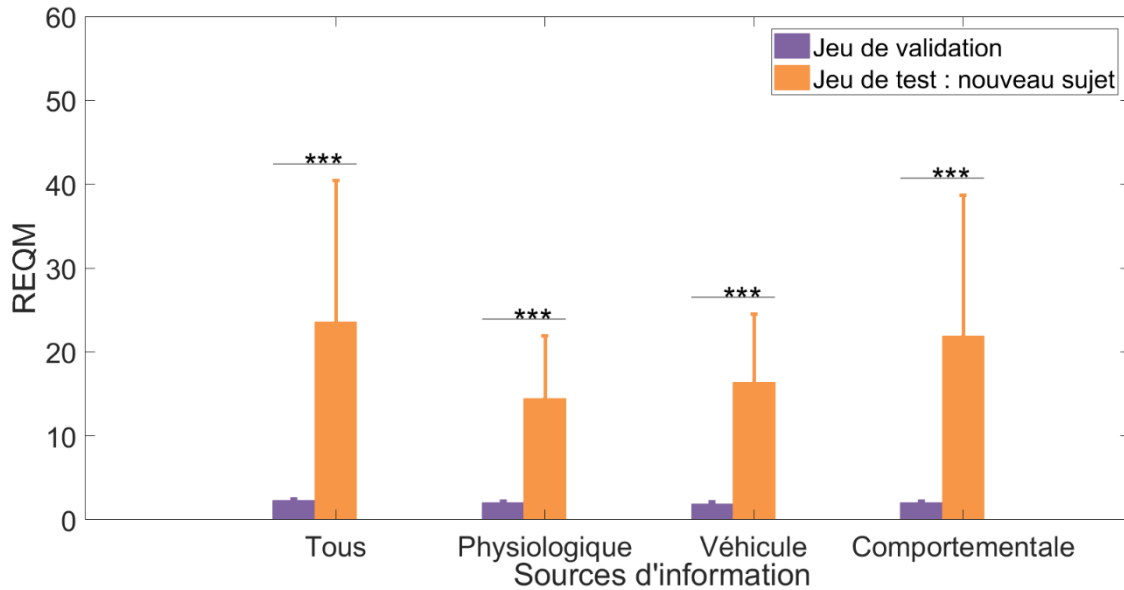


Figure 7-4 : REQM du jeu de validation et du jeu de test pour les différentes sources de données pour la détection du niveau de somnolence. Les étoiles représentent le degré de significativité (NS: $p > 0,05$; *: $p < 0,05$; **: $p < 0,01$; ***: $p < 0,001$).

7.5 Discussion

Le chapitre 7 étudie les performances de généralisation des modèles de ANN sur un nouveau participant. Nous avons donc comparé les performances des modèles sur le jeu de validation (20% de données tirées aléatoirement sur les 20 participants utilisés pour entraîner les modèles) avec les performances des modèles sur les données produites par un nouveau participant grâce à une validation croisée inter- individuelle. Pour chaque source d'informations (tous, véhicule, comportementale, physiologiques), nous observons une importante et significative augmentation de la moyenne et de l'écart-type de l'erreur du jeu de test par rapport au jeu de validation. Nous pouvons par conséquent dire que le modèle, qui a une bonne performance lorsqu'il est testé avec des données inédites provenant de participants utilisés pour l'entraînement, voit cette performance se dégrader lorsqu'on le teste avec des données provenant d'un nouveau participant. Ceci caractérise une faible généralisation, conformément à nos attentes. Concernant l'estimation du niveau de somnolence (détection), l'ordre d'erreur est en effet situé entre 1 ou 2 niveaux de somnolence. De plus, concernant le modèle de prédiction, l'erreur est située entre 14 et 24 minutes. Au niveau de l'erreur de généralisation (jeu de test), une grande variance est également observée.

Dans la littérature, à notre connaissance, une comparaison de résultats entre une validation croisée aléatoire et une validation croisée inter- individuelle n'a pas ou peu été réalisée. Toutefois, il a été

comparé une validation croisée intra-sujet (le jeu d'entraînement et de test viennent du même participant) et inter-participants, principalement avec comme mesure de l'EEG (Lin et al., 2012). En comparant la validation croisée inter-individuelle et intra-participant, les performances sont supérieures avec la validation croisée intra-participant. Une étude a également obtenu des performances intéressantes en validation croisée inter-individuelle avec, comme données d'entrées, des informations provenant de l'ECG (Vicente et al., 2016). Nous pouvons donc nous interroger sur les causes de la moins bonne performance observée à partir de nos propres résultats.

Ces mauvaises performances de généralisation pourraient s'expliquer de différentes manières. Tout d'abord, même si nous avons tenté de limiter les facteurs de variabilité (contrôlé la population, normalisé les variables, contrôlé l'heure de passation, etc.), nous observons tout de même une très grande variabilité du comportement et des données physiologiques, conformément à la littérature (Liu et al., 2009; Van Dongen, Baynard, et al., 2004; Van Dongen, Maislin, & Dinges, 2004). Cet état de fait pourrait expliquer les mauvaises performances car s'il n'existe pas de conducteurs « *similaires* » dans la base de données d'entraînement, le modèle ne permet pas une généralisation correcte sur tout nouveau conducteur en raison de la grande variabilité inter-individuelle des comportements. Deuxièmement, il se peut que le modèle créé soit trop complexe par rapport aux données présentes. En effet, ce modèle semble sensible à des variations au niveau des données. Plus formellement, cette mauvaise performance s'explique par le fait que chaque participant n'a pas la même distribution de données. Nos données ne sont en effet pas indépendantes et identiquement distribuées. Ainsi, les espaces de dimensions des données pour chaque participant sont très différents et, d'autre part, les données à l'instant t-1 ne sont pas indépendantes de l'instant t.

Il existe plusieurs solutions possibles afin de contrer ces différentes difficultés. Tout d'abord, il pourrait être nécessaire de recueillir des données supplémentaires avec davantage de conducteurs. Néanmoins, ce choix réclame du temps, tant pour récupérer davantage de données que pour les nettoyer et les labelliser. Des modèles plus simples et moins complexes que les réseaux de neurones pourraient également être testés, mais il faudrait, dans ce cas, se demander s'ils pourraient représenter le phénomène complexe de la somnolence au volant. De plus, une autre solution pourrait être envisagée comme changer le jeu de validation. Dans le cadre de cette étude, le jeu de validation est choisi aléatoirement parmi les données des 20 participants. Si le jeu de validation représentait un participant différent de celui du jeu de test, alors l'entraînement serait certainement arrêté plus tôt ce qui limiterait le sur-apprentissage. Toutefois, dans ce cas, la proximité du participant de validation et du jeu de test est très certainement encore nécessaire. Cette notion de proximité entre deux participants est difficilement définissable et nécessitera d'être également apprise par les modèles. Il y aurait alors deux apprentissages successifs : l'entraînement d'un ANN et la recherche d'une fonction

de similarité. Il y aurait ainsi, d'une part, l'apprentissage d'un ANN avec comme jeu de validation un participant A, puis un autre participant B permettrait de définir et de rechercher une fonction de distance en cherchant quel participant B est le plus proche du participant A. Finalement, un troisième participant C serait nécessaire pour tester la généralisation de la notion de distance apprise entre le participant A et B. Ces deux apprentissages successifs (ANN et distance entre participant) amènent donc à faire varier dans notre cas pour 21 participants :

- 21 fois le participant C ;
- pour chaque C, 20 fois le participant B ;
- pour chaque B, 19 fois le participant A ;

La modélisation doit donc être effectuée au total $21 \times 20 \times 19 = 7980$ fois. Ces différentes combinaisons complexifient grandement la mise en place d'un tel système. Une autre solution qui a également été étudiée dans la littérature est d'avoir un modèle spécifique pour chaque individu (Lin et al., 2012; Liu, Lin, Wu, Chuang, & Lin, 2016; Liu, Lin, Wu, Hsieh, & Lin, 2015; Rohit et al., 2017). Cependant, cette méthode réclame également du temps pour labelliser les données et obtenir un nombre de données suffisant pour être en mesure de construire un modèle adapté à chaque participant. Finalement, une cinquième possibilité est d'utiliser des méthodes de transfert d'apprentissage, ou d'apprentissage adaptatif, qui tirent parties des avantages des différentes méthodes précédentes. En effet, ces méthodes peuvent permettre de capitaliser la connaissance apprise sur un groupe d'individu puis ensuite de personnaliser, d'adapter et/ou de transférer cette connaissance sur un nouvel individu. Comme il existe une très grande variabilité inter-individuelle, posséder des modèles et algorithmes, qui performant correctement pour tous les individus, est encore un challenge important.

7.6 Conclusion

Ce chapitre a mis en exergue le fait que notre méthodologie et nos modèles ne permettent pas de généraliser les connaissances apprises à partir d'un ensemble d'individus sur un nouvel individu en raison de la grande variabilité inter-individuelle. Celle-ci est caractérisée par le fait que les données ne soient donc pas indépendantes et identiquement distribuées. Il est donc nécessaire de trouver les moyens de rendre plus performant le transfert d'un groupe de conducteurs à un nouveau conducteur.

Chapitre 8 : Adaptation personnalisée des modèles de détection et prédiction de la somnolence

Les résultats des deux chapitres précédents nous ont montré que les ANN sont très performants pour estimer un niveau de vigilance ou un temps avant l'occurrence d'un niveau critique lorsque les données du jeu de test sont tirées aléatoirement dans la base de données, composée de tous les participants. Toutefois, lorsque les données du jeu de test proviennent d'un nouveau participant, cette fois-ci les modèles ne sont pas performants. Ce problème est dû certainement à une grande variabilité des performances et des mesures entre les participants. Dans ce cas, il est envisageable d'utiliser les connaissances apprises grâce à un groupe de conducteurs, formant le jeu d'entraînement, et de transférer et d'adapter ensuite ces connaissances sur un autre conducteur, formant le jeu de test. Ce transfert pourrait être réalisable grâce à une faible quantité de données provenant du nouveau conducteur. Ce chapitre va exploiter cette idée.

Ce chapitre présentera tout d'abord ce qu'est le transfert et l'adaptation de l'apprentissage en le comparant à l'apprentissage classique. Puis dans une seconde partie, nous présenterons de nouveau la division et la source du jeu de données qui ont servies à l'adaptation personnalisée des modèles de détection et de prédiction de la somnolence vers un nouveau participant. Enfin, l'article qui a été l'objet d'une publication sera présenté ensuite. L'enjeu est double :

- observer et analyser les performances d'un ANN général appris sur 20 (n-1) conducteurs, adapté grâce aux premières données d'enregistrement (les premières minutes de conduite, supposées caractériser l'état initial de ce conducteur) d'un nouveau conducteur ;
- étudier l'impact de la quantité de données nécessaires sur l'amélioration des performances de généralisation sur le nouveau conducteur.

La méthodologie pour recueillir les données est décrite dans la partie 5. Quant à la méthodologie de modélisation, celle-ci est présentée et détaillée dans l'article suivant.

8.1 Apprentissage adaptatif

Intéressons-nous tout d'abord à la manière d'entraîner un ANN. Il existe en effet deux méthodes différentes d'entraînement : une méthode incrémentale et une méthode dite par lot (*batch* en anglais). Avec un entraînement incrémental, les poids et les biais des neurones sont mis à jour à chaque fois qu'une nouvelle donnée est présentée au ANN. Egalement appelée apprentissage adaptatif, cette méthode consiste à effectuer l'entraînement au fur et à mesure que les données deviennent disponibles. L'entraînement par lot consiste quant à lui à une mise à jour du réseau en présentant

simultanément toutes les données ou seulement un sous-ensemble de ces données. Cette seconde méthode est généralement considérée comme plus efficace (Beale et al., 1992).

Dans ces deux styles d'entraînement, les données peuvent être présentées une fois ou plusieurs fois au ANN afin de mettre à jour les poids et les biais ; ce nombre de fois est appelé « époque ». L'utilisation des deux méthodes combinées est également possible : présentation séquentielle de plusieurs blocs de données ou entraînement initial non adaptif, suivi d'un processus d'adaptation. Ces deux processus combinés sont utilisés dans le domaine du transfert d'apprentissage : un réseau profond pré-entraîné est ensuite adapté pour apprendre une autre tâche (LeCun, Bengio, & Hinton, 2015). Dans ce cas, la détermination des poids du réseau est plus rapide, plus simple et nécessite moins de données qu'un nouvel entraînement qui recommencerait à partir de zéro.

En apprentissage automatique traditionnel, un seul type de jeu de données est utilisé pour entraîner un modèle à une certaine tâche. Pour deux domaines différents, deux tâches seront apprises. Dans le cas de transfert d'apprentissage, un apprentissage de la tâche A sera réalisé grâce à un domaine « source » et ces connaissances seront transférées ensuite pour apprendre la tâche B sur le domaine « cible » (illustré dans la figure 8-1).

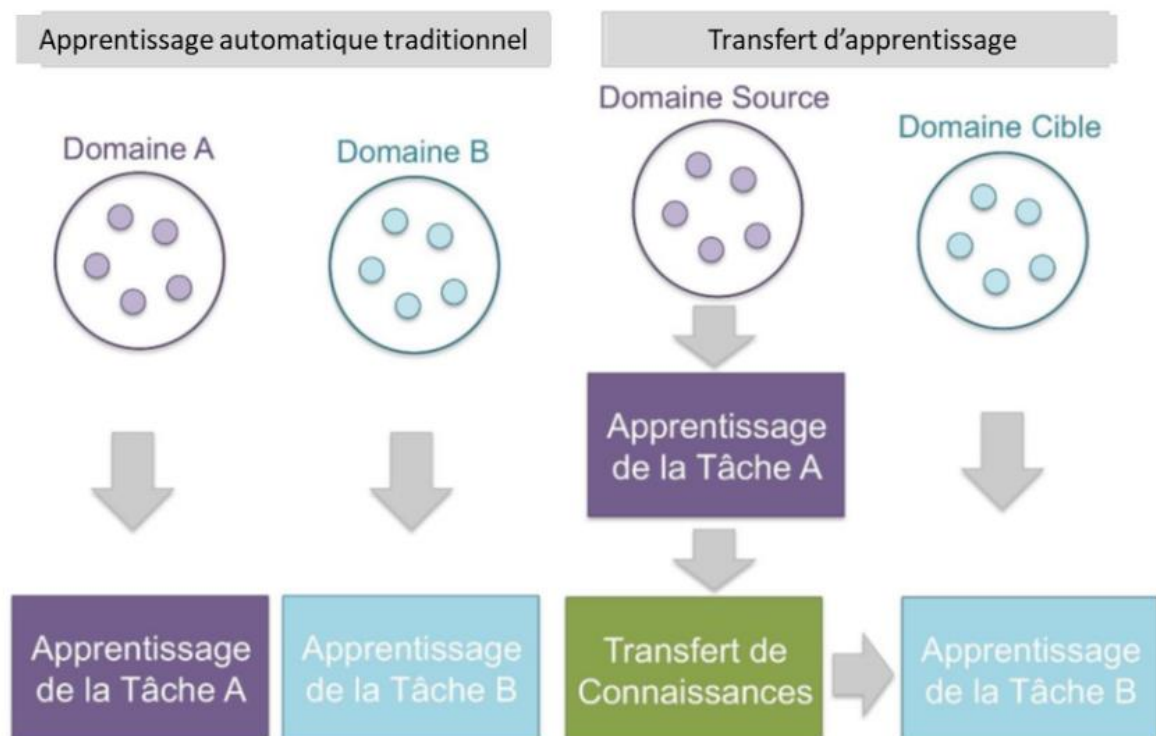


Figure 8-1 : Schéma explicatif entre Machine Learning Traditionnel et Transfer Learning (adapté de Pan & Yang, (2010))

Ce procédé peut être utilisé en l'occurrence lorsque peu d'exemples sont présents dans le domaine B/cible. Par rapport à ces définitions, le domaine fait référence à l'espace caractéristique de la tâche, c'est-à-dire aux données d'entrées ; la tâche est quant à elle caractérisée par la fonction prédictive. Dans notre cas, le domaine correspond aux données enregistrées sur un ou certains conducteurs. Quant à la tâche, elle correspond au fait d'estimer un certain niveau de somnolence ou de prédire le temps avant l'occurrence d'un état dégradé à l'aide d'ANN. La finalité de l'apprentissage par transfert est par conséquent d'améliorer les performances sur la deuxième tâche en se servant des connaissances apprises à partir du domaine source. Afin de permettre un transfert positif de connaissances, il est nécessaire que les deux tâches possèdent des points communs et des liens.

Dans notre cas, nous n'utilisons pas de réseaux de neurones artificiels profonds mais de simples réseaux de neurones artificiels à une ou deux couches. Néanmoins, la possibilité d'utiliser ce processus d'apprentissage en deux étapes est possible : un apprentissage global et un deuxième apprentissage grâce à l'utilisation de peu de données sur un nouveau participant. L'utilisation du transfert d'apprentissage pour la modélisation de la somnolence pourrait permettre de tirer profit de l'apprentissage basé sur d'autres conducteurs et, ainsi, réduire le temps de calibrage du système pour détecter et prédire la somnolence au volant sur un nouveau conducteur. L'utilisation de réseaux pré-entraînés est beaucoup employée dans la classification d'images car l'apprentissage profond a eu sa renommée grâce à ses performances très intéressantes sur la reconnaissance d'objets visuels, la détection d'objets ou la reconnaissance vocale (LeCun et al., 2015). Concernant l'homme et les mesures physiologiques et/ou comportementales, cette méthodologie a été, à notre connaissance, principalement utilisée dans le domaine des interfaces-cerveaux-machines (*Brain computer interface* en anglais, BCI) qui comprend généralement deux parties : une séance d'entraînement et une séance de test. L'apprentissage par transfert est choisi pour réduire l'étalonnage et le calibrage grâce à une utilisation de données enregistrées sur d'autres utilisateurs de l'interface homme-machine (Tu & Sun, 2012; Wang, Lu, Zhang, & Tang, 2015). Avec comme données des signaux provenant de l'EEG, des méthodes de *transfer learning* commencent à être utilisées pour détecter la somnolence au moyen d'algorithmes d'adaptation de domaine (Wei et al., 2015; Wu, Lawhern, Gordon, Lance, & Lin, 2016; Dongrui Wu, Chuang, & Lin, 2015). Une étude a, quant à elle, étudié la quantité de données nécessaires pour effectuer l'adaptation en sélectionnant au hasard, à chaque fois, 5 échantillons de plus (Dongrui Wu et al., 2015). L'étude de Dongrui Wu, Lawhern, Gordon, Lance, & Lin, (2016) a testé une méthode d'apprentissage hors-ligne dans l'objectif de sélectionner de manière optimale des données à labelliser afin que le modèle de régression puisse être construit pour étiqueter le reste des données. Ils proposent une méthode d'apprentissage actif en mode « batch » pour la régression comportant des données d'entrées d'EEG. Leurs méthodes améliorent la fiabilité, la représentativité et la diversité des

échantillons sélectionnés pour obtenir un meilleur calibrage. Une étude a également montré qu'en sélectionnant des sessions spécifiques pour le transfert d'apprentissage, il est possible d'atteindre une meilleure généralisation que celles des performances avec un apprentissage intra-sujet (c'est-à-dire apprentissage et test avec le même sujet). Toutefois la quantité de données, servant au transfert d'apprentissage, n'influence pas les performances (Wei et al., 2015). Cependant, ces différentes sélections de données ont été effectuées de manière non-chronologique, en temps réel, c'est-à-dire que quelques cas de somnolence pourraient ne pas avoir été détectés par les modèles. Ainsi, dans le cadre de cette étude, nous nous sommes intéressés à l'utilisation des premières données de conduite permettant un transfert d'apprentissage.

8.2 Division des jeux de données

Pour répondre à cette situation, nous allons réaliser une validation croisée inter-individuelle particulière. Tout d'abord, à chaque fois, un sujet sera extrait de la base de données. Le jeu de données, composé de 20 sujets sur 21, sera divisé en deux : jeu d'entraînement avec 80% des données et jeu de validation avec les 20% restants. Une validation croisée à 5 plis sera également réalisée afin de changer le jeu de validation. Ces deux jeux serviront à construire un ANN général. Le sujet mis de côté sera utilisé comme base de données d'adaptation de l'ANN et comme jeu de test. Les premières données permettront d'adapter l'ANN, les données restantes seront utiles, quant à elles, pour évaluer les performances de généralisation du modèle. La division des jeux de données est représentée schématiquement dans la figure 8-2. La méthodologie sera décrite plus en détails par la suite dans l'article suivant.

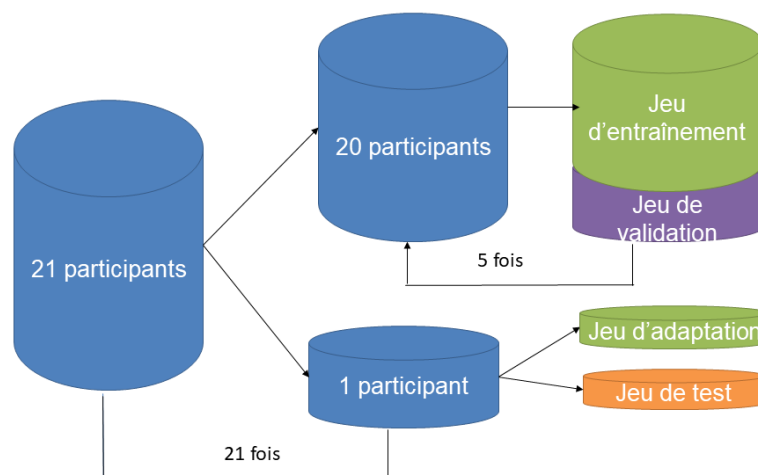


Figure 8-2 : Division des jeux de données

8.3 Sources du jeu de données

Comme vu dans le chapitre 6, le temps de conduite (le temps écoulé depuis le début de la conduite) et les informations concernant les participants permettent d'améliorer les performances des modèles. Ces variables seront donc présentes dans la base de données.

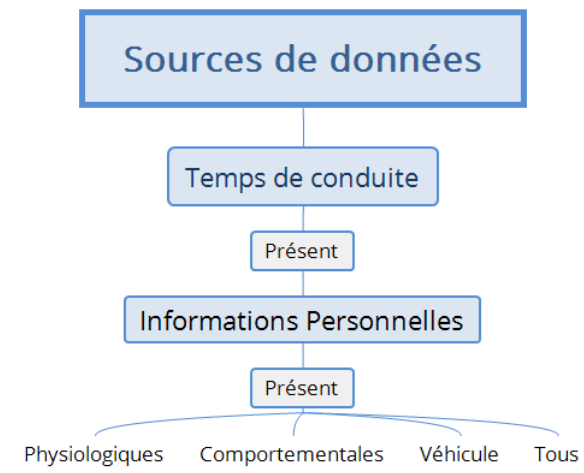


Figure 8-3 : Source de données pour l'adaptation des modèles de détection et de prédiction de la somnolence au volant sur un nouveau conducteur

8.4 Problématique du second article

L'objectif de ce second article est d'améliorer les performances des modèles d'apprentissage automatique, tant celui de détection du niveau de somnolence que celui de prédiction de la somnolence au volant. Ce dernier estime le temps restant avant un niveau critique. Nos modèles utilisent des réseaux de neurones artificiels adaptatifs (ANN). Nous émettons l'hypothèse que l'entraînement d'un ANN avec un groupe d'individus et ensuite l'adaptation de ce même modèle, grâce aux premières données d'enregistrement d'un nouvel individu (dont les données n'ont pas été vues par le modèle lors de la phase d'entraînement), améliore les performances de ce modèle adapté pour ce conducteur spécifique. Nous étudierons parallèlement la quantité de données requises pour l'adaptation des modèles ainsi que la source d'information provoquant les meilleures performances de généralisation.

