

Driver drowsiness monitoring using eye movement features derived from electrooculography

Von der Fakultät Informatik, Elektrotechnik und Informationstechnik
der Universität Stuttgart
zur Erlangung der Würde eines Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Abhandlung

vorgelegt von

Parisa Ebrahim

aus Teheran

Hauptberichter:
Mitberichter:

Prof. Dr.-Ing. Bin Yang
Assistant Prof. Dr. Dongpu Cao

Tag der mündlichen Prüfung: 06.06.2016

**Institut für Signalverarbeitung und Systemtheorie
Universität Stuttgart**

2016

Accknowledgement

I would like to express my genuine thanks to my supervisor, Prof. Dr.-Ing. Bin Yang for his smart guidance, warm encouragements and helpful comments, starting from my Studienarbeit, continuing to Diplomarbeit and finally to this thesis. He sharpened my mind towards a clear and scientific way of thinking and developing new ideas. For all of this I owe him more than I can describe.

Next I would like to thank my co-referee Assistant Prof. Dr. Dongpu Cao for accepting to be in the committee, and reviewing my thesis.

I am also grateful to Dr. Wolfgang Stolzmann for his great support starting from my Diplomarbeit at Daimler AG. I appreciate all his contributions of time, ideas and comments to make a productive and stimulating working experience.

My deep gratitude goes to Dr. Klaus-Peter Kuhn for providing a good atmosphere in his team at Daimler AG. His precious comments have opened new doors for research possibilities, from which my thesis benefited tremendously.

Much of the research in this thesis was carried out as part of Attention Assist project at Daimler AG. I would like to acknowledge all of my colleagues in this project, who provided a great team work atmosphere. It was indeed enriching to be part of this project. I am deeply indebted to Dipl.-Ing. Alexander Fürsich and Dipl.-Ing. Peter hermannstädtter for the proof reading, data collection and being ear to all my queries/problems both scientific and otherwise. Their help, support and permanent positive attitude of collaboration showed me that friendship knows no boundaries. My special thanks go to my predecessor Dipl.-Ing. Fabian Friedrichs for his immense support and consultation.

I also greatly appreciate all of my colleagues at ISS for their help, feedback and emergency assistance.

My acknowledgments would not be complete without giving thanks to my lovely parents and my brother. They made all my dreams come true. I am so much grateful for their loving support and admiration over the years. They have constantly backed my decisions and choices, and had unwavering faith in me. Last, but certainly not least, my tremendous and deep thanks are extended to my love, my husband MJ, without whom I could not have completed this journey. His patient, support and tremendous belief in me were the main source of confidence and motivation to complete this thesis.

Contents

Notation and abbreviations	v
Zusammenfassung	xi
Abstract	xv
1 Introduction	1
1.1 Problem statement and motivation	1
1.2 Definition of drowsiness and inattention	3
1.3 Countermeasures against drowsiness during driving	6
1.4 Driver drowsiness detection systems on the market	7
1.5 Thesis outline	9
1.6 Goals and new contributions of the thesis	10
2 Driver state measurement	13
2.1 Objective driver state measures	13
2.1.1 Driving performance measures	13
2.1.2 Driver physiological measures	16
2.2 Subjective estimation of the drowsiness	26
3 Human visual system	31
3.1 Visual attention	31
3.2 Structure of the human eye	31
3.3 Types of eye movements	32
4 In-vehicle usage of electrooculography and conducted experiments	37
4.1 Eye movement measurement during driving - a pilot study	37
4.1.1 Material	37
4.1.2 Test tracks	38
4.1.3 Comparing baselines of track 1 and track 2	39
4.1.4 Bumps and eye movements: related or unrelated?	41
4.1.5 Patterns of eye movements due to road curvature	43
4.2 Real road experiments	46
4.2.1 Daytime driving with no secondary tasks	46
4.2.2 Daytime driving with secondary tasks	46
4.2.3 Nighttime driving with no secondary tasks	48
4.3 Nighttime driving experiment in the driving simulator	50
5 Eye movement event detection methods	53
5.1 Eye movement detection using the median filter-based method	53
5.2 Eye movement detection using the derivative-based method	55