8.5 Adapting artificial neural networks to a specific driver enhances detection and prediction of drowsiness

Adapting artificial neural networks to a specific driver enhances detection and prediction of drowsiness

Charlotte Jacobé de Naurois^{a, b}, Christophe Bourdin^a, Clément Bougard^b, Jean-Louis Vercher^a

^a Aix Marseille Univ, CNRS, ISM, Marseille, France

^b Groupe PSA, Centre Technique de Vélizy, Vélizy-Villacoublay, Cedex, France

8.5.1 Abstract

Monitoring car drivers for drowsiness is crucial but challenging. The high inter-individual variability observed in measurements raises questions about the accuracy of the drowsiness detection process. In this study, we sought to enhance the performance of machine learning models (Artificial Neural Networks: ANNs) by training a model with a group of drivers and then adapting it to a new individual. Twenty-one participants drove a car simulator for 110 minutes in a monotonous environment. We measured physiological and behavioral indicators and recorded driving behavior. These measurements, in addition to driving time and personal information, served as the ANN inputs. Two ANN-based models were used, one to detect the level of drowsiness every minute, and the other to predict, every minute, how long it would take the driver to reach a specific drowsiness level (moderately drowsy). The ANNs were trained with 20 participants and subsequently adapted using the earliest part of the data recorded from a 21st participant. Then the adapted ANNs were tested with the remaining data from this 21st participant. The same procedure was run for all 21 participants. Varying amounts of data were used to adapt the ANNs, from 1 to 30 minutes, Model performance was enhanced for each participant. The overall drowsiness monitoring performance of the models was enhanced by roughly 40% for prediction and 80% for detection.

Key words: monitoring, ANN, adaptive learning, inter-individual variability, drowsiness

8.5.2 Introduction

Driving while drowsy is a safety issue, and a major cause of accidents. Numerous fundamental and applicative studies focus on detection of drowsiness as a way to improve accident prevention. However, simply detecting drowsiness is not enough: once the driver is drowsy, it is probably already too late to prevent the accident. The key challenge is to predict how and when drowsiness will occur, how often it will occur and who might become drowsy under which conditions. Prediction refers here to the timely identification of when a given event will occur within a given range of future states, in our case a given level of drowsiness. Watson & Zhou (2016) detected the occurrence of micro-sleep episodes with 96% accuracy and were able to predict the next micro-sleep between 15 seconds and 5 minutes in advance, although obviously not the time of occurrence of the first micro-sleep. A recent study (Jacobé de Naurois, Bourdin, Stratulat, Diaz, & Vercher, 2017) showed that an Artificial Neural Network (ANN) can not only detect the level of drowsiness but can also predict, in advance, the time at which this impaired driver's state will occur.

Various sources and types of information can be used to estimate the operator's functional state. For car driving, measurements must be easily recordable, not invasive, and reliable. The literature contains a variety of sources of information (Dong et al., 2011), mainly based on ocular and eyelid movements (J. Chen & Ji, 2012; C. C. Liu et al., 2009). For instance, PERCLOS (PERcentage of eye CLOSure, the percentage of time, generally during one minute, when eyes are closed more than 80%) indicates how long on average the eyes are closed. Physiological measurements are also often used to assess the driver's state through the central and the neuro-vegetative systems, offering the advantage of being continuously available, objective and fairly direct indicators of the functional state. The most commonly used physiological signal is the electroencephalogram (EEG). However, EEG recording during driving is rather intrusive and constraining (despite continuous technological advances), which can be a real disadvantage. Electrocardiogram (EKG) and respiration measurements are also often used. Yet it remains difficult to define a direct relationship between physiological features and a given cognitive state, since these physiological features vary with other states like stress, emotions, workload, physical effort and fatigue, or with the context.

Finally, driving behavior and performance, such as the standard deviation of car position relative to lane midline (also termed standard deviation of lane position (SDLP)) or steering wheel movements, (Arnedt et al., 2001; De Valck et al., 2003; C. C. Liu et al., 2009; Philip et al., 2004) are also common measures used to detect the driver's state. However, here again, driving performance and activity are not specific indicators of drowsiness.

To deal with the above limitations, recent research has sought to improve prediction through complex approaches combining multivariate, heterogeneous information via data fusion (Dong et al., 2011; Samiee et al., 2014). Findings from these studies show that this hybrid approach can provide better accuracy (Awais, Badruddin, & Drieberg, 2014).

However, current models need to deal with yet another challenge to their prediction power. It is now widely recognized that neurobehavioral and cognitive performance vary considerably from one individual to another (Van Dongen, Baynard, et al., 2004; Van Dongen, Maislin, et al., 2004, 2004). In car driving tasks, according to (Ingre et al., 2006), there is extensive inter-individual variability in driving behavior and eye behavior. Under similar conditions, individuals' patterns of drowsiness evolution over time can differ, and for a given self-declared drowsiness level, markers such as eye blink duration also vary considerably. Van Dongen, Rogers, & Dinges (2003) showed that individuals probably also differ in their vulnerability to sleep deprivation, and that this is partially predictable from individual cognitive performance without deprivation, i.e. from the individual cognitive profile. In driving simulator studies, drowsiness is often observed to develop in differing ways (Pierre Thiffault & Bergeron, 2003b). Situational and personality factors, sleeping habits and driving history help explain why some people fall asleep at the wheel while others do not. This confirms the need to consider drivers' traits or profiles to calibrate systems for the detection and prediction of drowsiness (Jacobé de Naurois et al., 2017).

Such large inter-individual variability makes creating algorithms that will perform well for all individuals a challenge. As most studies use machine-learning algorithms, the difficulty is finding a general model trained with a limited number of drivers which can then be applied to the majority of individual drivers (Karrer et al., 2004). One of the main issues with machine learning is uncertainty about the generalization of a given model to a new participant. To ascertain whether an algorithm generalizes well, the dataset is segregated into either two (training and testing) or three (training, validating and testing) datasets (in most cases, the segregation is randomly performed on the full set of recorded data). Thus, it is impossible to be sure that the algorithm will perform well for another participant whose data is unknown to the model.

This problem can be approached in different ways. One is to train the model with as large a population as possible: the more data, the better the model. However, this method is based on the assumption that for each new individual, the model has previously encountered a similar individual. This makes it difficult to determine the number of participants required to deal with the large inter-individual variability. Furthermore, the level of similarity between two individuals is hard to quantify. This method would thus be extremely time-consuming, not only in terms of training the model but

also in terms of data collection. A second solution is to have a specific model for each driver, but this obviously involves collecting and labeling sufficient data from each driver as well as training the specific model with these data, another time-consuming option. A third way is to use methods such as transfer learning or adaptive learning, which combine the advantages of the two preceding methods by permitting capitalization on a group of individuals and personalization for each new individual. In particular, these methods are applied on Brain Computer Interface systems (P. Wang et al., 2015). To detect driver drowsiness, studies applied such techniques on EEG signals (Wei et al., 2015; Dongrui Wu et al., 2015, 2016) and found that transfer learning applied to EEG significantly enhances model performance. Our aim here was to test a similar method based on adaptive learning but using non-intrusive measurements including eyelid movements, head movements, EKG, respiration rate, driving activity and performance, as in our previous study (Jacobé de Naurois et al., 2017).

The goal of the present study is to enhance the performance of machine-learning models both in detecting the level of driver drowsiness and in predicting when a given impaired state will be reached, by first training a model and then adapting it to each new individual. The model uses Artificial Neural Networks. We hypothesize that training an ANN with a group of individuals and then personalizing the ANN for a new individual (whose data were not encountered by the model during training) will improve the performance of the model for this specific individual. We also assess the amount of data required to enhance the generalization performance of the model.

8.5.3 Materials and methods

The participants and the protocol, including data collection and preprocessing, were the same as used for our previous study (Jacobé de Naurois et al., 2017). Data modeling methods were specifically developed for the present study.

8.5.3.1 Participants

Twenty-one participants were included in the study (mean age 24.09 ± 3.41 years; 11 men, 10 women). Inclusion criteria were: valid driver's license for at least 6 months, no visual correction needed to drive, not susceptible to simulator sickness, as assessed by the Motion Sickness Susceptibility Questionnaire, Short-form (MSSQ-Short, Golding, 1998), and an Epworth scale score (assessing susceptibility to drowsiness) below 14 (Johns, 1991) (for more detail, see (Jacobé de Naurois et al., 2017)). The following participant information was collected: Epworth scale score (assessing susceptibility to drowsiness (Murray W. Johns, 1991)), quality of the previous night's sleep (on a scale from 1 to 10), caffeine consumption (never, rarely, one or two cups per day, more than two cups per day), driving frequency (occasionally, several times a month/a week/a day), distance (kilometers)

driven per year and score on the Horne and Östberg morning/evening questionnaire (Horne & Östberg, 1975).

8.5.3.2 Protocol

The participants drove for 100 to 110 minutes in a static driving simulator in an air-conditioned room with temperature control set at 24° Celsius. They drove just after lunchtime, a time considered as risky in terms of drowsiness (Horne & Reyner, 1999). The road and traffic were generated with SCANer Studio®. A webcam located on top of the central screen of the simulator video-recorded the participants during the session to establish the ground truth (see below). The (static) simulator, provided by Oktal® and powered with SCANer Studio® software, is made of a real car seat, 3 video screens (24" in format 16/9 each, forming a tryptic), a steering wheel, pedals and a small screen (10") for the dashboard, located just behind the wheel. The driving environment was displayed at a resolution of 1280×1024 pixels onto the three forward screens providing a 210° horizontal forward field of view. A rear screen provided a 60° rear field of view, corresponding to the normal use of the central rearview and two side mirrors. A stereo sound system provided simulated engine, road, and traffic sounds. An example of the field of view is presented on the figure below, which has been added to figure 4. The simulated car had an automatic gearbox, so the driver had only access to the steering wheel, gas and brake pedals.

At the beginning of the session, the participants drove along a highway for roughly 90 minutes, then turned off the highway and drove for around 5 minutes to reach a city. Finally, they drove in an urban environment for roughly 5 minutes. There was no traffic during most of the highway stretch. The very monotonous environment (without event or traffic) was selected in order to induce drowsiness. Somewhere 2/3 of the way along, 22 cars appeared from the right of the highway, disappearing a few kilometers later. This sudden addition of traffic was intended to change the driver's level of drowsiness. Rossi and collaborators (2011) demonstrated that a driver is more susceptible to sleepiness in a

simulator with a monotonous scenario.

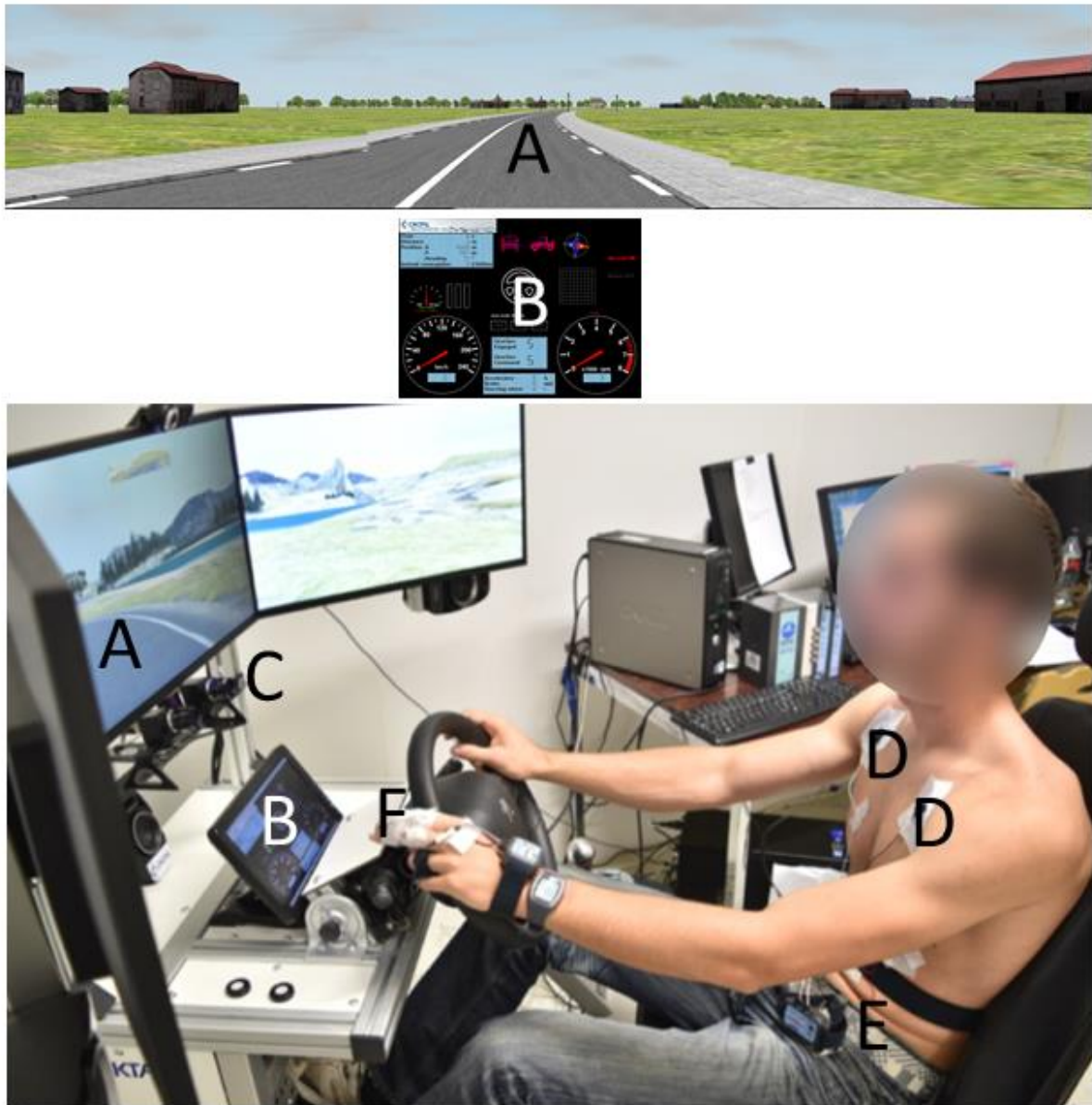


Figure 8-4 : The static driving simulator by Oktal®. A represents the road scene displayed on the 3 video screens. B represents the dashboard. C is the hardware faceLAB®. D are two of the three electrodes used for ECG, E is the respiratory belt. F are the electrodes for EDA (not used in this study because of important signal loss).

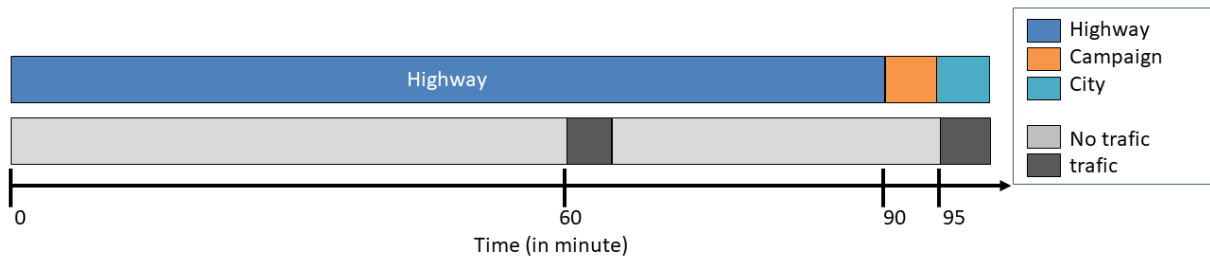


Figure 8-5: Diagram of the scenario with different types of road and the associated traffic

8.5.3.3 Data collection and processing (inputs)

During the driving, data on driving performance, eyelid and head movements, as well as physiological data, were recorded using the following hardware and software: SCANer Studio® for driving performance at 10 Hz, faceLAB® for sensorimotor signals at 60 Hz, and EKG, pulse plethysmography (PPG), respiration with the Biopac® MP150 system and Acqknowledge® software at 1000 Hz. Even if EEG is a gold standard, it is a method quite intrusive, so its use seems difficult in a context of industrialization in real cars. Several indicators were extracted from each source of information. The different indicators are summarized in table 1. The variables were recorded at a frequency of 1/60Hz (one by minute) because it is the lowest frequency common to all sources of information, including the ground truth. The participant information and the driving time (time elapsed since the beginning of the session) were both included as input based on the results of the previous study (Jacobé de Naurois et al., 2017).

Table 1: All the variables (grouped in columns by source of information) computed for each participant, averaged for each minute of driving time, and used as inputs for the ANNs.

Physiological data	Behavioral data	Car data
HR: Heart Rate (average and standard deviation) (beat/min)	Blink duration (average and standard deviation)	Lateral distance from the closest lane and the center of the car in m (average and standard deviation)
Svlf: HR signal Very Low Frequency Power (0.0-0.04 Hz)	Blink frequency (average and standard deviation) (per minute)	Time to lane crossing (average and standard deviation)
Slf: HR signal Low Frequency Power (0.04-0.15 Hz)	PERCLOS (average and standard deviation) (% of eye-closure time)	Steering angle (average and standard deviation)
Shf: HR signal High Frequency Power (0.15-0.4 Hz)	Head position x (average and standard deviation)	Steering angle velocity (average and standard deviation)
Svhf: HR signal Very High Frequency Power (0.4-3.0 Hz)	Head position y (mean and standard deviation)	Steering entropy (computed from steering angle)
Sympathetic ratio (Slf / (Svlf + Slf + Shf))	Head position z (average and standard deviation)	Number of direction changes (0-crossings) per minute (computed from steering angle)
Vagal ratio (Shf / (Svlf + Slf + Shf))	Head rotation x (average and standard deviation)	Accelerator pedal angle (average and standard deviation)
Sympathetic-vagal ratio (Slf / Shf)	Head rotation y (average and standard deviation)	Lateral shift of the vehicle center relative to the lane center (average and standard deviation)
Respiration Rate (average and standard deviation) (per minute)	Head rotation z (average and standard deviation)	Vehicle speed (km/h) (average and standard deviation)
	Saccade frequency (mean and standard deviation) (per minute)	Number of runs-off-road per minute

8.5.3.4 Model (ANN)

Driver's drowsiness was modeled with an artificial neural network created with the neural network toolbox (Beale et al., 1992) of Matlab R2013a. A feedforward neural network with one hidden layer optimized with the Levenberg-Marquardt algorithm (Levenberg, 1944) was used. The number of neural units in the hidden layer varied between 1 and 25 and was optimized via a grid-search method

(applied by steps of 2). This model was trained with a subset (n-1) of participants and adapted to a further participant according to the method described below (section 2.7).

8.5.3.5 Detection of the real level of drowsiness (ground truth)

The real (ground truth) level of drowsiness was determined based on a method proposed by Wierwille & Ellsworth (1994). Every minute of driving, two raters evaluated the driver’s state as ranging between 0 (alert state) and 4 (extremely drowsy). The mean of both raters was used as the drowsiness level. Inter-rater reliability was computed with Pearson's linear correlation (R=0.71 and p=0.00). Even if this subjective rating by a third (informed) person is validated by the “consistency and reliability in the rating produced” as stated by Wierwille & Ellsworth (1994), it is difficult to relate the ratings to the real state of drowsiness, thus one cannot conclude because the real drowsiness is a hidden measure, not directly accessible.

Table 2: Trained observer rating based on a scale by Wierwille and Ellsworth (1994) cited by Rost et al. (2015)

Level	Drowsiness State	Video image indicators
0	not drowsy	Normal fast eye blinks, often reasonably regular; Apparent focus on driving with occasional fast sideways glances; Normal facial tone; Occasional head, arm and body movements.
1	slightly drowsy	Increase in duration of eye blinks; Possible increase in rate of eye blinks; Increase in duration and frequency of sideways glances; Appearance of “glazed eye” look; Appearance of abrupt irregular movements – rubbing face/eyes, moving restlessly on the chair; Abnormally large body movements following drowsiness episodes; Occasional yawning.
2	moderately drowsy	Occasional disruption of eye focus; Significant increase in eye blink duration; Disappearance of eye blink patterns observed during alert state; Reduction on degree of eye opening; Occasional disappearance of facial tone;

		Episodes without any body movements.
3	very drowsy	Discernable episodes of almost complete eye closure, eyes never fully open; Significant disruption of eye focus; Periods without body movements (longer than for level 2) and facial tone followed by abrupt large body movements.
4	extremely drowsy	Significant increase in duration of eye closure; Longer duration of episodes of no body movement followed by large isolated "correction" movements.

8.5.3.6 Outputs of models

As in the previous study, the goal here was both to detect the current level of drowsiness and to continuously predict when the driver's state would reach a given threshold. We therefore used two ANNs, one for detection, the other for prediction, as follows. After appropriate training, the first ANN detected a level of drowsiness with an output in the range 0-4 by steps of 0.5. If detected drowsiness was lower than 1.5, the second ANN predicted (in minutes) when it would reach 1.5: this time was its output, otherwise 0. The threshold was set at 1.5, meaning that at a given time, one of the two raters evaluated the state of the participant as moderately drowsy (level 2) while the other rated it as slightly drowsy (level 1). The impaired state was defined according to the level of drowsiness and not driving performance or event detection performance because there is no direct and reciprocal relation between driving performance and driver's level of drowsiness. The driving performance is not necessarily impaired when the driver shows signs of sleepiness (Philip et al., 2005).

8.5.3.7 Methods for adaptive learning with the ANN

The present study set out to test whether the ANN can efficiently be adapted to each specific driver. The following methodology was used to create the adaptive ANN (hereafter termed Ad-ANN). Figure 6 presents the overall process, in two phases: a classic training phase (steps 1, 2a, 3, 4), similar to our previous study (Jacobé de Naurois et al., 2017), and an adaptation phase (steps 2b, 5 and 6) aimed at improving the performance of the system for a particular participant.

Step 1: a cross-subject validation (also known as a one-subject-left-strategy) was computed, testing a different participant at each validation cycle: one dataset was composed using data from 20 (n-1) participants to train the ANN and the other dataset was composed using data from the 21st

participant. This process was repeated 21 times, each time using a different participant's data as test set.

Step 2: the datasets were again divided into sub-datasets. The dataset composed of 20 participants was divided in two sub-datasets (step 2a) for a 5-fold cross-validation with random distribution: 80% for training the ANN and 20% for testing and validating the ANN (hereafter ANN-training dataset and ANN-validation dataset). The dataset from the remaining single participant/driver was also divided into two sub-datasets (step 2b): one composed of the first data segment recorded during the driving task, i.e. the first minutes of driving (hereafter Ad-ANN-training dataset) and the other composed of the remaining data (hereafter Ad-ANN-validation dataset). Between 1 and 30 minutes of recorded data were used in the Ad-ANN-training dataset, i.e. between 1 and 30 lines of the dataset. This time variable, hereafter τ , constituted the first parameter for the adaptation. Thus, if the first parameter was equal to 5 and the participant drove for 110 minutes, the Ad-ANN-training dataset contained 5 lines corresponding to the first 5 minutes, while the Ad-ANN-validation dataset contained 105 lines covering the 6th to the 110th minutes. This second step yielded four datasets: ANN-training dataset, ANN-validation dataset, Ad-ANN-training dataset, and Ad-ANN-validation dataset. The ANN-training dataset was used to train a general ANN on 20 drivers. The ANN-validation dataset was used to validate and choose the general ANN defining the number of neural units in the hidden layer. The Ad-ANN-training dataset was used to adapt the ANN (Ad-ANN), i.e. to personalize the ANN for each driver. Finally, the Ad-ANN-validation dataset was used to assess the performance of the adaptive ANN, i.e. its performance with data not previously encountered by the model.

Step 3: after dividing the datasets, the training dataset (ANN-training dataset) composed of 20 participants was used to train feedforward neural networks. This ANN is the general ANN (see Jacobé de Naurois et al., 2017 for more details).

Step 4: the ANN-validation-dataset was used to choose the best-performing neural networks (best number of neural units on hidden layer). The ANN with the lowest root mean square error (RMSE) was kept.

Step 5: the chosen ANNs were adapted using the AdANN-training dataset containing the first segment of recorded data from the 21st driver. This step restarts the training process, this time based on the AdANN-training dataset (hereafter, Ad-ANN). For the adaptation, different change ratios μ were tested from 10^{-5} to 10^3 . Thus, μ is the second parameter of adaptation, changing the weight of the Ad-ANN.

Finally (Step 6), the remaining dataset (Ad-ANN-validation dataset) was used to choose both best parameters (amount of data on Ad-ANN-training dataset τ and μ).

Adaptation personnalisée des modèles de détection et prédiction de la somnolence

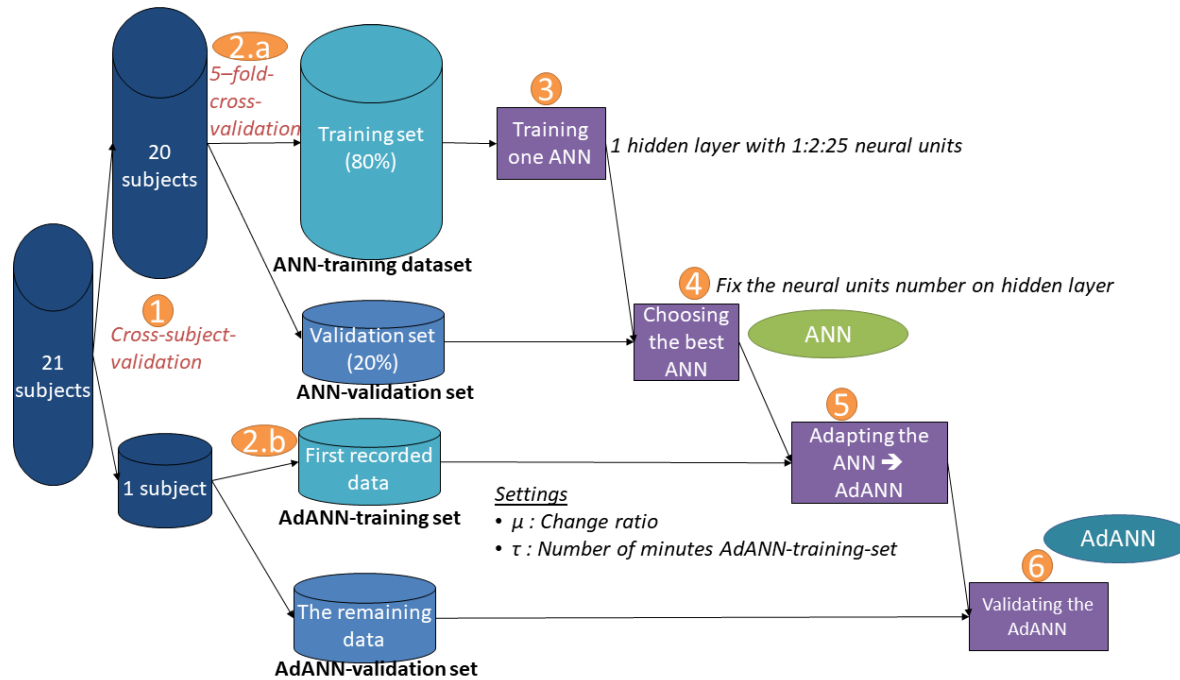


Figure 8-6: The training method used to adapt the ANNs. Cylinders represent different datasets. The small numbered circles represent the step of the process defined in part 2.7 Methods for adaptive learning with the ANN. Rectangles represent steps in the process.

8.5.3.8 Evaluation and performance

For each step described above, the performance function was the root mean square error (RMSE) between the network outputs and the target outputs. This performance metric was used to compare the error on each dataset: before the adaptation, using the general ANN and after the adaptation, using the Ad-ANN. The lower this metric, the better the model. After the adaptation, the RMSEs of the ANN-training dataset and the ANN-validation dataset were expected to increase while the RMSEs of the AdANN-training dataset and the AdANN-validation dataset were expected to decrease, because the model would now be adapted specifically to the data of the 21st driver. Both RMSE mean and standard deviation (SD) before (step 4) and after (step 6) the adaptation were compared to test this hypothesis, assuming that SD would also decrease after the adaptation. RMSE was also analyzed with respect to the amount of data used in the Ad-ANN-training dataset and in the AD-ANN-validation dataset.

8.5.3.9 Statistical analysis

In order to compare pre- and post-adaptation RMSE, a two-sample F-test for equal variances was performed before a two-sample matched T-test for means. Bonferroni corrections were performed for each value of τ to compare RMSE variation as a function of τ .

8.5.3.10 Subject-specific performance assessment of AdANN.

To assess the subject-specific performance of AdANN, the same learning process was repeated. This time, the ANN was trained with only 19 participants and adapted with the 20th participant (A). Finally, the performance of the ANN and the Ad-ANN were evaluated on the 21st participant (B). These two drivers (A and B) changed with each iteration. The ANN was adapted using the first segment of recorded data from driver A and the resulting AdANN was tested on driver B. To test subject-specific performance, RMSE on driver A and driver B before adaptation (general ANN) and after adaptation (Ad-ANN) were compared. Our hypothesis was that RMSE would decrease for driver A (since the model was specifically adapted to this driver) but would increase for driver B.

8.5.4 Results

One participant had sign of simulator sickness, so the session was stopped immediately, and the participant was excluded from the study. Some participants had road departure at different levels of drowsiness. Before analyzing the performance of models, it is important to note that all participants did not reach the same level of drowsiness at the same moment. Some participants reached the level “extremely drowsy” (level 4), while others only reached “moderately drowsy” (level 2) and this at different temporalities. Only one participant was a little particular because he reached at the maximum this level (moderately drowsy) after one hour of driving, which was a very long delay as compared to the others. We first (section 3.1) present the results with both best parameters, i.e.: those yielding the lowest RMSE (τ : amount of data in AdANN-training dataset and μ : change ratio during adaptation) with the different sources of information (“all”, “physiological”, “behavioral” and “car”). Next (section 3.2), we focus on the best source of information to analyze RMSE variation as a function of τ , i.e. the amount of data used to adapt the ANN for each participant. Finally (section 3.3), we examine the subject-specific performance of the procedure, presenting results on the adaptation performed with data from one driver and tested with data from another driver, again for the best source of information. Each section gives results on both detection and prediction modeling.

8.5.4.1 Best μ and best τ on AdANN-validation dataset

In this section, the lowest mean RMSE for the cross-subject validation is presented with both best τ and μ parameters, for each different source of information and for both detection and prediction ANNs.

8.5.4.1.1 Detection of level of drowsiness

The lowest average RMSEs pre- and post-adaptation for each source of information, and in each case for both best parameters, are presented in fig. 7. For all sources of information, mean and SD are significantly lower after adaptation than before adaptation. For the “all” category ($t=4.294$; $p<0.001$ and $F=25.826$; $p<0.001$ for RMSE mean and SD respectively), best performance is for $\tau=20$ min and $\mu=50$. For the “physiological” category ($t=2.335$; $p=0.027$ and $F=8.620$; $p<0.001$ for the mean and SD respectively), post-adaptation best performance is for $\tau=18$ min and $\mu=10^{-5}$. For the “car” category ($t=3.231$; $p=0.003$ and $F=24.420$; $p<0.001$ for the mean and SD respectively), best performance is for $\tau=19$ min and $\mu=5$. Finally, for the “behavioral” category ($t=3.231$; $p=0.003$ and $F=28.652$; $p<0.001$ for the mean and SD respectively), best performance is for $\tau=30$ min and $\mu=5$. RMSE is lowest when the “all” source of information is used. Furthermore, it is only after adaptation that the average RMSE (for

all participants) is lower than 1 level of drowsiness, i.e. model error is more or less one level of drowsiness. Model adaptation did not improve performance for some participants: 2 participants for the “all” source of information, 8 for “physiological”, 1 for “behavioral” and 5 for “car”, but it is worth noting that adaptation never resulted in higher error.

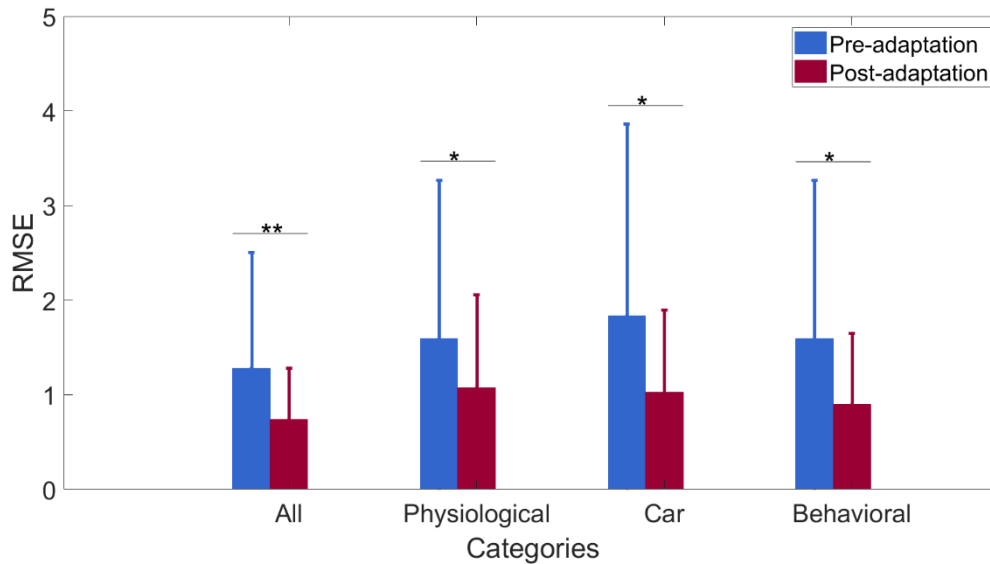


Figure 8-7: Root mean square error (RMSE) and standard error of the detected level of drowsiness, based on the AdANN-validation dataset, for different sources of information, before and after adaptation. Stars represent the level of significance (NS: $p > .05$; *: $p < .05$; **: $p < .01$; ***: $p < .001$).

8.5.4.1.2 Prediction of time of occurrence of impaired driver state

The lowest average RMSEs pre- and post-adaptation for both parameters (τ and μ) are presented in Fig. 8. In terms of prediction of driver impairment, once again, for all sources of information, mean and SD are significantly lower after adaptation. For the “all” source of information category ($t=3.349$; $p=0.003$ and $F=179.079$; $p<0.001$ for the mean and SD respectively), best performance is with $\tau=30$ min and $\mu=10$. For the “physiological” category ($t=4.262$; $p=0.004$ and $F=14.051$; $p<0.001$ for the mean and SD respectively), best performance is with $\tau=29$ min and $\mu=0.5$. For the “car” category ($t=3.938$; $p=0.001$ and $F=1235.924$; $p<0.001$ for the mean and SD respectively), best performance is with $\tau=30$ min and $\mu=10$. Finally, for the “behavioral” category ($t=3.938$; $p<0.001$ and $F=37.474$; $p<0.001$ for the mean and SD respectively), best performance is with $\tau=28$ min and $\mu=10^{-1}$. The lowest RMSE is achieved with “behavioral” data, but this is not statically different from other sources of information. Moreover, it is only after adaptation that the mean of RMSE is lower than 6 minutes, i.e. the model error is more or less 6 minutes. As with detection, however, adaptation did not improve the prediction performance of the model for some participants (9 participants for the “all” source of information, 3 for

“physiological”, 4 for “behavioral” and 4 for “car”). Again, model performance never decreased after adaptation.

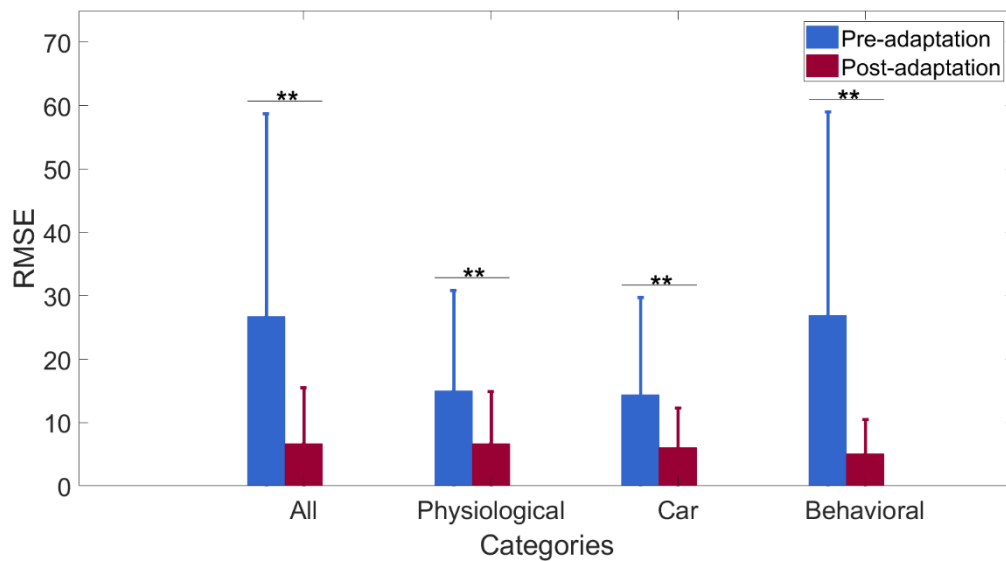


Figure 8-8: mean of RMSE and SD of the predicted time of occurrence of drowsiness level 1.5, based on AdANN-validation dataset, for different sources of information, before and after adaptation. Stars represent the level of significance of the difference in means (NS: $p > .05$; *: $p < .05$; **: $p < .01$; ***: $p < .001$).

8.5.4.2 Effect of varying the amount of data (τ) used for the adaptation

Detailed results on the best source of information with variations of τ , the amount of data used in the adaptation dataset (AdANN-training dataset), for both detection and prediction are outlined below. The “all” dataset yields best detection of drowsiness and the “behavioral” dataset best prediction of drowsiness. For each τ chosen (amount of data), the best μ was chosen. Both pre- and post-adaptation variations in RMSE are presented for the ANN-training dataset, the ANN-validation dataset, the AdANN-training dataset and the AdANN-validation dataset.

8.5.4.2.1 Detection of level of drowsiness

Figure 9 shows the mean and the SD of RMSE for the cross-subject validation using the different datasets as a function of τ (amount of data in the AdANN-training dataset used to adapt the ANN). First, for both the ANN-training and the ANN-validation datasets, RMSE mean and SD are significantly higher after adaptation, whatever the amount of data used (for the mean, t-values between -2.787 and -3.968; $p < 0.01$ for the first value of τ and t-values between -4.440 and -5.487; $p < 0.001$ for $\tau = 7$ to 30, for the SD, $p < 0.001$; F-values between 0.002 and 0.018 in all cases). For the AdANN-training dataset, RMSE mean and SD are significantly lower after adaptation for each τ (for the mean t-values

between 3.592 and 5.896; $p \leq 0.001$ for all cases, for the SD F-values between 3.440 and 11.713; $p < 0.001$ for all τ except 2, 3 and 6: $F=6.691$; $p=0.001$, $F=4.658$; $p=0.008$ and $F=4.469$; $p=0.002$). However, for the AdANN-validation dataset, RMSE is significantly lower when three minutes are used to adapt the ANN (t-values between 2.111 and 2.572; $p < 0.05$ for $\tau = 3$ to 7, t-values between 2.906 and 3.618; $p < 0.01$ for $\tau = 8$ to 12 and t-values between 3.869 and 4.292; $p < 0.001$ from the 13th minute to the end). SD is significantly smaller since $\tau=8$ to 30 (for $\tau=8$, $F=2.670$; $p=0.033$, for 9, $F=3.176$; $p < 0.05$, for $\tau=10$, $F=3.176$; $p = 0.013$, for $\tau=11$, $F=4.543$; $p < 0.01$ and from $\tau = 12$ to 30, F-values between 5.525 and 7.369; $p < 0.001$). Moreover, the mean RMSE and SD of the AdANN-validation dataset decreases as a function of the amount of data used in the AdANN-training dataset, affording more accurate performance. In addition, there is a roughly 40% improvement in the performance of the AdANN-validation dataset after adaptation.

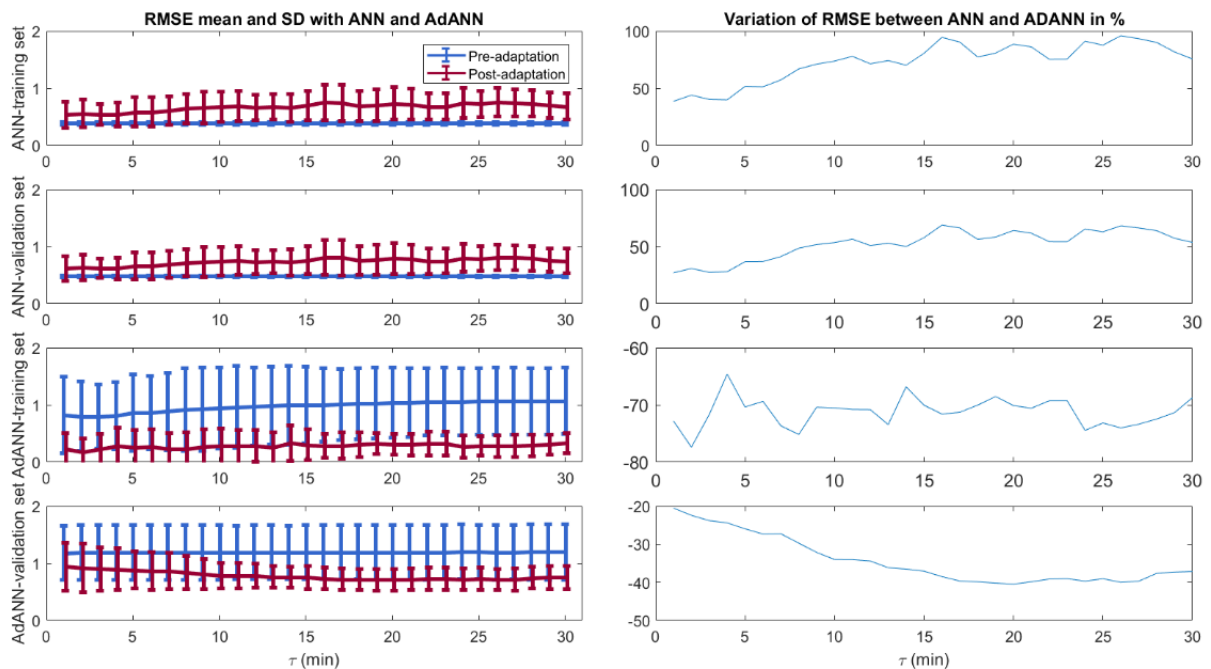


Figure 8-9: Detection: Mean and SD of RMSE for the different datasets (ANN-training set, ANN-validation set, AdANN-training set, AdANN-validation set) as a function of amount of data (τ , in minutes) used to adapt the ADANN-training dataset.

Comparison of RMSE between the different τ for the Ad-ANN-validation dataset (Fig. 10) shows that compared to $\tau=1$, RMSE mean is significantly higher for $\tau = 2$ to 30 (t-values between 3.945 and 4.517). Compared to $\tau=2$, RMSE mean is significantly higher for $\tau = 6$ to 27 (t-values between 3.774 and 4.228). Thus, more than 2 time-samples are required to significantly improve the performance of the ANNs.

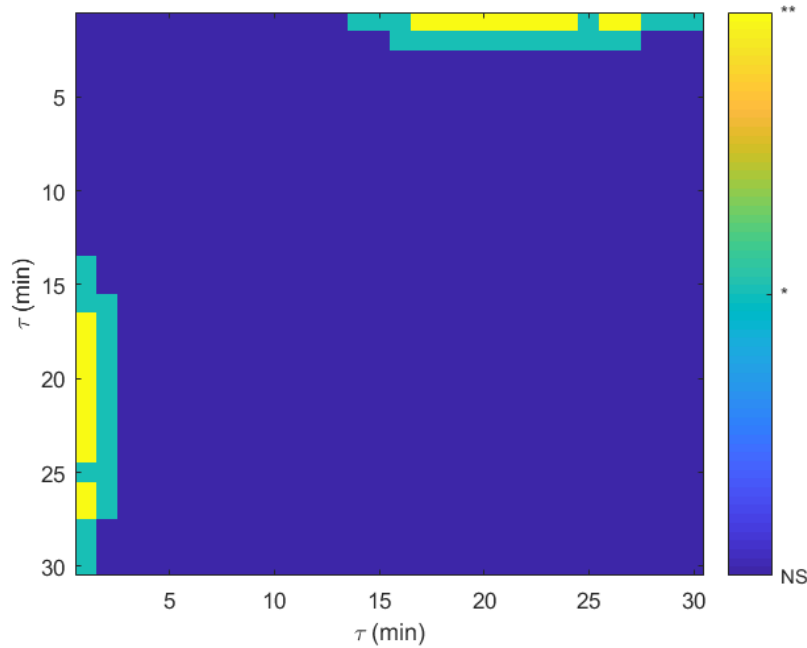


Figure 8-10: P values for RMSE means compared between the different τ for the Ad-ANN-validation dataset after adaptation.

8.5.4.2.2 Prediction of time of occurrence of impaired driver state

Figure 11 shows the mean and the SD of RMSE for the cross-subject validation for the different datasets as a function of the amount of data in the AdANN-training dataset used to adapt the ANN. For both the ANN-training and the ANN-validation datasets, RMSE increases significantly after adaptation (for $\tau = 1$ to 20, t-values between -5.239 and -3.340; $p < 0.01$ and for $\tau = 1$ to 21 to 30, t-values between -6.021 and -4.657; $p < 0.001$). SD is always significantly higher for both ANN datasets (F-values between 0.000 and 0.002; $p < 0.001$). For the AdANN-training dataset, the RMSE mean is always significantly lower after adaptation (t-values between 2.272 and 5.030; $p < 0.001$) but the SD is only lower for $\tau = 8$ to 30 (for $\tau = 8$ to 11, F-values between 2.642 and 3.300; $p < 0.05$, for $\tau = 12$ to 19, F-values between 3.300 and 3.245; $p < 0.01$, for $\tau = 20$ to 30, F-values between 4.797 and 9,255; $p < 0.01$). For the AdANN-validation dataset, mean RSME is significantly lower after adaptation (for $\tau = 1$ minute, $t = 3.535$; $p = 0.001$; t-values between 2.272 and 5.031; $p = 0.001$ for $\tau = 2$ to 30), while SD is significantly lower after adaptation for $\tau = 8$ to 30, (for $\tau = 8$ to 11, F-values between 2.642 and 3.300; $p < 0.05$, for $\tau = 12$ to 19, F-values between 3.378 and 3.245; $p < 0.01$, for $\tau = 20$ to 30, F-values between 4.797 and 9.255; $p < 0.001$). For the AdANN-validation dataset, there is a small decrease followed by a stabilization of RMSE as a function of τ .

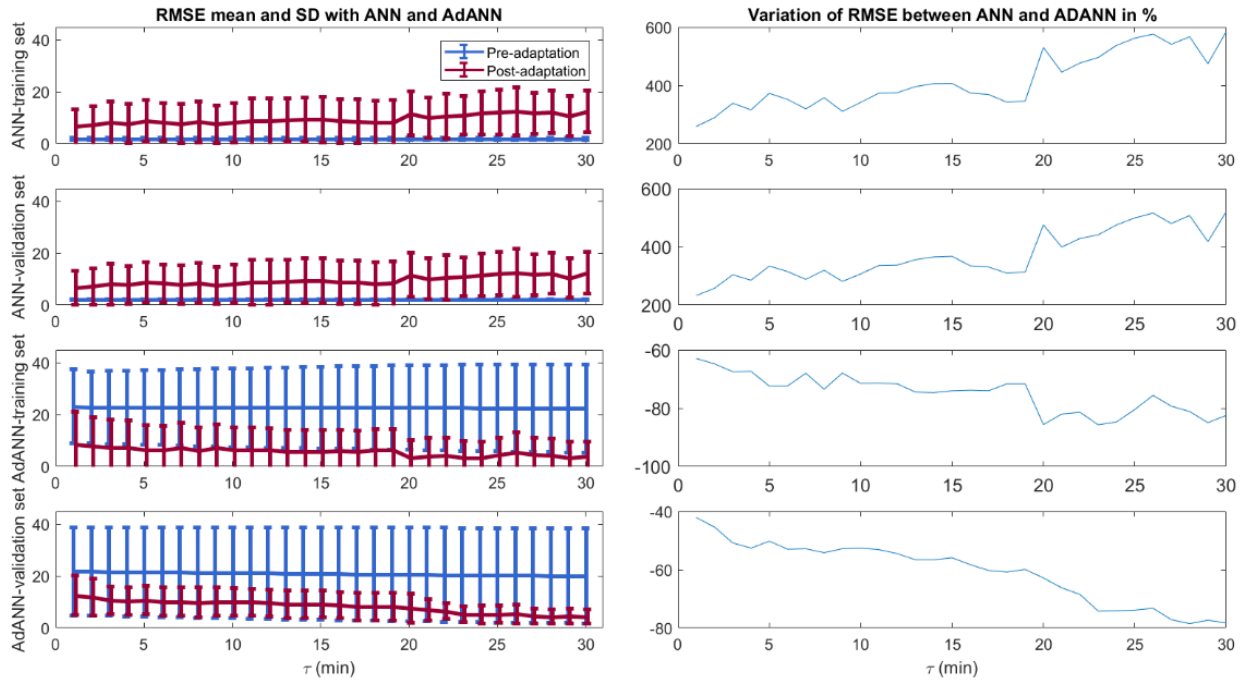


Figure 8-11: Prediction: Mean and SD of RMSE for the different datasets (ANN-training set, ANN-validation set, AdANN-training set, AdANN-validation set) as a function of amount of data (τ , in minutes) used to adapt the AdANN-training dataset.

This can be confirmed by determining the p-value, comparing each τ with each other τ (Fig. 12). Compared to $\tau = 1$, RMSE mean is significantly lower for $\tau = 21$ to 30 (t-values between 4.377 and 7.010). Compared to $\tau = 2$, RMSE mean is significantly lower for $\tau = 21$ to 30 (t-values between 4.658 and 6.102). Compared to $\tau = 3$, RMSE mean is significantly lower for $\tau = 23$ and 27 to 30 (t-values between 3.933 and 4.361). Compared to $\tau = 4$, RMSE mean is significantly lower for $\tau = 27$, 28 and 30 (t-values between 3.885 and 4.016). Compared to $\tau = 5$, RMSE mean is significantly lower for $\tau = 23$ to 30 (t-values between 3 and 4.444). Compared to $\tau = 6$, RMSE mean is significantly lower for $\tau = 28$ and 30 (t= 3.932 and 3.913, respectively). Compared to $\tau = 7$, RMSE mean is significantly lower for $\tau = 28$ and 30 (t=3.955 and t=3.936, respectively). Compared to $\tau = 8$ to 30, RMSE is not significantly different; thus RMSE can be considered as stable at this level. Moreover, there is an almost 80% improvement in performance in the AdANN-validation dataset after adaptation.

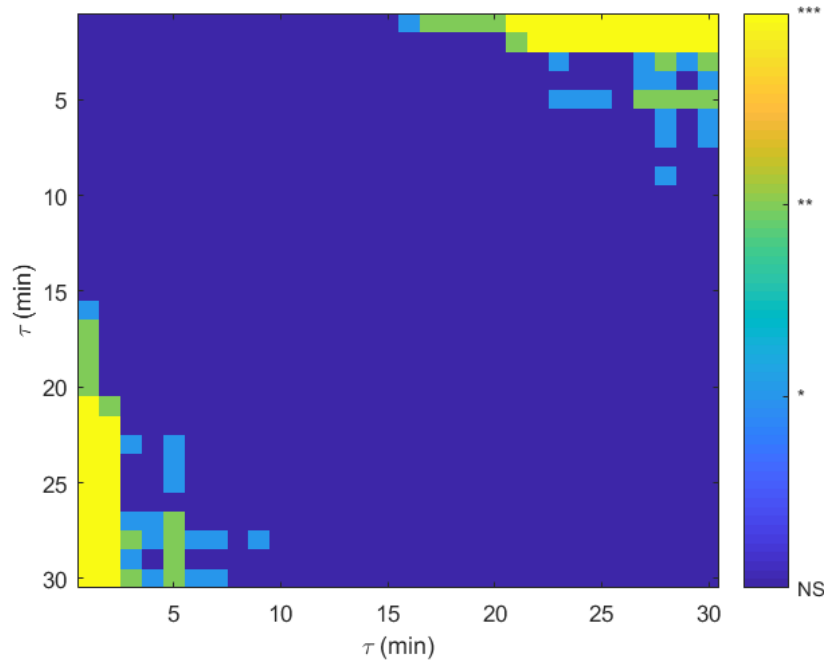


Figure 8-12: P-value for RMSE mean comparing each τ with each other τ for the AD-ANN-validation dataset after adaptation.

8.5.4.3 Subject-specific performance of ANN adapted to one driver and tested with another

In this session, the ANN was trained with data from 19 participants and adapted with a 20th called A. Then the AdANN was tested on the 21st participant, called B. Applied to all participants, both for the detection and the prediction of drowsiness, this procedure was performed 21x20 times (with a different A and B each time). Our objective was to determine whether the adaptation is specific to one participant or can improve model performance for another participant whose data is unknown to the ANN. To limit calculation time, the best parameters (number of neural units in the hidden layer, μ and τ values used for adaptation and best sources of information) defined in the first section of results (see 3.1) were chosen.

8.5.4.3.1 Detection of level of drowsiness

On average (Fig. 13), for the participant used for the adaptation (A), RMSE mean and SD are significantly lower after adaptation (for the τ first segment of the data, $t=10.696$; $p<0.001$, $F=14.213$; $p=$ and for the remaining data, $t=11.636$; $p<0.01$, $F=10.496$ $p=$ for mean and SD respectively). For the other tested participant (B), RMSE mean and SD are significantly higher after adaptation (for the τ first segment of the data, $t=-2.737$; $p=0.006$, $F=0.482$; $p<0.001$ and for the remaining data, $t=-2.028$; $p=0.043$, $F=0.670$; $p<0.001$ for mean and SD respectively).

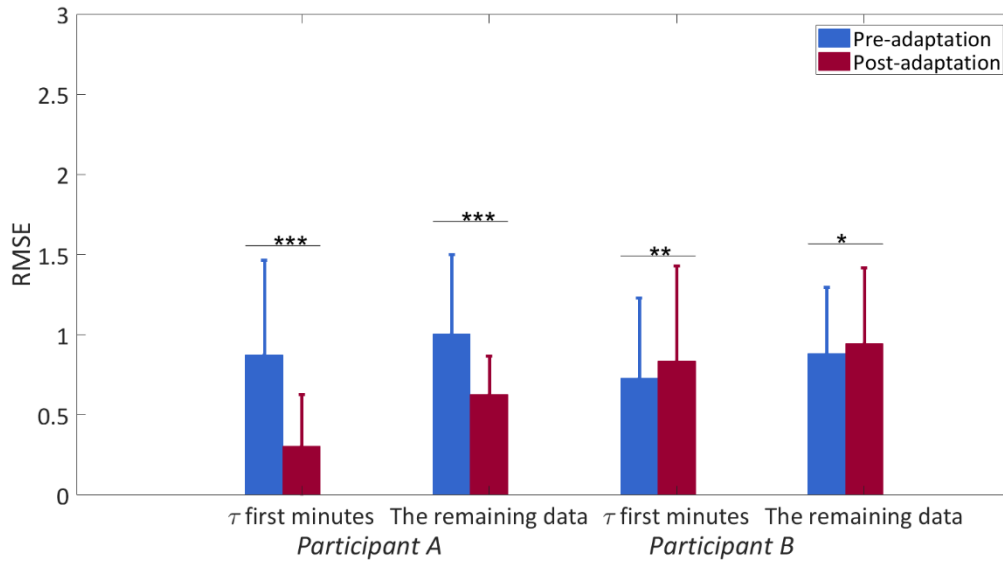


Figure 8-13: Detection of level of drowsiness: RMSE mean and SD for participants used for the adaptation (A) and other participants (B) never previously encountered by the ANN, before and after adaptation with the τ first data recorded on participant A. Best τ is used here. Stars represent the level of significance (NS: $p > 0.05$; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$).

8.5.4.3.2 Prediction of time of occurrence of impaired driver state

On average (Fig. 14), participant A's RMSE mean and SD are significantly lower after adaptation (for the τ first segment of the data, $t=10.999$; $p < 0.001$, $F=23.454$; $p < 0.001$ and for the remaining data, $t=8.431$; $p < 0.001$, $F=80.829$; $p < 0.001$ for mean and SD respectively). The other tested participant B's RMSE mean and SD are significantly higher after adaptation (for the τ first segment of the data, $t=-2.729$; $p=0.007$, $F=0.482$; $p < 0.001$ and for the remaining data, $t=-2.507$; $p=0.012$, $F=0.348$; $p < 0.001$ for mean and SD respectively). In the case of this B, an outlier participant was removed from the analysis because post-adaptation RMSE is too high (around 10^5), although adaptation also increases RMSE for this outlier participant.

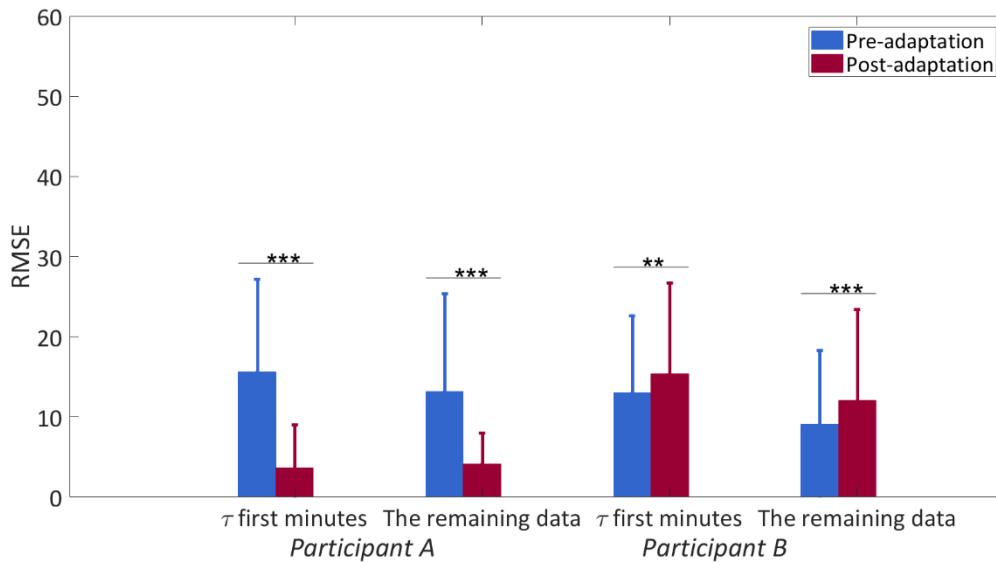


Figure 8-14: Prediction of impaired driver state: RMSE mean and SD for participants used for the adaptation (A) and other participants never previously encountered (B), before and after adaptation with the τ first data recorded on participant A. Stars represent the level of significance (NS: $p > .05$; *: $p < .05$; **: $p < .01$; ***: $p < .001$).

8.5.5 Discussion

To maintain performance and safety, driving a car requires physiological and cognitive resources. While much attention has been given to detect driver drowsiness and, more recently, to predict a degraded state (Jacobé de Naurois et al., 2017), predicting the time of occurrence of a given state of drowsiness remains a challenge. A major issue is the inter-individual variability found even in a very monotonous situation. Seeking a methodology for driver-specific detection and prediction of drowsiness, the objective of this study was to test whether a model trained with a limited set of driver data could be easily and efficiently adapted to a further driver, using only a limited set of data from this driver. The inter-rater reliability about the ground truth was high as the two independent ratings were quite close enough, the trend was the same for both raters, excepted that in a few cases one rater detected a change a little bit in advance (in a range between 1 to 5 minutes). Inter-rater correlation showed a similar value of R than in Weirwille and Ellsworth. (1994). Nonetheless, a potential bias in detection might be suspected from the fact that the ground truth is based on subjective evaluations from video recordings of the participant's motor behavior, which could be thought to explain the superior performance of the behavioral dataset. However, it is worth noting that these features are consensually described in the literature as the most objective and pertinent indicators of drowsiness. It is therefore difficult to conclude on whether the high performance of a model trained with behavioral data is due to the way ground truth is set or to the greater relevance of this particular set of data.

Secondly, not all types of data were found to have the same effect on adaptation. The model performs best on prediction using behavioral data alone, while the best performance on detection is achieved using the full dataset (behavioral, physiological, car data, as well as personal information and driving time). In this latter case, however, there is no significant difference between the different combinations of dataset. Similar to our previous study (Jacobé de Naurois et al., 2017), Daza et al., (2014) obtained better results with features extracted from eyelid movements (such as PERCLOS) than with features extracted from car data. Taken together, these results suggest a predominant role of oculomotor data compared to other types of data. Furthermore, when using different features, the neural network has to learn the dependencies between these different kinds of information. While this poses no problem in terms of detection, it is more of an issue for temporal prediction. When the aim is prediction, adding different types of information could be counterproductive due to the added complexity.

Thirdly, our innovative study shows that adapting an ANN to a participant whose data were not included in the initial ANN training, using only a few minutes of recorded data, can significantly improve the model's performance for this participant. This is consistent with the findings of another study in detection of drowsiness with transfer learning and EEG signals (Dongrui Wu et al., 2015). In other words, the performance of the ANN, for each new participant, is enhanced after adaptation. For detection of drowsiness, the performance of the model is improved by about 40%, and for prediction, improvement reaches 80%. The difference in performance improvement between the two models can be explained by the fact that the outputs have different scales, or granularities: drowsiness is coded with 9 levels, between 0 and 4 by steps of 0.5, while time before an impairment state is between 0 and 60, by steps of 1 minute. The improvement after adaptation corresponds to $\frac{1}{2}$ level for detection of drowsiness and 15 minutes for prediction. The improvement in performance is significant above a threshold of data input used to adapt the model: two minutes (for prediction) or three minutes (for detection). In addition, increasing data up to 15 minutes further improves performance, though using an even larger amount of data (from 16 to 30 min) does not significantly improve performance further. An important result is that performance is improved using only a limited amount of data for model adaptation. This can probably be explained by the fact that our ANNs are simple, not deep, networks. It is worth noting that the adapted ANN performs much better on the participant used for the adaptation. Its performance is worse when applied to data used for the initial (general) training, and even worse when applied to a previously unencountered participant, whose data were not used to adapt the ANN. Thus, adaptation with data from a particular participant makes the ANN very subject-

specific, and less efficient for others. To our knowledge, this particular angle has never before been explored in the context of adaptation, at least concerning driver drowsiness.

However, while adapting the model to a previously unencountered participant gives better results than testing this new participant with a model trained on data from others, does adaptation really improve ANN performance compared to its performance with the participants used for the training? To answer this question, we need to compare the present results (with adaptation) to results from the previous study (Jacobé de Naurois et al., 2017), which used random 10-fold cross-validation with the same protocol.

In this previous study, the average RMSE (between different sources of information) was 0.49 (1/2 level) for detection of drowsiness and 2.33 minutes for temporal prediction of drowsiness. In the present study, the average RMSE after adaptation to a new participant is 0.93 (less than 1 level) for detection of drowsiness and 6.15 minutes for prediction. Therefore, adapting the ANNs (with a limited amount of data) is not as efficient as training the ANN with a larger set of data from this same participant, among others. However, our results show that, although less efficient, the adaptation method is also more parsimonious, and allows the model to be rapidly tuned to a new driver.

The difference in performance between the two methods may also be due to the high inter-individual variability of the effects of drowsiness on driving performance and on differences in physiological signals (C. C. Liu et al., 2009; Van Dongen, Baynard, et al., 2004; Van Dongen, Maislin, et al., 2004). This may increase the gap between what is learned during the training phase and what is observed during the testing phase, due to overfitting. In machine learning, overfitting occurs when a greater difference in error is observed for the test dataset than for the validation dataset. One solution may be to train one ANN for each participant. However, this would obviously require a large amount of data from each driver, and thus quite a long driving time to collect this data, before the ANN was operational. Yet an ANN previously trained with a larger set of data only needs a limited dataset to be quickly (though less accurately) adapted (customized) to a new driver. It can be speculated that extending the adaptation, i.e. collecting more data from the new driver, might further improve performance and customization. This was not possible in our study, since using more data for adaptation means less data for the test phase (this is why here we limited the range of τ to only 1 to 30 minutes, and reserved the remaining data, from $\tau+1$ to 110 minutes, as the test set).

Other studies on transfer learning or adaptive learning, for example through a BCI (Brain Computer Interface), either randomly selected data from one participant as their dataset for transfer or adaptation, or used a dataset from another session (Wei et al., 2015; Dongrui Wu et al., 2015, 2016). To our knowledge, the present study is the first to use data initially recorded within the session to

adapt the ANN for drowsiness detection. Actually, using randomly selected lines of data in a real car would probably give more information about variability; however, it would significantly increase the time required for the adaptation since it would involve more data collection.

Finally, while the present study addressed inter-individual variability, not intra-individual variability, the latter is also a highly relevant issue. Obviously, people do not behave and react exactly in the same way on different days, nor do they present the same tendency to fall asleep during driving. For a given individual, a large range of factors may influence the risk of becoming drowsy, such as quality and duration of sleep, daytime activity, health, medication, drug and/or alcohol use, etc. Although one individual's performance was found to be stable after sleep deprivation repeated on different days (Van Dongen, Baynard, et al., 2004), a recent study on a car driving simulator (Nilsson et al., 2017) found intra-individual differences when comparing several day-time and night-time sessions. Here, with the dataset collected and the modeling approach used, it is impossible to decide whether adaptation is only necessary once or whether an adaptation phase is needed every time a given individual starts driving on a new day. To address this issue, the same participant would have to be tested at least a second time, on a different day, as proposed by Zhang, Wang, & Li (2017). This would involve a further study addressing intra-individual variability and assessing the optimal frequency of adaptation renewal.

Obviously, the number of participants used here to create the generalized model and test its customization was limited. We deliberately chose to study only a specific population (young people) at a particular time of day (post-lunch, attention dip) and in a static driving simulator. Different drowsiness dynamics are usually observed in different conditions (Anund, Ahlström, Fors, & Åkerstedt, 2018; Fors, Ahlstrom, & Anund, 2018). The transfer of knowledge from a younger population to an older one, from one-time slot to another, from a non-professional driver to a professional driver (Anund et al., 2018), or from a driving simulator to real road driving, are situations where transfer learning and adaptive learning could be further tested. Such an approach would allow the implementation of a general model (for instance in a new car), its progressive adaptation to a new driver, and longer term, its continuous adaptation to this driver to take account of changes in his/her behavior, mood, health, etc.

8.5.6 Conclusion

In this study, ANNs were adapted for a new participant from very limited data and used either to detect this driver's drowsiness level or to predict the onset of an impaired driving state. The ANNs were trained on a group of individuals and subsequently adapted for each specific participant to make

allowance for high inter-individual variability. Our results with this new method using adaptive learning confirm that the ANN can rapidly be made subject-specific. This adaptation to a specific driver's data provides a promising first response to the challenge of high inter-individual variability, although other issues like intra-individual variability and/or driving on different days, time-on-day or different road conditions remain to be addressed. Moreover, other individual information could also be tested and added in the models such as for instance quality of rest/hours since last drive, drive task demand, lack of demands/monotony, personality trait as other methods of transfer learning... In the future, the model, trained beforehand with a limited set of drivers and used for a detection and prediction of drowsiness will probably need to be adapted to each new driver, and also updated (for a given driver) at a frequency still to be determined, in order to maintain the model in phase with potential evolutions of the driver's characteristics and behavior.

8.6 Synthèse des résultats du second article

Une forte variabilité inter-individuelle est observée dans le comportement et dans les signaux recueillis. Elle soulève donc des questions sur les performances de détection et de prédiction de la somnolence au volant. Ainsi, lorsque le modèle est entraîné sur un ensemble de conducteurs et testé ensuite sur un autre conducteur, les performances de généralisation peuvent alors être mauvaises (chapitre 7). Pour contrer ce problème, cet article a mis en évidence l'intérêt d'entraîner un modèle général sur un groupe de conducteurs puis ensuite de le personnaliser sur un conducteur différent grâce à ces premières données d'enregistrement de la conduite. A cette fin, un ANN a été entraîné sur 20 conducteurs puis adapté sur un 21^{ème} conducteur (entraîné de nouveau grâce aux premières données d'enregistrement).

Ce processus d'adaptation a permis d'améliorer les performances de généralisation de 40 % pour la détection du niveau de somnolence et de 80 % pour la prédiction de la somnolence. Quant à la détection du niveau de somnolence, nous obtenons les meilleures performances en utilisant la combinaison de tous les jeux de données et une adaptation réalisée grâce aux 20 premières minutes d'enregistrement. Pour la prédiction de la somnolence, nous obtenons les meilleures performances par des données comportementales assorties d'une adaptation réalisée à l'aide des 28 premières minutes d'enregistrement. En outre, dès la première minute pour la prédiction ou dès trois minutes pour la détection, les performances sont significativement plus faibles après adaptation. Par ailleurs, nous avons montré qu'en entraînant le modèle sur 19 conducteurs, en l'adaptant sur le 20^{ème} et en le testant sur le 21^{ème}, nous avons mis en valeur que les performances se dégradent sur ce 21^{ème} conducteur. Ces tests montrent la spécificité de l'adaptation à un certain conducteur.

Chapitre 9 : Discussion et conclusion générale

9.1 Synthèse et avancées de l'ensemble des résultats

La conduite automobile requiert la mobilisation de ressources cognitives et physiologiques dans l'objectif de maintenir à la fois les performances et la sécurité du conducteur et des passagers. Il est donc aisément compréhensible que de nombreux chercheurs aient étudié la problématique de la somnolence au volant, considérée comme un superviseur majeur des capacités de conduite. Cette thématique revêt en effet un tel intérêt que détecter le moment où le conducteur est somnolent, et même prédire dans combien de temps il risque de le devenir sont devenus deux défis importants pour le développement de nouveaux systèmes d'aide à la conduite (Advanced Driving Assistance Systems, ou ADAS). Cette thèse s'est ainsi non seulement intéressée à la détection de la somnolence, c'est-à-dire l'estimation à un instant t du niveau de somnolence du conducteur, mais également à la prédiction de la somnolence, c'est-à-dire l'estimation du temps restant avant l'occurrence d'un certain niveau critique de somnolence.

Tout d'abord, nous avons mis en valeur qu'un environnement routier monotone amène les personnes à être somnolentes (Ahlström et al., 2018; Larue, 2010). Les changements de contexte tels que le type de route (route de campagne, ville, autoroute) ou la densité du trafic provoquent des variations de la dynamique de la somnolence (Ahlström et al., 2018). Des différences, concernant les dynamiques de la somnolence, sont également observées entre les individus ; elles font alors référence à la notion de variabilité inter-individuelle. Dans une même situation, tous les individus ne sont pas amenés à somnoler, de même que face à un niveau de somnolence défini, tous les individus ne vont pas manifester le même comportement ou les mêmes variations physiologiques (Van Dongen, Baynard, et al., 2004; Van Dongen, Maislin, et al., 2004) (chapitre 4). C'est pourquoi dans le chapitre 6, nous avons montré qu'à l'aide d'un ANN, il était non seulement possible d'estimer le niveau de somnolence (détection) mais également d'estimer le temps avant l'occurrence d'un certain niveau critique (prédiction). Ces deux estimations sont plus performantes lorsque les ANN possèdent comme données d'entrée le temps de conduite, des informations personnelles par rapport aux participants et des indicateurs sensorimoteurs relatifs aux mouvements de paupières et de tête (chapitre 6). Néanmoins, lorsque ces ANN sont entraînés sur un groupe d'individus et testés sur un autre individu, nous avons alors observé que les performances de généralisation de ces modèles sur ce nouveau participant, servant de test, ne sont pas satisfaisantes (chapitre 7). Pour cette raison, nous avons mis en place une adaptation des modèles : l'ANN a été entraîné sur un groupe de conducteurs puis adapté grâce aux premières données d'enregistrement sur un nouveau conducteur. Dans cette dernière configuration, nous obtenons de bien meilleures performances des modèles sur un nouveau

conducteur. Nous observons alors un transfert de connaissances apprises par les modèles sur un nouveau conducteur grâce à seulement quelques minutes de données d'enregistrement (chapitre 8).

Les enjeux et les objectifs de nos travaux ont pour rôle de mettre en valeur la possibilité de pouvoir anticiper une dégradation de l'état du conducteur grâce à une modélisation nécessitant des capteurs peu intrusifs pouvant se généraliser sur de nouveaux conducteurs. Les avancées et résultats majeurs de ces travaux sont donc le fait de savoir désormais détecter l'état du conducteur comme continu (d'alerte à extrêmement somnolent) et de pouvoir prédire le temps d'occurrence de l'épisode de somnolence grâce une modélisation performante, à l'aide de mesures relativement peu intrusives, par l'ajout d'informations contextuelles et par une adaptation accompagnée d'un transfert de connaissances. Chacune de ces avancées fera l'objet d'une discussion. Par la suite, nous envisagerons les principales limites de ces travaux, c'est-à-dire le nombre restreint de participants, la vérité de terrain (le véritable état du conducteur), les conditions de l'expérience en elle-même et le type de modélisation choisie. Enfin, nous finirons par une mise en relief de différentes perspectives possibles concernant la variabilité intra-individuelle, le niveau d'automatisation du véhicule et l'intégration de contre-mesures.

9.2 Détecter l'état du conducteur de manière continue

Dans le cadre de la modélisation de l'état du conducteur, nous avons fait le choix d'estimer l'état du conducteur comme une variable continue, variant de façon continue au cours du temps. Il était donc nécessaire de connaître le véritable état du conducteur au cours du temps de manière continue, ce qui imposait une évaluation fréquente (une fois par minute dans notre cas). Nous avons utilisé une méthode d'évaluation subjective du comportement du conducteur par deux personnes externes en se basant sur une vidéo du conducteur. Le *ground truth*, la vérité de terrain concernant la somnolence, a été déterminée entre 0 « alerte » et 4 « extrêmement somnolent » avec un pas de 0,5 minute après minute. La connaissance sur l'état du conducteur était donc plus riche qu'un simple état binaire alerte/somnolent. Notre but est de modéliser cet état à partir de mesures indirectes grâce à des modèles d'apprentissage automatique. En apprentissage supervisé, il existe deux types de problèmes : les problèmes de classification (la sortie du modèle est catégorielle) et les problèmes de régression (la sortie du modèle est continue). Ainsi, nos choix nous ont imposé d'utiliser des modèles de régression plutôt que de classification, contrairement à une précédente étude utilisant également cette vérité de terrain qui a été ensuite divisée en alerte (niveaux 0 et 1) et somnolent (niveaux 2, 3 et 4) (McDonald, Schwarz, Lee, & Brown, 2012). En effet, ce choix nous a permis d'augmenter la précision sur l'état du

conducteur. Tout d'abord, nous verrons en quoi il n'est pas possible de comparer nos résultats à ceux de la littérature par rapport au modèle de régression, puis nous mettrons en lumière les avantages et un seul inconvénient de considérer l'état du conducteur comme continu, d'alerte à extrêmement somnolent.

Concernant la somnolence au volant, dans la littérature, nous avons observé que des modèles de régression ont été utilisés dans des cas où la sortie à prédire était un temps de réaction (Larue, 2010; Lin et al., 2012; Liu et al., 2016; Wu et al., 2016). Néanmoins, nous ne pouvons pas comparer les métriques de performances, comme la racine carrée de l'erreur moyenne quadratique (REQM ou RMSE en anglais), utilisées dans ces articles car l'ordre de grandeur des sorties est différent : entre 0,150 et 2 secondes maximum pour un temps de réaction environ et entre 0 et 4 niveaux dans notre cas pour la détection et entre 0 et 60 minutes pour la prédiction.

L'avantage de considérer l'état du conducteur comme une variable continue, plutôt que binaire, nous permet d'obtenir une connaissance plus précise et plus détaillée de l'état réel du conducteur et d'agir ainsi en conséquence grâce à cette information enrichie. Afin de ne pas prendre le chemin du paradoxe du choix (Schwartz, 2009) (plus il y a de possibilité de choix, plus la décision d'un choix est difficile), nous nous sommes orientés vers un modèle continu qui va donc apprendre que les niveaux « alerte » et « légèrement somnolent » sont plus proches que les niveaux « alerte » et « extrêmement somnolent ». Ainsi, grâce à nos propositions de modélisation, nous pouvons savoir si le conducteur est proche de l'un des niveaux « alerte », « légèrement somnolent », « modérément », « très » ou « extrêmement somnolent » ou entre deux états consécutifs. Cette augmentation de précision par rapport à un état binaire alerte/somnolent présente divers avantages lors de son utilisation. Tout d'abord, cette granularité pourra permettre d'alerter le conducteur au fur et à mesure de sa conduite du changement de son état et également lui proposer, selon son état, certaines contre-mesures plutôt que d'autres. En effet, lorsque son état est critique, il est nécessaire de lui proposer une contre-mesure plus efficace, comme une pause avec la prise d'un café (Anund et al., 2008). Alors que lorsqu'il commence tout juste à s'écarter d'un état « alerte », essayer de le garder éveillé grâce à des contre-mesures dont le but est de maintenir le conducteur dans un état « alerte » pourraient être des options simples et intéressantes. Les contre-mesures seront plus détaillées dans le passage centré sur les perspectives.

De plus, le fait de connaître de manière plus précise son degré de somnolence pourrait plus facilement permettre de laisser le conducteur atteindre un niveau de somnolence déterminé en fonction du niveau d'automatisation de son véhicule. En effet, comme observée sur la figure 9-1, le conducteur doit surveiller la route au niveau 0 (conduite normale), au niveau 1 (conduite assistée) et

au niveau 2 (automatisation partielle). Dans ce cas, le niveau de somnolence « légèrement » ou « modérément somnolent » serait critique. Toutefois, au niveau 3 (automatisation conditionnelle), le conducteur n'a plus besoin de superviser lui-même la route. Même si , à ce stade, il n'a pas le droit de dormir car il doit être dans la capacité de reprendre la main à n'importe quel moment. Le niveau « extrêmement somnolent » pourrait être alors défini comme critique. Au niveau 4 (automatisation élevée), il a la possibilité de s'endormir en toute sérénité. La période de reprise de conduite pouvant être programmée à l'avance. Enfin, au niveau 5 (automatisation totale), le véhicule n'a plus besoin de conducteur pour circuler. On voit donc qu'on ne peut se contenter d'une connaissance binaire (oui/non) de l'état de somnolence du conducteur et que la détermination du niveau de somnolence de manière plus fine offre la possibilité de recommandations ou de contre-mesures plus adaptées aux circonstances de conduite.

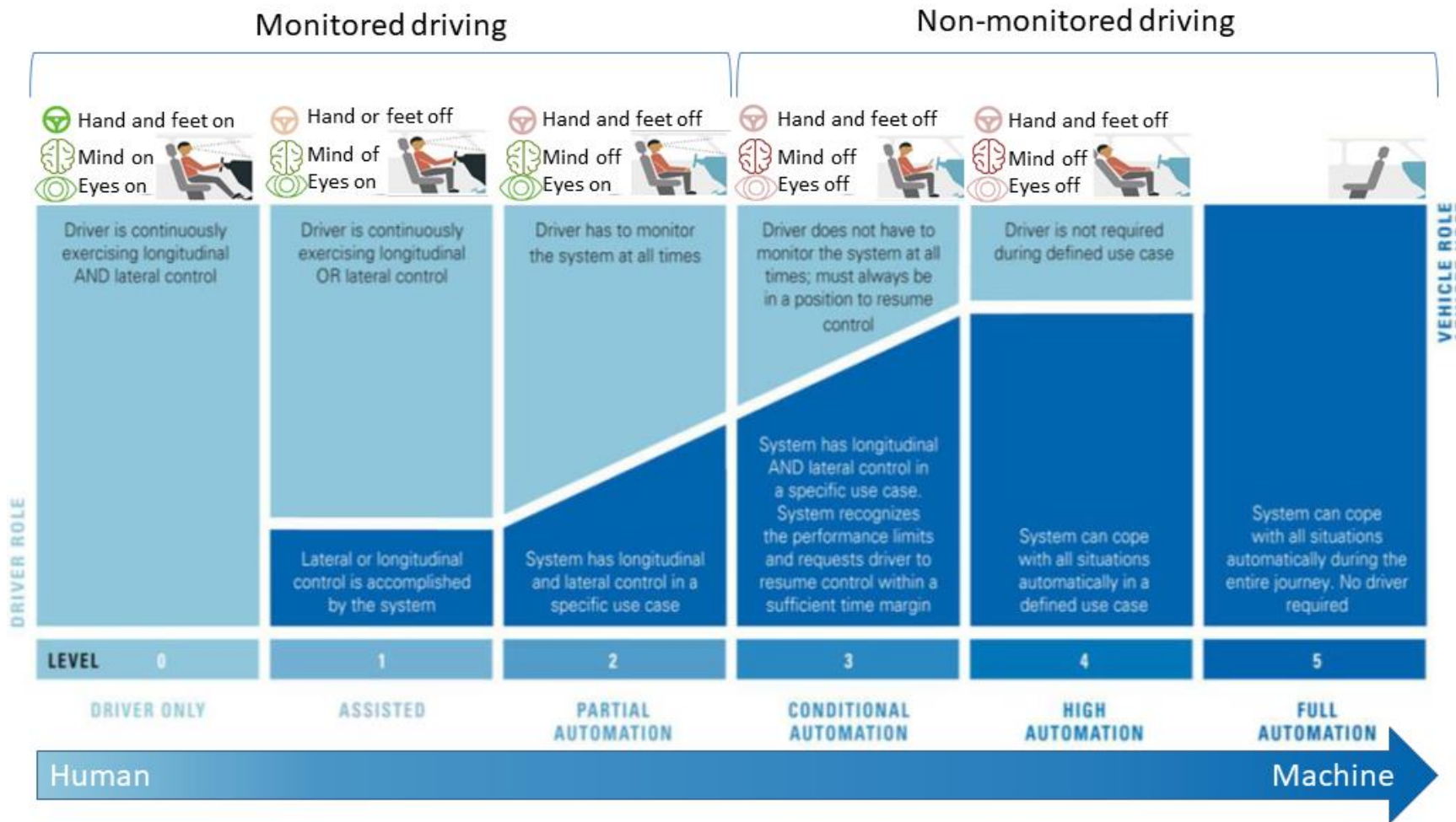


Figure 9-1 : Les différents niveaux d'automatisation

Cette augmentation de précision sur l'état du conducteur pourrait donc contribuer à la diminution des accidents causés par une baisse de vigilance. Cependant, le fait de détecter de fines variations de l'état du conducteur risque de diminuer la confiance du conducteur dans le système d'alerte. En effet, en lui annonçant qu'il est très légèrement somnolent ou qu'il sera très certainement somnolent dans un temps défini, mais que pour l'instant il est encore alerte, le conducteur risque de considérer le nombre de fausses alarmes plus important, même si cela n'est pas le cas. Ce problème des fausses alarmes est cependant déjà très présent dans les systèmes actuels de détection de la somnolence (Karrer-Gauß & Zawistowski, 2013). Dans ce contexte-là, le risque que le conducteur désactive son système de supervision reste important. Une proposition liée aux contre-mesures pour contrer ce problème de manque de confiance est d'ailleurs proposée dans un passage ultérieur centré sur les perspectives.

Nos travaux autour de l'état du conducteur ont permis de mettre en valeur que non seulement il était possible d'estimer avec une bonne précision le niveau de somnolence de celui-ci (modèle de détection), mais également qu'il était possible d'estimer le temps restant avant la survenue d'un niveau critique de somnolence (modèle de prédiction). En effet, peu d'études se sont intéressées à ces deux notions : au continuum de l'état du conducteur et au temps restant avant un niveau défini comme critique. Abordons maintenant les raisons pour lesquelles nous nous sommes intéressés à la prédiction.

9.3 Prédire la somnolence, mais pour quoi faire ?

Un enjeu majeur de ces travaux était de pouvoir anticiper une dégradation des performances du conducteur. En effet, savoir à un instant t que le conducteur est en train de s'endormir peut s'avérer trop tardif en terme de sécurité. Il vaut mieux anticiper. C'est pourquoi nous avons cherché à estimer le temps restant avant la survenue d'un niveau critique.

Ainsi, notre modélisation pouvant prédire le temps d'occurrence de la somnolence présente plusieurs avantages. Connaître le temps restant avant un certain niveau critique plutôt que d'estimer un état somnolent, 20 secondes avant (Murata et al., 2016), une minute avant (Kaida et al., 2007), 5 minutes avant (Damousis, Cester, Nikolaou, & Tzovaras, 2007) ou jusqu'à 10 minutes avant avec des performances amoindries (Larue, 2010; Larue, Rakotonirainy, & Pettitt, 2015), permet d'être plus précis sur le risque et le temps de dégradation de l'état du conducteur. Notre choix permet d'une part d'alerter le conducteur avant même l'arrivée de la situation critique et/ou de mettre en place une

contre-mesure adaptée selon le temps restant. Dans le cadre de cette étude, nous avons fixé l'état dégradé comme étant supérieur à un niveau de somnolence de 1,5 sur notre échelle. Il aurait sans doute été intéressant de faire varier cette valeur et d'étudier si un certain niveau peut être prédit avec une meilleure performance par exemple. Or, pendant la conduite, le niveau 1,5 est déjà critique, c'est pourquoi nous avons choisi celui-là prioritairement.

Ce modèle de prédiction permet d'anticiper le déclenchement d'une alerte afin d'annoncer au conducteur qu'il risque d'être somnolent, et ainsi lui proposer une pause à l'aire de repos la plus proche. Le fait d'anticiper qu'un état dégradé risque d'arriver pourrait permettre au véhicule de mettre en place des mécanismes nécessaires afin d'éviter l'accident en essayant de garder le conducteur dans la boucle de conduite et, si possible, contrer ainsi la dégradation des performances grâce à une contre-mesure. En effet, une conséquence négative du désengagement du conducteur dans la tâche de conduite risque d'impliquer une dégradation, voire une perte de la conscience de la situation (*situation awareness* : la connaissance de l'état du véhicule et/ou de l'environnement à un moment donné). On parle alors de conducteur « hors de la boucle de commande » (*out-of-the-loop state*) (Kaber & Endsley, 1997; Louw, Merat, & Jamson, 2015).

Nous venons de rappeler les avancées de nos travaux concernant les objectifs de la modélisation (détection et prédiction de la somnolence), intéressons-nous de plus près maintenant au caractère intrusif de la modélisation sur le conducteur.

9.4 Caractère intrusif de la modélisation sur le conducteur

Notre étude a été réalisée dans des conditions de laboratoire. Néanmoins, une contrainte importante que nous avons prise en compte réside dans le fait que la modélisation devait être peu intrusive pour pouvoir être utilisée facilement et de manière non intrusive dans un futur véhicule. A la vue de notre modélisation, trois points par rapport au caractère intrusif de la modélisation doivent être discutés : les signaux, la reconnaissance du conducteur et l'adaptation.

Tout d'abord, il était nécessaire que les signaux enregistrés dans le cadre de l'expérience du laboratoire puissent l'être également dans un véhicule. En effet, il était impératif que les indicateurs soient facilement exploitables et utilisables dans un véhicule. Nous pouvons noter que les signaux utilisés (mesures comportementales, mesures véhicules et mesures physiologiques) ne sont pas ou peu intrusifs et permettent de bonnes capacités de généralisation sans gêner et/ou perturber le conducteur. Par exemple, les mesures comportementales tels que les mouvements de paupières et de

tête sont observables grâce à une caméra. De nombreuses recherches s'intéressent à améliorer les performances de détection des clignements des yeux, en condition réelle, grâce une caméra comme cité dans la revue de littérature de Sahayadhas, Sundaraj, & Murugappan, (2012). Cependant, les caméras disposant forcément d'un angle de vision limité, si le conducteur ne regarde pas la route ou vers la/les caméras alors il ne sera pas possible de détecter ses yeux. Les mesures provenant du véhicule sont quant à elles disponibles via le CAN du véhicule et à l'aide d'une ou plusieurs caméras placées à l'extérieur du véhicule. Quant aux mesures physiologiques, en n'enregistrant pas l'EEG et en s'intéressant seulement au rythme cardiaque et à la respiration, nous limitons l'installation de capteurs car ces signaux peuvent être enregistrés grâce à une ceinture connectée, une montre connectée ou également une caméra (Bousefsaf, 2014).

Par ailleurs, nous avons utilisé plusieurs sources d'informations différentes : comportementales (mouvement de paupières et de tête), physiologiques (respiration et rythme cardiaque), activité et performance de conduite. Nous avons également testé la combinaison de toutes ces mesures. Par définition lors d'une conduite autonome, l'activité et les performances de conduite ne peuvent pas être mesurées, ou alors partiellement, si le véhicule est partiellement autonome. Or, dans un cas de conduite autonome sans supervision du conducteur, une étude a mis en évidence que les mesures comportementales utilisées pour analyser généralement la somnolence n'ont pas des patterns similaires pendant la conduite effective et pendant une conduite autonome (Schmidt, Braunagel, Stolzmann, & Karrer-Gauß, 2016; Schmidt, Laarousi, Stolzmann, & Karrer-Gauß, 2017). Ainsi, il n'était pas possible lors de la phase de conduite autonome de savoir si un PERCLOS élevé signifie que la personne s'ennuie, se relaxe ou est somnolente. Par conséquent, la fiabilité de ces mesures comportementales en conduite autonome risque donc d'être, dans une certaine mesure, compromise. Les mesures physiologiques qui peuvent ressortir comme plus contraignantes ont donc un grand intérêt dans ce cas.

De plus, nos travaux ont démontré qu'il était nécessaire d'avoir accès à un certain nombre d'informations personnelles sur le conducteur telles que l'âge, le genre, le score sur l'échelle de somnolence d'Epworth (Johns, 1991), le score sur le questionnaire de typologie circadienne de Horne et Östberg (Horne & Östberg, 1975), la qualité du sommeil, le nombre d'heures de sommeil de la nuit précédente, la fréquence de consommation de café, la fréquence de conduite, le nombre de kilomètres parcourus par an. Concernant les informations personnelles, le score sur l'échelle de somnolence d'Epworth (Johns, 1991), le score sur le questionnaire de typologie circadienne de Horne et Östberg (Horne & Östberg, 1975) et la consommation de café devront être demandés au conducteur la première fois qu'il conduit un véhicule possédant un système de supervision. Ces deux questionnaires sont rapides. Concernant les informations sur le sommeil comme la qualité du sommeil

et le nombre d'heures, il peut être envisagé que ces informations soient recueillies grâce à une montre connectée ou tout autre objet connecté apte à analyser le sommeil. Quant aux informations sur la conduite, celles-ci pourront être demandées une seule fois lors de l'acquisition du véhicule puis mises à jour au fur et à mesure de l'utilisation du véhicule. Ainsi, la plupart des informations sur lesquelles nous nous sommes appuyés pour développer nos modèles sont déjà, ou seront, facilement accessibles et ne nécessitent pas de modifications majeures du véhicule.

Cependant, pour l'adaptation et la récupération d'informations personnelles déjà existantes, il est également essentiel d'identifier le conducteur du véhicule à l'instant t . Cette reconnaissance de conducteur pourra être effectuée de manière manuelle par le conducteur en l'invitant à s'identifier, ou automatiquement. Il existe différentes façons de reconnaître un individu. Les informations provenant du CAN avec les accélérations et les freinages ont d'ailleurs permis d'identifier un conducteur unique (Enev, Takakuwa, Koscher, & Kohno, 2016). L'utilisation de mesures biométriques est également possible (Jain, Hong, & Pankanti, 2000) comme la reconnaissance de visage et/ou son thermogramme, l'empreinte digitale (« Vehicle security system including fingerprint and eyeball part identification », 1996), la géométrie de la main, l'iris (« Vehicle security system including fingerprint and eyeball part identification », 1996), la rétine, la voix (« Digital voice profiles », 2006) ou encore grâce à l'utilisation de l'ECG (Israel, Irvine, Cheng, Wiederhold, & Wiederhold, 2005; Singh & Gupta, 2008). La reconnaissance du conducteur pourrait donc être véritablement réalisée de manière non intrusive.

Toutefois, le caractère intrusif de la modélisation ne réside pas seulement dans l'acquisition de signaux mais également dans le type de modélisation choisie. En effet, pour le processus d'adaptation, il est nécessaire de connaître le véritable état du conducteur pendant la phase d'adaptation. Or, cet indicateur peut lui-même être intrusif et/ou contraignant. En effet, l'analyse vidéo est contraignante car elle réclame la connaissance d'experts *a posteriori*. L'utilisation de l'EEG est une solution moins contraignante car elle n'implique pas de tiers, mais elle nécessite également du matériel spécifique. De plus, il aurait été possible d'utiliser les écarts de conduite par rapport à la route, néanmoins, cette mesure est loin d'être spécifique à la somnolence au volant, d'autres causes, y compris externes, pouvant causer des écarts de conduite. Afin d'avoir des modèles performants même sur des nouveaux participants, nous avons identifié qu'il existait deux méthodologies, soit l'apprentissage d'un modèle par participant, soit l'adaptation ou transfert de connaissances. Cette dernière option est celle que nous avons choisie. L'avantage de cette proposition par rapport à la première est que l'adaptation nécessite moins de données à étiqueter que l'apprentissage d'un modèle propre à chacun.

En plus d'être non intrusifs, il était important que les modèles soient performants et permettent d'être généralisé sur tout nouveau conducteur. Afin de rendre les modèles performants, nous avons mis en valeur que l'ajout d'informations supplémentaires et la transférabilité des connaissances permettaient d'accroître les performances de manière importante.

9.5 L'ajout d'informations supplémentaires

Le chapitre 6 a mis en évidence l'intérêt d'ajouter des informations supplémentaires telles que des informations personnelles sur le conducteur et le temps de conduite (le temps écoulé depuis le début de l'expérience). Dans la suite des résultats (chapitre 7 et 8), ces indicateurs ont été toujours présents dans les données d'entrée des modèles. Au début de ces travaux nous avons voulu ajouter également le type de route et le trafic : deux éléments contextuels qui ont également varié au cours de l'expérience. Cependant, la conduite a été trop brève dans ces différentes zones pour en tirer des conclusions définitives. De plus, une variation du niveau de somnolence observé a été similaire pour tous les participants dans ces différentes zones, c'est-à-dire que nous avons constaté pour tous une diminution du niveau de somnolence avec la présence d'un trafic important et un retour au niveau d'alerte lors de la conduite en ville. En effet, le trafic ou la ville contrastaient avec la conduite par ailleurs très monotone sur autoroute, ce qui engendrait un changement du niveau de vigilance des conducteurs. Nous n'avons pas ajouté ces informations aux modèles car nous aurions risqué un sur-apprentissage. Le modèle aurait toujours assimilé la conduite en ville à un état « alerte », alors qu'il est possible de rencontrer des épisodes de somnolence même en ville (Philip et al., 2010).

Nous avons démontré au chapitre 4 que les informations personnelles que nous avons recueillies sur les participants ne permettaient pas d'expliquer véritablement la dynamique de la somnolence. Lorsque ces informations personnelles sont couplées aux sources de données physiologiques, comportementales, à l'activité et aux performances de conduite (avec un ANN, chapitre 6) alors cette fois-ci, leur importance est révélée de manière évidente, se traduisant par une amélioration des performances. Ce constat peut s'expliquer par le fait que la combinaison de ces informations grâce aux ANN peut cette fois être utile pour donner du poids au résultat.

Quant à la variable temps de conduite, comme vu et analysé dans le chapitre 6, nous avons mis en évidence que cette variable améliore grandement les différentes performances de généralisation. Evidemment, plus le temps de conduite est long sous des conditions monotones, plus la probabilité d'être somnolent augmente (Philip, Taillard, Guilleminault, et al., 1999). C'est pourquoi il est conseillé de faire une pause toutes les deux heures de conduite (Philip, Taillard, Quera-Salva, Bioulac, &

Åkerstedt, 1999). Dans le scénario de conduite mis en place, un élément était présent dans le but de faire varier la dynamique de la somnolence malgré l'augmentation du temps de conduite : le trafic au 2/3 de l'autoroute. Cette perturbation n'arrivait pas forcément en même temps pour tous les conducteurs car étant contrôlée par les conducteurs, la vitesse était variable.

9.6 Transférabilité des connaissances

Une particularité de ces travaux réside, d'une part, dans la comparaison des performances selon le jeu de données utilisé et, d'autre part, dans la division des jeux de données afin de connaître les performances de généralisation des modèles dans différentes situations. Ces deux éléments sont également liés et dépendants de la finalité des modèles. Lorsque les données sont réparties aléatoirement entre les différents jeux de données (entraînement, validation et test), les meilleures performances sont obtenues à l'aide des données comportementales aussi bien pour la détection que pour la prédiction. Lors de l'adaptation, les meilleures performances sont obtenues à l'aide du jeu de données comportementales pour la prédiction et à l'aide de la combinaison des différentes sources d'information pour la détection. Nos travaux ont permis de mettre en avant la notion de transférabilité des connaissances et la généralisation des modèles. En effet, en apprentissage automatique, savoir si les connaissances sur un domaine A seront utilisables sur un domaine B, similaire au domaine A mais différent, est un important challenge ; par exemple, savoir si les connaissances apprises pour reconnaître un chat sont utilisables pour reconnaître un chien. Ainsi, nos travaux ont mis en valeur d'une part la possibilité de transférer à un nouveau conducteur des connaissances apprises pour détecter et prédire la somnolence au volant, apprises sur un ensemble de conducteurs, grâce à une faible quantité de données d'enregistrement seulement. Cette notion de transfert avait déjà été utilisée dans la littérature concernant les interfaces-cerveau-machine (Wang et al., 2015) et la détection de la somnolence grâce à de l'EEG (Wei et al., 2015; Wu et al., 2016; Dongrui Wu et al., 2015, 2015; Dongrui Wu, Lawhern, Gordon, Lance, & Lin, 2016). Toutefois à notre connaissance, la modélisation de la somnolence grâce à du transfert d'apprentissage n'avait jamais été réalisée en utilisant des indicateurs indirects moins intrusifs que l'EEG.

Cette transférabilité des connaissances sur la modélisation de la somnolence au volant permet de mettre en valeur aussi bien la variabilité que l'immuabilité inter-individuelle. En effet, l'idiosyncrasie montre le fait que les individus ne sont pas tous identiquement sensibles à la somnolence au volant. Sans celle-ci, il n'y aurait pas besoin d'adaptation et de transfert des connaissances. Cependant, afin que le transfert de connaissances soit bénéfique, seulement très peu de données suffisent pour

augmenter les performances. Cette faible quantité de données nécessaires montre une immuabilité inter-individuelle indispensable pour que les modèles apprennent et transfèrent leurs connaissances sur un nouvel individu.

Afin d'étudier la quantité de données nécessaires à l'adaptation, l'adaptation personnalisée a été réalisée grâce aux premières données d'enregistrement, contrairement à une autre étude qui a sélectionné 5 données supplémentaires, choisies aléatoirement, au fur et à mesure de l'adaptation (Dongrui Wu et al., 2015). L'avantage de leurs propositions repose sur le fait que l'ajout de données, non proches temporellement, permet d'ajouter certainement plus de variabilité. En effet, il est fort probable que des données proches temporellement soient également proches en valeur et correspondent à un niveau de somnolence voisin, n'apportant pas forcément des informations supplémentaires pour personnaliser le modèle. Néanmoins, l'avantage de notre méthodologie est qu'elle nous permet de ne pas manquer des événements de somnolence (faux négatifs), survenus avant les données choisies aléatoirement lors de l'adaptation.

Par ailleurs, supposons que le modèle personnalisé à chaque conducteur s'adapte au fur et à mesure dans le temps. Le modèle sera donc de plus en plus personnalisé et s'éloignera de plus en plus du modèle général. Nous pourrions nous demander alors si cela est nécessairement bénéfique ou non, faut-il ajouter de l'oubli au modèle ? En effet, en devenant de plus en plus spécifique, ses capacités de généralisation sur une nouvelle situation seront peut-être dégradées.

Finalement, dans le cadre de ces travaux, nous avons utilisé une modélisation employant des ANN, et nous avons justifié ce choix. D'autres méthodes de transfert d'apprentissage et d'adaptation de domaine pourraient être également envisagées. En effet, la problématique de la détection et de la prédiction du niveau de somnolence offre un grand champ de possibles pour innover dans ce domaine.

9.7 Les limites de l'ensemble de ces travaux

En vue de l'analyse de ces travaux, nous avons soulevé différentes limites concernant tant la méthodologie de recueil de données que la modélisation. Les quatre principales limites vont donc être discutées : le nombre de participants, la vérité de terrain, les conditions de l'expérience, le type de modélisation.

9.7.1 Le nombre de participants

Tout d'abord, une première limite que nous pouvons identifier est liée au faible effectif de conducteurs testés dans cette étude, et donc au nombre limité des données acquises. En effet, 21 participants conduisant entre 105 et 120 minutes, âgés en moyenne de 24 ans, restent un nombre faible. De plus, dans le domaine de l'apprentissage automatique, il est souvent dit que plus il y a de données, mieux c'est. En effet, l'apprentissage profond a eu ces dernières années un grand intérêt du fait de l'augmentation de la quantité de données accessibles et de la puissance de calcul disponible. La faible quantité de données utilisée dans notre cas pour modéliser la somnolence peut également expliquer pourquoi la généralisation des modèles sur un nouveau conducteur n'est pas satisfaisante. En effet, même si les ANN modélisant sont dans notre cas relativement simples, il se peut tout de même qu'ils soient trop complexes à la vue de la quantité de données disponibles. Malgré ce nombre relativement faible de participants, nous avons obtenu des performances de généralisation élevées, notamment grâce à un jeu de test aléatoire (chapitre 6), mais également grâce à l'adaptation des modèles (chapitre 8).

Par ailleurs, le nombre de participants qui permet de construire l'ANN général lors du processus d'adaptation a toujours été fixe. Il aurait été intéressant toutefois de faire varier cette quantité afin de savoir si un nombre plus faible de conducteurs et surtout quels conducteurs sont pertinents dans les données du jeu d'entraînement du ANN général pour améliorer le transfert de connaissances, comme cela a été étudié par exemple sur la détection de la somnolence avec de l'EEG (Wei et al., 2015).

9.7.2 La vérité de terrain

Le choix de notre « vérité de terrain » (c'est-à-dire le véritable état du conducteur, ce que le modèle devra apprendre) a été déterminé au moyen d'une analyse subjective effectuée toutes les minutes par deux personnes tierces en se basant sur des vidéos de conducteurs. La fiabilité de cette mesure a déjà été démontrée dans la littérature. L'étiquetage des données a d'ailleurs été décrit comme plus riche que celles utilisant de l'EEG ou le PERCLOS (Millemann et al., 2001). En effet, il a déjà été démontré que la survenue et le nombre de micro-sommeils par exemple, objectivés par enregistrements de l'EEG ne correspondaient pas nécessairement à des sorties de route, et inversement. Par ailleurs, bien que cette mesure nécessite un post-traitement coûteux, nous nous sommes orientés vers cette mesure pour la facilité de la mise en place. Un autre aspect lié à l'observation des personnes tierces est que cette dernière s'est orientée de manière indirecte vers le comportement de la personne. Ce fait pourrait donc expliquer que nous obtenons de meilleurs résultats avec le jeu de données comportementales dans le cas de la détection et prédiction de la somnolence au volant avec

répartition aléatoire des données. Cette différence est présente, mais faible, lorsque nous comparons aux autres sources d'information (physiologiques et provenant du véhicule).

9.7.3 Les conditions de l'expérience

Les données recueillies au cours de l'expérimentation et qui ont servies comme données d'entrées et de test des différents modèles ont été enregistrées sur un simulateur de conduite, en début d'après-midi, et sur une population jeune. Il a été démontré que les modèles développés sur des simulateurs n'étaient pas transférables directement en situation réelle (Philip et al., 2005). Cependant, une étude récente a mis en exergue que la dynamique concernant la variation des variables liées à la somnolence telles que les clignements de paupières, la vitesse du véhicule, le score sur l'échelle KSS, sont cohérentes et similaires. En comparant ces variations sur un simulateur de conduite et en situation réelle, il en ressort que les différences relatives sont plus importantes (Fors et al., 2018). Ainsi, étudier la somnolence au volant sur un simulateur de conduite est plus simple et plus sûr, mais cela n'est pas si différent de la conduite en condition réelle. Une contrainte à prendre en compte néanmoins afin d'envisager un transfert en situation réelle, est relative à l'ensemble des caractéristiques dynamiques de la situation de conduite (vibrations, changements de luminosité, mouvements du conducteur...) qui peuvent nuire à la qualité des enregistrements disponibles. De plus, l'adaptation de domaine et l'apprentissage par transfert pourraient servir non seulement pour améliorer les performances sur un nouveau conducteur mais également à une conduite en condition réelle, à une autre population, à un autre créneau horaire etc. Ce transfert de connaissances pourrait donc permettre une diminution du temps d'acquisition et de labellisation de la base de données qui sert à l'entraînement d'un modèle. En effet, comme les dynamiques sont similaires, l'utilisation de transfert d'apprentissage d'un simulateur de conduite à une situation réelle semble tout à fait envisageable.

9.7.4 Le type de modélisation choisie

Afin de détecter et de prédire la somnolence au volant, nous nous sommes orientés vers des réseaux de neurones artificiels, des modèles qui ont révélés leurs pertinences et leurs bons résultats (Beale et al., 1992; Mahesh M. Bundele & Banerjee, 2009; Kohonen, Barna, & Chrisley, 1988; Tjolleng et al., 2017; Wilson & Russell, 1999; Zhao, Wang, Zhang, Qi, & Wang, 2017). Cependant, il aurait été intéressant, comme cela est beaucoup fait dans le domaine de l'apprentissage automatique, de comparer différents algorithmes entre eux et également de tester différentes techniques d'apprentissage par transfert. De plus, nos modèles sont statiques, c'est-à-dire qu'ils ne prennent pas en compte l'historique des données de la personne. Il pourrait donc être intéressant de tester par la

suite des modèles dynamiques comme l'ont fait Yang et al. (2010). Par ailleurs, d'autres types de méthodes de transfert d'apprentissage seraient également intéressantes à tester (comme l'ont proposé Torrey & Shavlik, 2009) et à comparer afin de déterminer la méthode la plus efficace.

Malgré ces différentes limites, nos travaux ont mis en évidence de bonnes performances pour la détection et la prédiction de la somnolence au volant. Ces travaux amènent également diverses perspectives pour des recherches futures.

9.8 Perspectives et futures recherches

Nos travaux amènent une proposition de réponse à la variabilité inter-individuelle dans le cas de la conduite. Ainsi, nous proposerons deux perspectives à ces travaux concernant la variabilité intra-individuelle et la prise en compte du niveau d'automatisation du véhicule. De plus, dans le cas de notre étude, nous n'avons pas alerté le conducteur à propos de la survenue de la somnolence lors de la conduite. Nous détaillerons alors en quoi il serait intéressant d'intégrer des contre-mesures et de tester également l'évolution du modèle de prédiction

9.8.1 Intégrer les contre-mesures dans les modèles et utilisation en temps réel

Notre étude s'est focalisée sur la création de modèles performants dans l'objectif d'estimer l'état de la personne de manière précise. Cependant, nous n'avons pas effectué d'essai en condition réelle et nous ne nous sommes pas intéressés volontairement à l'utilisation de cette information sur l'état du conducteur. Ainsi, pouvoir tester les conducteurs en condition réelle avec un environnement routier riche tout en analysant leur niveau de somnolence, ainsi que déterminer le temps restant avant l'atteinte d'un niveau critique est une perspective fortement intéressante. Cette information sur l'état du conducteur peut en effet servir au développement de contre-mesures afin de lutter contre la somnolence au volant. Par exemple, un simulateur de conduite « augmenté » a été développé en détectant l'hypovigilance à partir des clignements, bâillements et clignement de la tête. Lorsqu'un état critique était détecté, le conducteur était alerté grâce à un système audio, visuel et des vibrations sur le volant (Benoit et al., 2009). Le développement et le choix d'une contre-mesure imposent en effet de bien maîtriser la problématique de la somnolence au volant. Il existe d'ailleurs une large variété de contre-mesures qui ont été mises en valeur par la littérature, comme (c'est trivial) une sieste,

l'exposition à une lumière brillante ou bleue, la prise de caféine, ouvrir la fenêtre (Anund, Fors, Kecklund, Leeuwen, & Åkerstedt, 2015). Cependant, certaines contre-mesures populaires comme mettre la radio ou ouvrir la fenêtre n'ont pas révélé d'effets significatifs sur la modification de l'état de somnolence (Reyner & Home, 1998; Schwarz et al., 2012). La contre-mesure la plus prometteuse semble être de s'arrêter pour prendre une pause et de faire la sieste (Anund et al., 2008) ou bien encore une combinaison d'une exposition à une lumière vive à dominante bleue, assortie d'une sieste, permet de réduire de manière significative la somnolence (Leger, Philip, Jarriault, Metlaine, & Choudat, 2009). Nous pourrions également citer les campagnes de prévention menées par la Fondation VINCI Autoroutes, qui préconisent de s'arrêter, prendre un café, se reposer pendant une vingtaine de minutes et effectuer quelques mouvements éveillant de remobilisation avant de reprendre la route. Comme mise en valeur précédemment, l'augmentation de précision sur le niveau de somnolence pourra permettre d'utiliser une contre-mesure adaptée à son niveau de somnolence et d'augmenter le nombre de modalités sensorielles permettant d'alerter le conducteur au fur et à mesure que son niveau de somnolence augmente. Cela pourrait se traduire, lorsque la personne devient très somnolente, par lui proposer une pause avec une sieste et un café grâce aussi bien par un système sonore, visuel qu'haptique.

Lorsqu'un système alerte le conducteur ou que le conducteur prend une pause ou boit un café, sa dynamique de somnolence peut être modifiée et, par là même, le temps restant avant un certain niveau critique évolue. La réponse du modèle de prédiction doit donc savoir évoluer, s'ajuster. Il semble alors nécessaire de tester cette potentialité et de, peut-être, réapprendre un modèle par le moyen de l'utilisation de contre-mesures.

Dans des circonstances où le conducteur est alerté de sa baisse de vigilance, il peut toutefois choisir de continuer à conduire. Une étude a ainsi révélé que 41% des conducteurs ayant été alertés par un système de vigilance d'un évènement de somnolence, et la trouvant pertinente, ont toutefois décidé de poursuivre la conduite (Karrer-Gauß & Zawistowski, 2013). Il peut donc se révéler nécessaire, soit de contrôler la conduite afin qu'elle soit fiable en cas de défaillance du conducteur, soit de faire en sorte que le conducteur reste vigilant. Pour le premier cas, mettre en place des aides à la conduite comme un régulateur de vitesse adaptatif au trafic par exemple pourra être particulièrement pertinent (Zhang et al., 2016). Néanmoins afin de ne pas être contre-productif, ce système pourrait intervenir uniquement en cas d'erreur de conduite importante. Pour le deuxième cas, face au risque que le conducteur continue de conduire alors qu'il devient somnolent, des aides pour le maintenir dans un état « alerte » peuvent se révéler nécessaires comme l'utilisation d'huiles essentielles de cannelle ou de menthe poivrée (Raudenbush, Grayhem, Sears, & Wilson, 2009), l'implication dans des tâches actives telles qu'un jeu à base de QCM (Oron-Gilad, Ronen, & Shinar, 2008), l'activation d'une lumière

bleue (Taillard et al., 2012). Par ailleurs, jouer sur la motivation du conducteur à rester en état d'alerte pourrait également être intéressant car, par exemple il a été montré à partir d'une autre tâche que la conduite que les performances de participants n'étaient pas dégradées avec la durée de la tâche lorsqu'on les informait que leurs performances étaient évaluées (Bonfond, Doignon-Camus, Hoeft, & Dufour, 2011). Ces différentes propositions ont pour objectif de conserver le conducteur dans un état de vigilance sécuritaire pour la conduite.

9.8.2 Prendre en compte la variabilité intra-individuelle

L'adaptation des modèles permet d'offrir une première réponse au problème de la variabilité inter-individuelle. La variabilité intra-individuelle n'est, quant à elle, pas directement abordée dans le cadre de cette thèse. En effet, une récente étude a mis en exergue une variabilité intra-individuelle en faisant conduire des conducteurs à quelques jours d'intervalle dans des conditions similaires (Nilsson et al., 2017). Il peut donc être difficile de savoir, sans de nouvelles expériences, à quelle fréquence l'adaptation, comme une forme de mise à jour afin de contrecarrer la variabilité intra-individuelle, est nécessaire : quotidiennement, hebdomadairement, annuellement... Pour résoudre, partiellement ce problème, le même participant, ayant conduit une fois dans nos études, devrait au moins conduire une nouvelle fois, un jour différent, comme proposé par (J. Zhang et al., 2017). Cela impliquerait cependant une étude plus approfondie sur la variabilité intra-individuelle et l'évaluation de la fréquence optimale de renouvellement de l'adaptation.

9.8.3 Prendre en compte le niveau d'automatisation du véhicule

Ces travaux de thèse ont étudié principalement le conducteur lors de la conduite manuelle en simulateur, sans aucune assistance. Or, actuellement, la technologie propose des systèmes avancés pouvant apporter une assistance aux conducteurs en présentant des informations supplémentaires comme le GPS ou en automatisant des processus (régulation longitudinale et/ou latérale). En effet, la supervision du conducteur diminue à mesure que le niveau d'automatisation augmente, mais le principal problème vient du fait que le niveau de sécurité dépend de la combinaison entre le système d'automatisation et l'homme (Merat & Lee, 2012). Néanmoins, ces systèmes d'aide à la conduite (*Advanced Driving Assistance Systems*, ou ADAS) peuvent avoir des conséquences négatives sur les performances de conduite en étant source de distraction ou en provoquant une baisse de vigilance (Inagaki, 2008; Li et al., 2012; Nikolova, Radev, & Vukov, 2003; Young et al., 2008) avec une perte de la conscience de la situation du conducteur lorsque celui est « hors de la boucle » (Kaber & Endsley, 1997; Louw et al., 2015). Or, la supervision des systèmes automatisés en général peut être

accompagnée d'une perte de vigilance et de la difficulté à anticiper des événements extérieurs (Körber, Cingel, Zimmermann, & Bengler, 2015). En effet, une conduite pendant le niveau 1 d'automatisation augmente le risque de somnolence et les conducteurs mettent plus de temps à réagir à des événements soudains (Dufour, 2014). Afin d'éviter d'être affecté par les conséquences négatives liées à l'automatisation, ces systèmes d'aide à la conduite doivent plutôt être en mesure de coopérer avec le conducteur (Hoc, 2007; Inagaki, 2008; Li et al., 2012) en s'adaptant, par exemple à l'état de celui-ci à l'instant t pour lui apporter une aide utile dans un contexte précis (Hong, Suh, & Kim, 2009; Navarro, 2010). Ainsi, avec l'arrivée des véhicules autonomes, un besoin de partage d'autorité entre l'homme et le véhicule apparaît désormais incontournable. En effet, avec des degrés d'autonomie importants, des nouvelles problématiques surgissent. Sans stimuli, le conducteur n'agit plus sur son véhicule, sa distraction et sa somnolence risquent alors de croître dangereusement. En devenant passif, pourrait-il toutefois continuer à rester vigilant sur la route ? Que se passerait-il également s'il devait reprendre la main lors d'une interruption brutale du système de conduite autonome ?

Par ailleurs, concernant la prédiction de la somnolence au volant, nous avons choisi de fixer le niveau critique à 1,5. Mais si le conducteur n'agit pas sur le véhicule, ce seuil pourrait évoluer selon la sûreté du niveau d'automatisation et le temps possible avant que le conducteur reprenne la main. Si une supervision est nécessaire (niveau 2 ou partiellement 3 par exemple), si la personne n'a plus besoin de superviser la route, il peut alors être envisagé que la personne devienne somnolente ou dorme (pour le niveau 4 ou 5 d'automatisation). Toutefois, nous pouvons supposer que le véhicule autonome pourrait également anticiper la sortie des zones d'autonomie pour ne plus autoriser dans ce cas un niveau de somnolence trop élevé, ceci un certain temps avant la reprise en main. Il peut être envisageable, si le conducteur éprouve des difficultés à rester éveillé dans le cas où la conduite autonome sans supervision est possible, qu'elle s'active automatiquement et propose au conducteur de se reposer afin de reprendre la main plus tard et d'être dans les meilleures conditions de conduite. Avant la reprise en main, le système de supervision du conducteur devra alors s'assurer que celui-ci se trouve dans un état optimal pour la reprise de la conduite.

Un monde empli de véhicules 100 % autonomes se profile à un horizon toutefois encore lointain (ou tout du moins qui semble reculer au fur et à mesure que la technologie avance, tant ces avancées révèlent en fait de nouveaux problèmes à résoudre). Il s'agit maintenant de poursuivre des recherches dans tous les méandres et ramifications des domaines nécessaires à la synthèse que représente un acte de conduite par un être humain ou une machine, ou la transition entre les deux. La lutte contre

la somnolence est un enjeu actuel dans un monde où le poids du quotidien risque de nous amener à somnoler par le biais de cet enfermement, apparemment passif, dans un véhicule de plus en plus confortable (à la limite du *cocooning*) et qui réclame pourtant de l'attention et de multiples ressources cognitives. Qu'en sera-t-il de ce nouveau monde quand la confiance en l'automatisation sera définitivement instaurée et l'être humain définitivement libéré de la conduite ? Profitera-t-il de ce temps pour récupérer la fatigue accumulée et dormir sereinement ou, au contraire, déploiera-t-il une nouvelle créativité ou de nouvelles activités, étroitement liées à ce nouvel espace offrant du temps de liberté et de nouvelles potentialités à son imaginaire ? L'avenir, peut-être plus proche qu'on ne l'imagine, nous le dira.

Références

- Ahlström, C., Anund, A., Fors, C., & Åkerstedt, T. (2018). Effects of the road environment on the development of driver sleepiness in young male drivers. *Accident; Analysis and Prevention*, 112, 127-134.
- Alhazmi, S. (2013). *Towards Context-based Fatigue Detection System in Vehicular Area Network*.
- Amato, J.-N. (2011). Effet de la prise de médicaments psychotropes et de l'âge du conducteur sur la conduite, l'éveil et les fonctions attentionnelles. I. Effet dose d'un analgésique, II. Effet combiné d'une dose d'analgésique et d'un hypnotique. (Doctoral dissertation, Université de Caen).
- Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52(1-2), 29-37.
- Anund, A., Ahlström, C., Fors, C., & Åkerstedt, T. (2018). Are professional drivers less sleepy than non-professional drivers? *Scandinavian Journal of Work, Environment & Health*, 44(1), 88-95.
- Anund, A., Fors, C., Kecklund, G., Leeuwen, W. van, & Åkerstedt, T. (2015). *Countermeasures for fatigue in transportation: a review of existing methods for drivers on road, rail, sea and in aviation*.
- Anund, A., Kecklund, G., Peters, B., & Åkerstedt, T. (2008). Driver sleepiness and individual differences in preferences for countermeasures. *Journal of Sleep Research*, 17(1), 16-22.
- Apparies, R. J., Riniolo, T. C., & Porges, S. W. (1998). A psychophysiological investigation of the effects of driving longer-combination vehicles. *Ergonomics*, 41(5), 581-592.
- Appelhans, B. M., & Luecken, L. J. (2006). Heart rate variability as an index of regulated emotional responding. *Review of general psychology*, 10(3), 229.
- Arnedt, J. T., Wilde, G. J., Munt, P. W., & MacLean, A. W. (2001). How do prolonged wakefulness and alcohol compare in the decrements they produce on a simulated driving task? *Accident Analysis & Prevention*, 33(3), 337-344.
- Awais, M., Badruddin, N., & Drieberg, M. (2014). A non-invasive approach to detect drowsiness in a monotonous driving environment. In *TENCON 2014 - 2014 IEEE Region 10 Conference* (p. 1-4).
- Barr, L. C., Yang, C. Y., Hanowski, R. J., & Olson, R. (2011). *An assessment of driver drowsiness, distraction, and performance in a naturalistic setting*.
- Barron, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information theory*, 39(3), 930-945.
- Beale, M., Hagan, M. T., & Demuth, H. B. (1992). Neural network toolbox. *Neural Network Toolbox, The Math Works*, 5, 25.
- Belz, S. M., Robinson, G. S., & Casali, J. G. (2001). *An On-Road Investigation of Commercial Motor Vehicle Operator Self Assessment of Fatigue as an Indicator of Driver Fatigue* (p. 1576-1580).
- Benoit, A., Bonnaud, L., Caplier, A., Ngo, P., Lawson, L., Trevisan, D. G., Levacic, V., Manca, C. & Chanel, G. (2009). Multimodal focus attention and stress detection and feedback in an augmented driver simulator. *Personal and Ubiquitous Computing*, 13(1), 33-41.
- Benoit, A., & Caplier, A. (2005). Hypovigilance analysis: open or closed eye or mouth? Blinking or yawning frequency? In *IEEE Conference on Advanced Video and Signal Based Surveillance, 2005. AVSS 2005* (pp. 207-212).
- Bergasa, L. M., Nuevo, J., Sotelo, M. A., Barea, R., & Lopez, M. E. (2006). Real-time system for monitoring driver vigilance. *IEEE Transactions on Intelligent Transportation Systems*, 7(1), 63-77.
- Besson, P. (2007). A multimodal pattern recognition framework for speaker detection.
- Besson, P., Bourdin, C., Bringoux, L., Dousset, E., Maiano, C., Marqueste, T., ... Vercher, J.-L. (2013). Effectiveness of Physiological and Psychological Features to Estimate Helicopter Pilots'

- Workload: A Bayesian Network Approach. *IEEE Transactions on Intelligent Transportation Systems*, 14(4), 1872-1881. <https://doi.org/10.1109/TITS.2013.2269679>
- Bhowmick, B., & Chidanand Kumar, K. S. (2009). Detection and classification of eye state in IR camera for driver drowsiness identification. In *2009 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* (pp. 340–345).
- Boucsein, W., Haarmann, A., & Schaefer, F. (2007). Combining Skin Conductance and Heart Rate Variability for Adaptive Automation During Simulated IFR Flight. In D. Harris (Éd.), *Engineering Psychology and Cognitive Ergonomics* (p. 639-647). Springer Berlin Heidelberg.
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, 44, 58–75.
- Bousefsaf, F. (2014). *Mesure sans contact de l'activité cardiaque par analyse du flux vidéo issu d'une caméra numérique : extraction de paramètres physiologiques et application à l'estimation du stress*. (Doctoral dissertation Université de Lorraine).
- Brandt, T., Stemmer, R., & Rakotonirainy, A. (2004, October). Affordable visual driver monitoring system for fatigue and monotony. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on* (Vol. 7, pp. 6451-6456). IEEE.
- Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3), 199–231.
- Briest, S., Karrer-Gauß, K., & Schleicher, R. (2006). Driving without awareness: Examination of the phenomenon. *Vision in Vehicles XI*, 89–141.
- Brown, I. D. (1991). Highway hypnosis: implications for road traffic researchers and practitioners. In: Gale, A.D., et al. (Eds.), *Vision in Vehicle III. Elsevier, North-Holland*, 459–465.
- Brown, I. D. (1994). Driver fatigue. *Human factors*, 36(2), 298-314.
- Brown, I. D. (1997). Prospects for technological countermeasures against driver fatigue. *Accident Analysis & Prevention*, 29(4), 525-531.
- Bundele, Mahesh M., & Banerjee, R. (2009). Detection of Fatigue of Vehicular Driver Using Skin Conductance and Oximetry Pulse: A Neural Network Approach. In *Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services* (p. 739–744). New York, NY, USA: ACM.
- Bundele, M.M., & Banerjee, R. (2009). An SVM Classifier for Fatigue-Detection Using Skin Conductance for Use in the BITS-Lifeguard Wearable Computing System. In *2009 2nd International Conference on Emerging Trends in Engineering and Technology (ICETET)* (p. 934-939).
- Caffier, P. P., Erdmann, U., & Ullsperger, P. (2003). Experimental evaluation of eye-blink parameters as a drowsiness measure. *European Journal of Applied Physiology*, 89(3-4), 319-325.
- Calcagnini, G., Biancalana, G., Giubilei, F., Strano, S., & Cerutti, S. (1994). Spectral analysis of heart rate variability signal during sleep stages. In *Engineering in Medicine and Biology Society, 1994. Engineering Advances: New Opportunities for Biomedical Engineers. Proceedings of the 16th Annual International Conference of the IEEE* (Vol. 2, p. 1252–1253). IEEE.
- Caldwell, J. A., Hall, K. K., & Erickson, B. S. (2002). EEG data collected from helicopter pilots in flight are sufficiently sensitive to detect increased fatigue from sleep deprivation. *The International Journal of Aviation Psychology*, 12(1), 19–32.
- Carskadon, M. A., & Dement, W. C. (1982). Nocturnal determinants of daytime sleepiness. *Sleep: Journal of Sleep Research & Sleep Medicine*.
- Chauhan, A., Saroliya, A., & Sharma, V. (2015). Design & Analysis of KNN algorithm for fatigue detection in vehicular drivers using Pulse Oximetry parameter. *International Journal of Engineering Technology and Management*, 2(3), 107-110.
- Chen, J., & Ji, Q. (2012). Drowsy Driver Posture, Facial, and Eye Monitoring Methods. In A. Eskandarian (Éd.), *Handbook of Intelligent Vehicles* (p. 913-940). Springer London.
- Chen, R. (2013, avril 1). *Sitting behaviour-based pattern recognition for predicting driver fatigue*. Deakin University.

- Christoff, K. (2012). Undirected thought: neural determinants and correlates. *Brain research*, 1428, 51–59.
- Chua, C.-P., McDarby, G., & Heneghan, C. (2008). Combined electrocardiogram and photoplethysmogram measurements as an indicator of objective sleepiness. *Physiological measurement*, 29(8), 857.
- Clarion, A. (2009, décembre 3). *Recherche d'indicateurs électrodermiques pour l'analyse de la charge mentale en conduite automobile* (Doctoral dissertation, Université Claude Bernard - Lyon I).
- Clarion, A., Ramon, C., Petit, C., Dittmar, A., Bourgeay, J. P., Guillot, A., ... Collet, C. (2009). An integrated device to evaluate a driver's functional state. *Behavior research methods*, 41(3), 882–888.
- Cohen, R. A. (2011). Yerkes–Dodson Law, 2737-2738.
- Croo, H. D., Bandmann, M., Mackay, G. M., Rumar, K., & Vollenhoven, P. (2001). The role of driver fatigue in commercial road transport crashes. *Eur. Transp. Safety Council, Brussels, Belgium, Tech. Rep.*
- Dababneh, L., & Gindy, M. E. (2015). Driver vigilance level detection systems: a literature survey. *International Journal of Vehicle Performance*, 2(1), 1.
- Damousis, I., Cester, I., Nikolaou, S., & Tzovaras, D. (2007). Physiological indicators based sleep onset prediction for the avoidance of driving accidents. In *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007. EMBS 2007* (p. 6699-6704).
- Davies, D. R., & Parasuraman, R. (1982). *The psychology of vigilance*. Academic Press.
- Dawson, M. E., Schell, A. M., & Fillion, D. L. (2007). The electrodermal system. *Handbook of psychophysiology*, 2, 200–223.
- Daza, I. G., Hernandez, N., Bergasa, L. M., Parra, I., Yebes, J. J., Gavilan, M., ... Sotelo, M. A. (2011). Drowsiness monitoring based on driver and driving data fusion. In *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on* (p. 1199–1204). IEEE.
- Daza, Iván G., Bergasa, L. M., Bronte, S., Yebes, J. J., Almazán, J., & Arroyo, R. (2014). Fusion of optimized indicators from Advanced Driver Assistance Systems (ADAS) for driver drowsiness detection. *Sensors*, 14(1), 1106–1131.
- De Gennaro, L., Ferrara, M., Curcio, G., & Cristiani, R. (2001). Antero-posterior EEG changes during the wakefulness–sleep transition. *Clinical Neurophysiology*, 112(10), 1901-1911.
- De Valck, E., De Groot, E., & Cluydts, R. (2003). Effects of slow-release caffeine and a nap on driving simulator performance after partial sleep deprivation. *Perceptual and motor skills*, 96(1), 67–78.
- Di Milia, L., Smolensky, M. H., Costa, G., Howarth, H. D., Ohayon, M. M., & Philip, P. (2011). Demographic factors, fatigue, and driving accidents: An examination of the published literature. *Accident; Analysis and Prevention*, 43(2), 516-532.
- Dong, Y., Hu, Z., Uchimura, K., & Murayama, N. (2011). Driver inattention monitoring system for intelligent vehicles: A review. *Intelligent Transportation Systems, IEEE Transactions on*, 12(2), 596–614.
- Dufour, A. (2014). Driving assistance technologies and vigilance: Impact of speed limiters and cruise control on drivers' vigilance. In Presentation at the International Transport Forum (Paris, France).
- Ebrahim, I. O., Shapiro, C. M., Williams, A. J., & Fenwick, P. B. (2013). Alcohol and sleep I: effects on normal sleep. *Alcoholism: Clinical and Experimental Research*, 37(4).
- Enev, M., Takakuwa, A., Koscher, K., & Kohno, T. (2016). Automobile driver fingerprinting. *Proceedings on Privacy Enhancing Technologies*, 2016(1), 34–50.
- Elsenbruch, S., Harnish, M. J., & Orr, W. C. (1999). Heart rate variability during waking and sleep in healthy males and females. *Sleep*, 22(8), 1067–1071.
- Eskandarian, A., Sayed, R., Delaigue, P., Mortazavi, A., & Blum, J. (2007). Advanced Driver Fatigue Research.
- Eysenck, H. J. (1967). *The Biological Basis of Personality*. Transaction Publishers.

- Fairclough, S. H., & Graham, R. (1999). Impairment of driving performance caused by sleep deprivation or alcohol: a comparative study. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 41(1), 118–128.
- Fors, C., Ahlstrom, C., & Anund, A. (2018). A comparison of driver sleepiness in the simulator and on the real road. *Journal of Transportation Safety & Security*, 10(1-2), 72-87.
- Fowles, D. C., Christie, M. J., Edelberg, R., Grings, W. W., Lykken, D. T., & Venables, P. H. (1981). Publication Recommendations for Electrodermal Measurements. *Psychophysiology*, 18(3), 232-239.
- Friedrichs, F., & Yang, B. (2010). Drowsiness monitoring by steering and lane data based features under real driving conditions. In *Proceedings of the European Signal Processing Conference, Aalborg, Denmark* (p. 23–27).
- Fuletra, J. D., & Bosamiya, D. (2013). A survey on driver's drowsiness detection techniques. *International Journal on Recent and Innovation Trends in Computing and Communication*, 1(11), 816-819.
- Gabaude, C. (2010). Les défauts d'attention au volant : définitions, état des connaissances et pistes de recherches pour la prévention des risques routiers. *Archives des Maladies Professionnelles et de l'Environnement*, 71(3), 533-536. <https://doi.org/10.1016/j.admp.2010.03.047>
- Golding, J. F. (1998). Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness. *Brain research bulletin*, 47(5), 507–516.
- Grandjean, E. (1979). Fatigue in industry. *British Journal of Industrial Medicine*, 36(3), 175-186.
- Hagan, M. T., & Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE transactions on Neural Networks*, 5(6), 989–993.
- Hajinorozi, M., Mao, Z., & Huang, Y. (2015). Prediction of driver's drowsy and alert states from EEG signals with deep learning. In *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)* (p. 493-496).
- Hallvig, D., Anund, A., Fors, C., Kecklund, G., Karlsson, J. G., Wahde, M., & Åkerstedt, T. (2013). Sleepy driving on the real road and in the simulator—A comparison. *Accident Analysis & Prevention*, 50, 44-50.
- Hancock, P. A., & Parasuraman, R. (1992). Human factors and safety in the design of intelligent vehicle-highway systems (IVHS). *Journal of Safety Research*, 23(4), 181-198.
- Hargutt, V., & Kruger, H. P. (2001). Eyelid movements and their predictive value for fatigue stages. In *international conference on traffic and transport psychology - ICTTP 2000, HELD 4-7 September 2000, Bern, Switzerland*.
- Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *Intelligent Transportation Systems, IEEE Transactions on*, 6(2), 156–166.
- Hoc, J.-M. (2007). Human and automation: a matter of cooperation. *Human* 07, 277–285.
- Hockey, G. R. J. (2003). *Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks*. IOS Press.
- Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: a new approach. *Psychophysiology*, 10(4), 431–436.
- Hong, J., Suh, E., & Kim, S.-J. (2009). Context-aware systems: A literature review and classification. *Expert Systems with Applications*, 36(4), 8509-8522.
- Horne, J. A., & Östberg, O. (1975). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International journal of chronobiology*, 4(2), 97–110.
- Horne, J. A., & Reyner, L. (1999). Vehicle accidents related to sleep: a review. *Occupational and environmental medicine*, 56(5), 289–294.
- Horne, J. A., & Reyner, L. A. (1996). Counteracting driver sleepiness: Effects of napping, caffeine, and placebo. *Psychophysiology*, 33(3), 306-309.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5), 359–366.

- Inagaki, T. (2008). Smart collaboration between humans and machines based on mutual understanding. *Annual Reviews in Control*, 32(2), 253-261.
- Ingre, M., Åkerstedt, T., Peters, B., Anund, A., & Kecklund, G. (2006). Subjective sleepiness, simulated driving performance and blink duration: examining individual differences. *Journal of sleep research*, 15(1), 47-53.
- Israel, S. A., Irvine, J. M., Cheng, A., Wiederhold, M. D., & Wiederhold, B. K. (2005). ECG to identify individuals. *Pattern Recognition*, 38(1), 133-142.
- Jacobé de Naurois, C., Bourdin, C., Stratulat, A., Diaz, E., & Vercher, J.-L. (s. d.). Detection and prediction of driver drowsiness using artificial neural network models. *Accident Analysis & Prevention*.
- Jagannath, M., & Balasubramanian, V. (2014). Assessment of early onset of driver fatigue using multimodal fatigue measures in a static simulator. *Applied Ergonomics*.
- Jain, A., Hong, L., & Pankanti, S. (2000). Biometric identification. *Communications of the ACM*, 43(2), 90-98.
- Ji, Q., Lan, P., & Looney, C. (2006). A probabilistic framework for modeling and real-time monitoring human fatigue. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 36(5), 862-875.
- Ji, Q., Zhu, Z., & Lan, P. (2004). Real-time nonintrusive monitoring and prediction of driver fatigue. *Vehicular Technology, IEEE Transactions on*, 53(4), 1052-1068.
- Johns, M. (1998). Rethinking the assessment of sleepiness. *Sleep Medicine Reviews*, 2(1), 3-15.
- Johns, M. W., Tucker, A., Chapman, R., Crowley, K., & Michael, N. (2007). Monitoring eye and eyelid movements by infrared reflectance oculography to measure drowsiness in drivers. *Somnologie - Schlaforschung Und Schlafmedizin*, 11(4), 234-242.
- Johns, Murray W. (1991). A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *sleep*, 14(6), 540-545.
- Ju, J. H., Park, Y. J., Park, J., Lee, B. G., Lee, J., & Lee, J. Y. (2015). Real-Time Driver's Biological Signal Monitoring System. *Sensors and Materials*, 27(1), 51-59.
- Kaber, D. B., & Endsley, M. R. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress*, 16(3), 126-131.
- Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology*, 117(7), 1574-1581.
- Kaida, K., Åkerstedt, T., Kecklund, G., Nilsson, J. P., & Axelsson, J. (2007). Use of subjective and physiological indicators of sleepiness to predict performance during a vigilance task. *Industrial Health*, 45(4), 520-526.
- Karrer, K., Vöhringer-Kuhnt, T., Baumgarten, T., & Briest, S. (2004). The role of individual differences in driver fatigue prediction. In *Third International Conference on Traffic and Transport Psychology, Nottingham, UK* (p. 5-9). Citeseer.
- Karrer-Gauß, K., & Zawistowski, P. (2013). Driver Fatigue Systems-How Do They Change the Drivers' Behaviour? In *6th International on Conference Driver Behaviour and Training*.
- Keshan, N., Parimi, P. V., & Bichindaritz, I. (2015). Machine learning for stress detection from ECG signals in automobile drivers. In *Big Data (Big Data), 2015 IEEE International Conference on* (p. 2661-2669). IEEE.
- Khushaba, R. N., Kodagoda, S., Lal, S., & Dissanayake, G. (2011). Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm. *IEEE Transactions on Biomedical Engineering*, 58(1), 121-131.
- Khushaba, R. N., Kodagoda, S., Lal, S., & Dissanayake, G. (2013). Uncorrelated fuzzy neighborhood preserving analysis based feature projection for driver drowsiness recognition. *Fuzzy Sets and Systems*, 221, 90-111.

- Kohonen, T., Barna, G., & Chrisley, R. (1988). Statistical pattern recognition with neural networks: benchmarking studies. In , *IEEE International Conference on Neural Networks, 1988* (p. 61-68 vol.1).
- Körber, M., Cingel, A., Zimmermann, M., & Bengler, K. (2015). Vigilance Decrement and Passive Fatigue Caused by Monotony in Automated Driving. *Procedia Manufacturing, 3*, 2403-2409.
- Krajewski, J., Batliner, A., & Golz, M. (2009). Acoustic sleepiness detection: Framework and validation of a speech-adapted pattern recognition approach. *Behavior Research Methods, 41*(3), 795-804.
- Krajewski, J., Sommer, D., Trutschel, U., Edwards, D., & Golz, M. (2009). Steering wheel behavior based estimation of fatigue. In *Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 118–124)
- Kurt, M. B., Sezgin, N., Akin, M., Kirbas, G., & Bayram, M. (2009). The ANN-based computing of drowsy level. *Expert Systems with Applications, 36*(2), 2534–2542.
- Lal, S. K., & Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biological psychology, 55*(3), 173-194.
- Lal, S. K., & Craig, A. (2002). Driver fatigue: electroencephalography and psychological assessment. *Psychophysiology, 39*(3), 313–321.
- Lal, S. K. L., Craig, A., Boord, P., Kirkup, L., & Nguyen, H. (2003). Development of an algorithm for an EEG-based driver fatigue countermeasure. *Journal of Safety Research, 34*(3), 321-328.
- Larue, G. S. (2010). *Predicting effects of monotony on driver's vigilance* (Doctoral dissertation, . Queensland University of Technology).
- Larue, G. S., Rakotonirainy, A., & Pettitt, A. N. (2011). Real-time evaluation of driver's alertness on highways (p. 553-564).
- Larue, G. S., Rakotonirainy, A., & Pettitt, A. N. (2015). Predicting Reduced Driver Alertness on Monotonous Highways. *IEEE Pervasive Computing, 14*(2), 78-85.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature, 521*(7553), 436.
- Lee, B. L., Lee, B. G., & Chung, W. Y. (2016). Standalone Wearable Driver Drowsiness Detection System in a Smartwatch. *IEEE Sensors Journal, 16*(13), 5444-5451.
- Lee, B.-G., & Chung, W.-Y. (2012). Driver Alertness Monitoring Using Fusion of Facial Features and Bio-Signals. *IEEE Sensors Journal, 12*(7), 2416-2422.
- Lee, B.-G., Lee, B.-L., & Chung, W.-Y. (2014). Mobile Healthcare for Automatic Driving Sleep-Onset Detection Using Wavelet-Based EEG and Respiration Signals. *Sensors, 14*(10), 17915-17936.
- Leger, D., Philip, P., Jarriault, P., Metlaine, A., & Choudat, D. (2009). Effects of a combination of napping and bright light pulses on shift workers' sleepiness at the wheel: a pilot study. *Journal of sleep research, 18*(4), 472–479.
- Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics, 2*(2), 164–168.
- Li, G., & Chung, W.-Y. (2013). Detection of Driver Drowsiness Using Wavelet Analysis of Heart Rate Variability and a Support Vector Machine Classifier. *Sensors, 13*(12), 16494-16511.
- Li, L., Wen, D., Zheng, N.-N., & Shen, L.-C. (2012). Cognitive Cars: A New Frontier for ADAS Research. *IEEE Transactions on Intelligent Transportation Systems, 13*(1), 395-407.
- Li, L., Werber, K., Calvillo, C. F., Dinh, K. D., Guardie, A., & König, A. (2014). Multi-Sensor Soft-Computing System for Driver Drowsiness Detection. In V. Snášel, P. Krömer, M. Köppen, & G. Schaefer (Éd.), *Soft Computing in Industrial Applications* (p. 129-140). Springer International Publishing.
- Li, X., Seignez, E., Lambert, A., & Loonis, P. (2014). Real-time driver drowsiness estimation by multi-source information fusion with Dempster–Shafer theory. *Transactions of the Institute of Measurement & Control, 36*(7), 906-915.
- Liang, Y., Reyes, M. L., & Lee, J. D. (2007). Real-Time Detection of Driver Cognitive Distraction Using Support Vector Machines. *IEEE Transactions on Intelligent Transportation Systems, 8*(2), 340-350.

- Lin, F.-C., Ko, L.-W., Chuang, C.-H., Su, T.-P., & Lin, C.-T. (2012). Generalized EEG-based drowsiness prediction system by using a self-organizing neural fuzzy system. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 59(9), 2044–2055.
- Lin, S. D., Lin, J.-J., & Chung, C.-Y. (2013). Sleepy eye's recognition for drowsiness detection. In *Biometrics and Security Technologies (ISBAST), 2013 International Symposium on* (p. 176–179). IEEE.
- Liu, C. C., Hosking, S. G., & Lenné, M. G. (2009). Predicting driver drowsiness using vehicle measures: Recent insights and future challenges. *Journal of safety research*, 40(4), 239–245.
- Liu, Y. T., Wu, S. L., Chou, K.-P., Lin, Y. Y., Lu, J., Zhang, G., ... Lin, C. T. (2016). Driving fatigue prediction with pre-event electroencephalography (EEG) via a recurrent fuzzy neural network. In *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (p. 2488-2494).
- Liu, Y.-T., Lin, Y.-Y., Wu, S.-L., Chuang, C.-H., & Lin, C.-T. (2016). Brain dynamics in predicting driving fatigue using a recurrent self-evolving fuzzy neural network. *IEEE transactions on neural networks and learning systems*, 27(2), 347–360.
- Liu, Y.-T., Lin, Y.-Y., Wu, S.-L., Hsieh, T.-Y., & Lin, C.-T. (2015). Assessment of Mental Fatigue: An EEG-Based Forecasting System for Driving Safety. In *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on* (p. 3233–3238). IEEE.
- Louw, T., Merat, N., & Jamson, H. (2015). Engaging with Highly Automated Driving: To be or Not to be in the Loop?
- Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, 4(11), 432-440.
- Lumley, M., Roehrs, T., Asker, D., Zorick, F., & Roth, T. (1987). Ethanol and caffeine effects on daytime sleepiness/alertness. *Sleep*, 10(4), 306–312.
- Lyznicki, J. M., Doege, T. C., Davis, R. M., & Williams, M. A. (1998). Sleepiness, driving, and motor vehicle crashes. *Jama*, 279(23), 1908-1913.
- Maia, Q., Grandner, M. A., Findley, J., & Gurubhagavatula, I. (2013). Short and long sleep duration and risk of drowsy driving and the role of subjective sleep insufficiency. *Accident Analysis & Prevention*, 59, 618–622.
- Marin-Lamellet, C., Paire-Ficout, L., Lafont, S., Amieva, H., Laurent, B., Thomas-Antérion, C., ... Fabrigoule, C. (2003). Mise en place d'un outil d'évaluation des déficits attentionnels affectant les capacités de conduite au cours du vieillissement normal et pathologique: L'étude SÉROVIE. *Recherche - Transports - Sécurité*, 81, 177-189.
- Matthews, G. (2000). *Human performance: Cognition, stress, and individual differences*. Psychology Press.
- McDonald, A. D., Lee, J. D., Schwarz, C., & Brown, T. L. (2013). Steering in a Random Forest Ensemble Learning for Detecting Drowsiness-Related Lane Departures. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 0018720813515272.
- McDonald, A. D., Schwarz, C., Lee, J. D., & Brown, T. L. (2012). Real-Time Detection of Drowsiness Related Lane Departures Using Steering Wheel Angle. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56(1), 2201-2205.
- Mehler, B., Reimer, B., Coughlin, J., & Dusek, J. (2009). Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. *Transportation Research Record: Journal of the Transportation Research Board*, 2138, 6-12.
- Merat, N., & Lee, J. D. (2012). Preface to the special section on human factors and automation in vehicles: Designing highly automated vehicles with the driver in mind. *Human factors*, 54(5), 681–686.
- Miles, W. (1929). Sleeping with the Eyes Open. *Scientific American*, 140(6), 489-492.
- Millemann, S., Panis, S., Loslever, P., Popieul, J.-C., Simon, P., & Dillies-Peltier, M.-A. (2001). Car driving activity: Human and vehicle behaviour on a long monotonous journey (p. 69-78). Présenté à Driving simulation conference.

- Murata, A., & Naitoh, K. (2015). Multinomial Logistic Regression Model for Predicting Driver's Drowsiness Using Only Behavioral Measures. *Journal of Traffic and Transportation Engineering*, 3, 80–90.
- Murata, A., Ohta, Y., & Moriwaka, M. (2016). Multinomial Logistic Regression Model by Stepwise Method for Predicting Subjective Drowsiness Using Performance and Behavioral Measures. In R. Goonetilleke & W. Karwowski (Éd.), *Advances in Physical Ergonomics and Human Factors* (Vol. 489, p. 665-674). Cham: Springer International Publishing.
- Nakayama, O., Futami, T., Nakamura, T., & Boer, E. R. (1999). *Development of a Steering Entropy Method for Evaluating Driver Workload* (SAE Technical Paper No. 1999- 01-0892). Warrendale, PA: SAE International. f
- Navarro, J. (2010). *Coopération homme-machine en conduite automobile assistée: Contrôle cognitif et contrôle de la trajectoire*. Université de Nantes.
- Neboit, M. (1980). L'exploration visuelle dans l'apprentissage de tâches complexes. L'exemple de la conduite automobile.
- Nikolova, R., Radev, V., & Vukov, M. (2003). Functional determination of the operator state in the interaction of humans with automated systems.
- Nilsson, E., Ahlström, C., Barua, S., Fors, C., Lindén, P., Svanberg, B., ... Anund, A. (2017). *Vehicle driver monitoring: sleepiness and cognitive load*. Statens väg-och transportforskningsinstitut.
- Oron-Gilad, T., Ronen, A., & Shinar, D. (2008). Alertness maintaining tasks (AMTs) while driving. *Accident Analysis & Prevention*, 40(3), 851-860.
- Pack, A. I., Pack, A. M., Rodgman, E., Cucchiara, A., Dinges, D. F., & Schwab, C. W. (1995). Characteristics of crashes attributed to the driver having fallen asleep. *Accident Analysis & Prevention*, 27(6), 769–775.
- Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345–1359.
- Patel, M., Lal, S. K., Kavanagh, D., & Rossiter, P. (2011). Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert systems with Applications*, 38(6), 7235–7242.
- Peiris, M. T. R., Jones, R. D., Davidson, P. R., Carroll, G. J., Signal, T. L., Parkin, P. J., ... Bones, P. J. (2005). Identification of vigilance lapses using EEG/EOG by expert human raters.
- Philip, P., Sagaspe, P., Lagarde, E., Leger, D., Ohayon, M. M., Bioulac, B., ... Taillard, J. (2010). Sleep disorders and accidental risk in a large group of regular registered highway drivers. *Sleep medicine*, 11(10), 973–979.
- Philip, P., Sagaspe, P., Taillard, J., Valtat, C., Moore, N., Åkerstedt, T., ... Bioulac, B. (2005). Fatigue, sleepiness, and performance in simulated versus real driving conditions. *Sleep*, 28(12), 1511.
- Philip, P., Taillard, J., Guilleminault, C., Quera, S., Bioulac, B., & Ohayon, M. (1999). Long distance driving and self-induced sleep deprivation among automobile drivers. *Sleep*, 22(4), 475–480.
- Philip, P., Taillard, J., Quera-Salva, M., Bioulac, B., & Åkerstedt, T. (1999). Simple reaction time, duration of driving and sleep deprivation in young versus old automobile drivers. *Journal of Sleep research*, 8(1), 9–14.
- Philip, P., Taillard, J., Sagaspe, P., Valtat, C., Sanchez-Ortuno, M., Moore, N., ... Bioulac, B. (2004). Age, performance and sleep deprivation. *Journal of Sleep Research*, 13(2), 105–110.
- Picot, A. (2009). *Détection d'hypovigilance chez le conducteur par fusion d'informations physiologiques et vidéo*. (Doctoral dissertation, Grenoble, INPG).
- Pressman, M. R., & Fry, J. M. (1989). Relationship of autonomic nervous system activity to daytime sleepiness and prior sleep. *Sleep*, 12(3), 239-245.
- Pribram, K. H., & McGuinness, D. (1975). Arousal, activation, and effort in the control of attention. *Psychological review*, 82(2), 116.
- Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A. S., McNamara, J. O., & Williams, S. M. (2001). *Neuroscience 2nd Edition*. Sunderland (MA) Sinauer Associates. Inc.

- Raudenbush, B., Grayhem, R., Sears, T., & Wilson, I. (2009). Effects of peppermint and cinnamon odor administration on simulated driving alertness, mood and workload. *N Am J Psychol*, *11*, 245–256.
- Rabat, A., Gomez-Merino, D., Roca-Paixao, L., Bougard, C., Van Beers, P., Dispersyn, G., ... & Sauvet, F. (2016). Differential kinetics in alteration and recovery of cognitive processes from a chronic sleep restriction in young healthy men. *Frontiers in behavioral neuroscience*, *10*, 95.
- Rebolledo-Mendez, G., Reyes, A., Paszkowicz, S., Domingo, M. C., & Skrypchuk, L. (2014). Developing a Body Sensor Network to Detect Emotions During Driving. *IEEE Transactions on Intelligent Transportation Systems*, *15*(4), 1850.
- Reimer, B., Coughlin, J. F., & Mehler, B. (2009). Development of a driver aware vehicle for monitoring, managing & motivating older operator behavior. *Proceedings of the ITS-America*, 1–9.
- Rétey, J. V., Adam, M., Gottselig, J. M., Khatami, R., Dürr, R., Achermann, P., & Landolt, H.-P. (2006). Adenosinergic mechanisms contribute to individual differences in sleep deprivation-induced changes in neurobehavioral function and brain rhythmic activity. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(41), 10472-10479.
- Reyner, L. A., & Home, J. A. (1998). Evaluation of 'in-car' countermeasures to sleepiness: cold air and radio. *Sleep*, *21*(1), 46–51.
- Richard, J.-B., Thélot, B., & Beck, F. (2013). Les accidents en France: évolution et facteurs associés. *Revue d'Épidémiologie et de Santé Publique*, *61*(3), 205–212.
- Riemersma, J. B. J., Sanders, A. F., Wildervanck, C., & Gaillard, A. W. (1977). Performance decrement during prolonged night driving. In *Vigilance* (p. 41–58). Springer.
- Rodriguez Ibañez, N., García González, M. Á., Ramos Castro, J. J., & Fernández Chimeno, M. (2011). Drowsiness detection by thoracic effort signal analysis with professional drivers in real environments. In *Driver Distraction & Inattention 2011: program, presentations & reviewed papers 2011*.
- Rohit, F., Kulathumani, V., Kavi, R., Elwarfalli, I., Kecojevic, V., & Nimbarte, A. (2017). Real-time drowsiness detection using wearable, lightweight brain sensing headbands. *IET Intelligent Transport Systems*, *11*(5), 255-263.
- Rossi, R., Gastaldi, M., & Gecchele, G. (2011). Analysis of driver task-related fatigue using driving simulator experiments. *Procedia - Social and Behavioral Sciences*, *20*, 666-675.
- Rost, M., Zilberg, E., Xu, Z. M., Feng, Y., Burton, D., & Lal, S. (2015). Comparing Contribution of Algorithm Based Physiological Indicators for Characterisation of Driver Drowsiness. *Journal of Medical and Bioengineering Vol*, *4*(5), 391–398.
- Sabet, M., Zoroofi, R. A., Sadeghniiat-Haghighi, K., & Sabbaghian, M. (2012). A new system for driver drowsiness and distraction detection. In *2012 20th Iranian Conference on Electrical Engineering (ICEE)* (p. 1247-1251).
- Sahayadhas, A., Sundaraj, K., & Murugappan, M. (2012). Detecting Driver Drowsiness Based on Sensors: A Review. *Sensors*, *12*(12), 16937-16953.
- Samiee, S., Azadi, S., Kazemi, R., Nahvi, A., & Eichberger, A. (2014). Data Fusion to Develop a Driver Drowsiness Detection System with Robustness to Signal Loss. *Sensors (14248220)*, *14*(9), 17832.
- Schaefer, F., Haarmann, B., & Boucsein, W. (2008). The usability of cardiovascular and electrodermal measures for adaptive automation. *Probing Experience: From Assessment of User Emotions and Behavior to Development of Products. Philips Research Book Series*, *8*, 235-243.
- Schlutz, A. P. (2006). *Driving without Awareness. Eine experimentelle Studie zumVigilanzabfall während des Autofahrens*. Technische Universität Berlin.
- Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., & Kincses, W. E. (2009). Drivers' misjudgement of vigilance state during prolonged monotonous daytime driving. *Accident Analysis & Prevention*, *41*(5), 1087-1093.

- Schmidt, J., Braunagel, C., Stolzmann, W., & Karrer-Gauß, K. (2016). Driver drowsiness and behavior detection in prolonged conditionally automated drives. In *Intelligent Vehicles Symposium (IV), 2016 IEEE* (p. 400–405). IEEE.
- Schmidt, J., Laarousi, R., Stolzmann, W., & Karrer-Gauß, K. (2017). Eye blink detection for different driver states in conditionally automated driving and manual driving using EOG and a driver camera. *Behavior research methods*, 1–14.
- Schwartz, B. (2009). *The paradox of choice*. HarperCollins.
- Schwarz, J. F. A., Ingre, M., Fors, C., Anund, A., Kecklund, G., Taillard, J., ... Åkerstedt, T. (2012). In-car countermeasures open window and music revisited on the real road: popular but hardly effective against driver sleepiness. *Journal of Sleep Research*, 21(5), 595-599.
- Shahid, A., Wilkinson, K., Marcu, S., & Shapiro, C. M. (2011). Karolinska Sleepiness Scale (KSS). In A. Shahid, K. Wilkinson, S. Marcu, & C. M. Shapiro (Éd.), *STOP, THAT and One Hundred Other Sleep Scales* (p. 209-210). Springer New York.
- Shen, K.-Q., Li, X.-P., Ong, C.-J., Shao, S.-Y., & Wilder-Smith, E. P. (2008). EEG-based mental fatigue measurement using multi-class support vector machines with confidence estimate. *Clinical Neurophysiology*, 119(7), 1524–1533.
- Singh, Y. N., & Gupta, P. (2008). ECG to Individual Identification. In *2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems* (p. 1-8).
- Smiley, A. (1998). Fatigue management: lessons from research. In *INTERNATIONAL CONFERENCE ON FATIGUE AND TRANSPORTATION, 3RD, 1998, FREMANTLE, WESTERN AUSTRALIA*.
- Smith, D., Tong, J., & Leigh, G. (1977). Combined effects of tobacco and caffeine on the components of choice reaction-time, heart rate, and hand steadiness. *Perceptual and Motor Skills*, 45(2), 635-639.
- Smith, P., Shah, M., & da Vitoria Lobo, N. (2000). Monitoring head/eye motion for driver alertness with one camera. In *15th International Conference on Pattern Recognition, 2000. Proceedings* (Vol. 4, p. 636-642 vol.4).
- Stein, A. C. (1995). *Detecting fatigued drivers with vehicle simulators*.
- Stein, P. K., & Pu, Y. (2012). Heart rate variability, sleep and sleep disorders. *Sleep Medicine Reviews*, 16(1), 47–66.
- Stutts, J. C., Wilkins, J. W., & Vaughn, B. V. (1999). Why do people have drowsy driving crashes. *Input from drivers who just did*, (202/638), 5944.
- Summala, H., & Mikkola, T. (1994). Fatal accidents among car and truck drivers: effects of fatigue, age, and alcohol consumption. *Human factors*, 36(2), 315–326.
- Sun, Y., & Yu, X. B. (2014). An Innovative Nonintrusive Driver Assistance System for Vital Signal Monitoring. *IEEE Journal of Biomedical & Health Informatics*, 18(6), 1932.
- Taillard, J., Capelli, A., Sagaspe, P., Anund, A., Åkerstedt, T., & Philip, P. (2012). In-Car Nocturnal Blue Light Exposure Improves Motorway Driving: A Randomized Controlled Trial. *PLoS ONE*, 7(10), e46750.
- Takahashi, I., & Yokoyama, K. (2011). Development of a feedback stimulation for drowsy driver using heartbeat rhythms. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE* (p. 4153–4158). IEEE.
- Tango, F., Calefato, C., Minin, L., & Canovi, L. (2009). Moving attention from the road: A new methodology for the driver distraction evaluation using machine learning approaches. In *2nd Conference on Human System Interactions, 2009. HSI '09* (p. 596-599).
- Thiffault, P., & Bergeron, J. (1997). Vigilance et conduite sur autoroute: Différences interindividuelles et résistance à la monotonie. In *Proceedings of the Xth Canadian Multidisciplinary Conference on Road Safety. Toronto* (p. 454–464).
- Thiffault, Pierre, & Bergeron, J. (2003a). Fatigue and individual differences in monotonous simulated driving. *Personality and Individual Differences*, 34(1), 159-176.
- Thiffault, Pierre, & Bergeron, J. (2003b). Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis & Prevention*, 35(3), 381–391.

- Thorslund, B. (2004). *Electrooculogram analysis and development of a system for defining stages of drowsiness*. Statens väg-och transportforskningsinstitut.
- Tjolleng, A., Jung, K., Hong, W., Lee, W., Lee, B., You, H., Son, J. & Park, S. (2017). Classification of a Driver's cognitive workload levels using artificial neural network on ECG signals. *Applied Ergonomics*, 59, Part A, 326-332.
- Torkkola, K., Gardner, M., Schreiner, C., Zhang, K., Leivian, B., Zhang, H., & Summers, J. (2008). Understanding Driving Activity Using Ensemble Methods. In D. Prokhorov (Ed.), *Computational Intelligence in Automotive Applications* (pp. 39–58). Springer Berlin Heidelberg.
- Torrey, L., & Shavlik, J. (2009). Transfer learning. *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, 1, 242.
- Tu, W., & Sun, S. (2012). A subject transfer framework for EEG classification. *Neurocomputing*, 82, 109–116.
- Vaez Mousavi, M., Barry, R. J., Rushby, J., & Clarke, A. (2007). Evidence for differentiation of arousal and activation in normal adults.
- Van Dongen, H. P. A., Baynard, M. D., Maislin, G., & Dinges, D. F. (2004). Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. *Sleep*, 27(3), 423–433.
- Van Dongen, H. P. A., Maislin, G., & Dinges, D. F. (2004). Dealing with inter-individual differences in the temporal dynamics of fatigue and performance: importance and techniques. *Aviation, Space, and Environmental Medicine*, 75(3), A147–A154.
- Van Dongen, H. P. A., Rogers, N. L., & Dinges, D. F. (2003). Sleep debt: Theoretical and empirical issues. *Sleep and Biological Rhythms*, 1(1), 5-13.
- Vehicle security system including fingerprint and eyeball part identification. (1996, juin 21).
- Verwey, W. B., & Zaidel, D. M. (2000). Predicting drowsiness accidents from personal attributes, eye blinks and ongoing driving behaviour. *Personality and Individual Differences*, 28(1), 123-142.
- Vicente, J., Laguna, P., Bartra, A., & Bailón, R. (2016). Drowsiness detection using heart rate variability. *Medical & biological engineering & computing*, 54(6), 927–937.
- Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, p. I–I). IEEE.
- Vural, E., Cetin, M., Ercil, A., Littlewort, G., Bartlett, M., & Movellan, J. (2007). Drowsy driver detection through facial movement analysis. *Human-computer interaction*, 6–18.
- Wang, J., Li, K., & Lu, X.-Y. (2014). Chapter 5 - Effect of Human Factors on Driver Behavior. In Y. C. Li (Éd.), *Advances in Intelligent Vehicles* (p. 111-157). Boston: Academic Press
- Wang, P., Lu, J., Zhang, B., & Tang, Z. (2015). A review on transfer learning for brain-computer interface classification. In *Information Science and Technology (ICIST), 2015 5th International Conference on* (p. 315–322). IEEE.
- Wang, J., Sun, S., Fang, S., Fu, T., & Stipanovic, J. (2017). Predicting drowsy driving in real-time situations: using an advanced driving simulator, accelerated failure time model, and virtual location-based services. *Accident Analysis & Prevention*, 99, 321-329.
- Watson, A., & Zhou, G. (2016). Microsleep Prediction Using an EKG Capable Heart Rate Monitor. In *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)* (p. 328-329).
- Wei, C.-S., Lin, Y.-P., Wang, Y.-T., Jung, T.-P., Bigdely-Shamlo, N., & Lin, C.-T. (2015). Selective transfer learning for EEG-based drowsiness detection. In *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on* (p. 3229–3232). IEEE.
- Wesensten, N. J., Belenky, G., Thorne, D. R., Kautz, M. A., & Balkin, T. J. (2004). Modafinil vs. caffeine: Effects on fatigue during sleep deprivation. *Aviation Space and Environmental Medicine*, 75(6), 520–525.
- Wickens, C. D., & Hollands, J. G. *Engineering psychology and human performance*. 2000. ISBN: 0-321-04711-7.

- Wiegand, D. M., McClafferty, J., McDonald, S. E., Hanowski, R. J. (2009). Development and evaluation of a naturalistic observer rating of drowsiness protocol: final report.
- Wierwille, W. W., & Ellsworth, L. A. (1994). Evaluation of driver drowsiness by trained raters. *Accident Analysis & Prevention*, 26(5), 571-581.
- Williams, G. W., & Shor, R. E. (1970). An historical note on highway hypnosis. *Accident Analysis & Prevention*, 2(3), 223–225.
- Wilson, G. F., & Russell, C. (1999). Operator Functional State Classification Using Neural Networks with Combined Physiological and Performance Features. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 43(20), 1099-1102.
- Wu, D., Lawhern, V. J., Gordon, S., Lance, B. J., & Lin, C. T. (2016). Driver Drowsiness Estimation from EEG Signals Using Online Weighted Adaptation Regularization for Regression (OwARR). *IEEE Transactions on Fuzzy Systems*, PP(99), 1-1.
- Wu, Dongrui, Chuang, C.-H., & Lin, C.-T. (2015). Online driver's drowsiness estimation using domain adaptation with model fusion. In *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on* (p. 904–910). IEEE.
- Wu, Dongrui, Lawhern, V. J., Gordon, S., Lance, B. J., & Lin, C.-T. (2016). Offline EEG-based driver drowsiness estimation using enhanced batch-mode active learning (EBMAL) for regression. In *Proc. IEEE Int'l Conf. on Systems, Man and Cybernetics*.
- Wylie, C. D., Shultz, T., Mitler, M. M., & Mackie, R. R. (1996). *Commercial motor vehicle driver fatigue and alertness study: Technical summary*.
- Yang, G., Lin, Y., & Bhattacharya, P. (2010). A driver fatigue recognition model based on information fusion and dynamic Bayesian network. *Information Sciences*, 180(10), 1942–1954.
- Yeo, M. V. M., Li, X., Shen, K., & Wilder-Smith, E. P. V. (2009). Can SVM be used for automatic EEG detection of drowsiness during car driving? *Safety Science*, 47(1), 115-124.
- Yin, Z., & Zhang, J. (s. d.). Cross-subject recognition of operator functional states via EEG and switching deep belief networks with adaptive weights. *Neurocomputing*.
- Young, K., Lee, J. D., & Regan, M. A. (2008). *Driver Distraction: Theory, Effects, and Mitigation*. CRC Press.
- Zhang, J., Wang, Y., & Li, S. (2017). Cross-subject mental workload classification using kernel spectral regression and transfer learning techniques. *Cognition, Technology & Work*, 19(4), 587-605.
- Zhang, Y., Owechko, Y., & Zhang, J. (2004). Driver cognitive workload estimation: a data-driven perspective. In *The 7th International IEEE Conference on Intelligent Transportation Systems, 2004. Proceedings* (p. 642-647).
- Zhang, Z., Luo, D., Rasim, Y., Li, Y., Meng, G., Xu, J., & Wang, C. (2016). A Vehicle Active Safety Model: Vehicle Speed Control Based on Driver Vigilance Detection Using Wearable EEG and Sparse Representation. *Sensors*, 16(2), 242.
- Zhao, C., Zhao, M., Liu, J., & Zheng, C. (2012). Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. *Accident Analysis & Prevention*, 45, 83–90.
- Zhao, L., Wang, Z., Zhang, G., Qi, Y., & Wang, X. (2017). Eye state recognition based on deep integrated neural network and transfer learning. *Multimedia Tools and Applications*, 1-24.
- Zuckerman, M. (1979). *Sensation seeking: Beyond the optimal level of arousal*. 1979. Hillsdale, New Jersey: Erlbaum.

Résumé

La conduite automobile requiert la mobilisation des ressources cognitives et physiologiques dans l'objectif de maintenir les performances et la sécurité du conducteur ainsi que des passagers. C'est pourquoi de nombreux chercheurs ont étudié la problématique de la somnolence au volant. Détecter quand le conducteur est somnolent ou même prédire dans combien de temps il risque de le devenir sont devenus des défis importants pour le développement de nouveaux systèmes d'aide à la conduite (Advanced Driving Assistance Systems, ou ADAS). Cette thèse s'est non seulement intéressée à la détection de la somnolence, c'est-à-dire l'estimation à un instant t du niveau de somnolence du conducteur, mais également à la prédiction de la somnolence, c'est-à-dire l'estimation du temps avant l'occurrence d'un certain niveau critique de somnolence.

La somnolence, état intermédiaire entre le sommeil et la veille, représente un état dégradé pour la conduite et affecte donc potentiellement les performances de conduite. Ces travaux se sont intéressés à la modélisation de la somnolence au volant à partir de réseaux de neurones artificiels (ANN) et à l'aide de mesures physiologiques (rythme cardiaque et respiratoire et leurs variabilités), comportementales (mouvement des paupières et de tête) et de l'activité et performances de conduite (vitesse, angle au volant et déviation de la position sur la voie). Une première étude a montré qu'un ANN pouvait non seulement détecter le niveau de somnolence du conducteur compris entre 0 et 4 (alerte à extrêmement somnolent) avec une racine carrée de l'erreur quadratique moyenne de 0,40, mais aussi prédire dans combien de temps un état dégradé risquait d'arriver avec une racine carrée de l'erreur quadratique moyenne de 2,23 minutes. Le temps de conduite et les informations par rapport aux participants permettent d'accroître les performances. Ces modèles ont été par la suite testés sur un nouvel individu. Mais, dans ce cas, nous observons de mauvaises performances de généralisation. Ainsi nous avons testé une adaptation personnalisée de ces ANN, c'est-à-dire, qu'un ANN a été entraîné sur un ensemble de participants, puis il a été ensuite adapté à un nouveau conducteur, jamais vu précédemment par le modèle, grâce à ces premières données d'enregistrement. Grâce à cette adaptation personnalisée du ANN, nous observons une amélioration des performances de 40 et 80% pour la détection et la prédiction de la somnolence au volant respectivement. Ce processus d'adaptation grâce aux premières données d'enregistrement est une première réponse au problème de la variabilité inter-individuelle.

Abstract

Driving a car is a complex, multifaceted and potentially risky activity requiring full mobilization of physiological and cognitive resources to maintain performance over time and the security of driver and passengers. That is why many researchers study the driver drowsiness. Detecting when the driver is drowsy but also predicting when the driver's operational state begins to degrade has become one ambitious challenge for the development of new *Advanced Driving Assistance Systems (ADAS)*. This thesis is not only interested in detecting drowsiness (the estimation at an instant t of the driver's level of drowsiness), but also, estimation of time before the occurrence of an impaired driver state.

Drowsiness, the intermediate state between sleep and awake, represents an impaired state for driving and its potential effect on the driving performance. This work focuses on developing a driver drowsiness model by using artificial neural networks (ANN) and physiological measures (heart and respiratory rate and their variability), behavioral (eyelids and head movement) and driving and performance activity (speed, time-to-lane-crossing, speed, steering wheel angle, position on the lane). The first study shows that a model can detect the level of drowsiness between 0 and 4 (alert and extremely drowsy) with a root mean square error of 0.4 and also predict when the impaired state will occur with a root mean square error of 2.23 minutes. The driving time, the time elapsed since the beginning of the driving session, and personal information can enhance the performance. These models were then tested on a different participant, but in this case, we observe poor generalization performance. We then tested a personalized adaptation of this ANN, where the ANN was trained on a group of participants and then adapted to a new driver. With this personalized adaptation of the ANN, we observe a performance improvement of 40% and 80% for the detection and the prediction of driver drowsiness respectively. This adaptation process to the first data recorded is an initial response to the problem of inter-individual variability.