5.3	Eye movement detection using the wavelet transform-based method	61
5.3.1	Discrete Fourier transform	61
5.3.2	Continuous wavelet transform	62
5.3.3	Discrete wavelet transform	71
5.4	Comparison of event detection methods	83
5.4.1	Median filter-based versus derivative-based method	84
5.4.2	Derivative-based method versus wavelet transform-based method	86
6	Blink behavior in distracted and undistracted driving	91
6.1	Time-on-task analysis of the saccade rate during the visuomotor task	92
6.2	Time-on-task analysis of the blink rate	92
6.3	Saccades time-locked to blinks during the visuomotor task	93
6.4	Direction-dependency of blinks time-locked to saccades	93
6.5	Blink rate analysis during the secondary and primary tasks	95
6.6	Impact of the visuomotor task on the blink behavior	96
6.7	Amount of gaze shift vs. the occurrence of gaze shift-induced blinks	97
7	Extraction and evaluation of the eye movement features	103
7.1	Preprocessing of eye movement features	103
7.1.1	KSS input-based feature aggregation	104
7.1.2	Drive time-based feature aggregation	105
7.1.3	Feature baselining	106
7.2	Eye blink features	107
7.3	Saccade features	120
7.4	Event-based analysis of eye blink features	122
7.4.1	Event 1: Lane departure	123
7.4.2	Event 2: Microsleep	127
7.5	Correlation-based analysis of eye blink features	130
7.5.1	Case 1: Correlation between a feature and KSS values	130
7.5.2	Case 2: Correlation between features	134
7.6	Eye blink feature's quality vs. sampling frequency	135
8	Driver state detection by machine learning methods	141
8.1	Introduction to machine learning	141
8.1.1	Supervised classification	142
8.1.2	Metrics for evaluating the performance of classifiers	144
8.1.3	Subject-dependent classification	145
8.1.4	Subject-independent classification	145
8.1.5	Imbalanced class distributions	146
8.2	Artificial neural network classifier	149
8.2.1	Network's architecture	150
8.2.2	Training of the network	151
8.2.3	Classification results of subject-dependent data sets	153
8.2.4	Classification results of the subject-independent data sets	158
8.3	Support vector machine classifier	159
8.3.1	Hard margin support vector machines	159
8.3.2	Soft margin support vector machines	161
8.3.3	Kernel trick	162

8.3.4	Model construction	163
8.3.5	Multi-class classification approaches	165
8.3.6	Dealing with imbalanced data	167
8.3.7	Classification results of subject-dependent data sets	168
8.3.8	Classification results of the subject-independent data sets	171
8.4	<i>k</i> -nearest neighbors classifier	172
8.4.1	Background theory	172
8.4.2	Classification results of the subject-dependent data sets	173
8.4.3	Classification results of the subject-independent data sets	175
8.5	Comparison of the supervised classifiers for driver state classification	176
8.6	Features of the driving simulator versus real road driving	179
8.6.1	Generalization of the simulator data to real road driving	179
8.6.2	Classification of drop-outs under real road driving conditions	182
8.7	Feature dimension reduction	183
8.7.1	Sequential floating forward selection	184
8.7.2	Margin influence analysis	186
8.7.3	Correlation-based feature selection	187
9	Summary, conclusion and future work	191
9.1	Summary and conclusion	191
9.2	Perspective of future work	194
A	Derivation of sawtooth occurrence frequency during curve negotiation	197
B	Description of a boxplot	199
C	<i>k</i>-means clustering	201
D	Statistical test	203
D.1	Paired-sample <i>t</i> -test	203
D.2	Normal distribution test: Lilliefors test	204
D.3	Test of significance for the Pearson correlation coefficient	204
D.4	Comparison of two Pearson correlation coefficients	205
D.5	One-way repeated measures ANOVA	205
D.6	Homogeneity of variance: Levene's test	207
D.7	Wilcoxon signed-rank test	208
D.8	Pearson's chi-square test	209
E	Mother wavelets	211
F	Additional results	213
F.1	Analysis of statistical measures of features	213
F.2	Boxplot of drive time-based features versus KSS values	214
F.3	Correlation between features using the Spearman's rank correlation coefficient	254
G	Gradient descent approach for training the ANN	257
H	On the understanding of the dual form of the optimization problem	259
H.1	Karush-Kuhn-Tucker theorem	259
H.2	Extraction of the dual problem for the soft margin classifier	260

List of figures	263
List of tables	271
Bibliography	273

Notation and abbreviations

Notations

x	scalar
\mathbf{x}	column vector
\mathbf{X}	matrix
X	space
\mathbb{Z}, \mathbb{R}	set of integer and real numbers

Mathematical operations

$ x $	absolute value of scalar
$\ \mathbf{x}\ $	2-norm of a vector
x^*	complex conjugate
$\mathbf{x}^T, \mathbf{X}^T$	transpose of vector or matrix
$\mathbf{A} \oplus \mathbf{B}$	direct sum of the vector spaces \mathbf{A} and \mathbf{B}
$\text{std}(\mathbf{x})$	standard deviation of variable \mathbf{x}
$\text{mean}(\mathbf{x})$	average of variable \mathbf{x}
$\text{cov}(\mathbf{x})$	covariance matrix of random vector \mathbf{x}
\in	is element of
$\mathcal{F}(.)$	discrete Fourier transform
$\mathcal{W}(.)$	continuous wavelet transform
$\text{rank}(\mathbf{x})_i$	the i -th rank of \mathbf{x}

Symbols

a	scale of the wavelet transform
A	amplitude of a correctly detected blink
A_1	closing amplitude of a blink
A_2	opening amplitude of a blink
amp	amplitude of detected potential blinks
amp_{em}	amplitude of a detected eye movement but not a fast blink
b	translation of the wavelet transform
b_i	bias term of the i -th hyperplane
c	true class label
\hat{c}	estimated class label

c_j	approximation coefficient of the DWT at stage j
C	margin parameter
C_{opt}	optimal margin parameter
C_+	margin parameter of the majority class
C_-	margin parameter of the minority class
d_{x_i, x_j}	distance between sample x_i and x_j in the feature space
d_j	detail coefficient of the DWT at stage j
D	dimension of the feature matrix
\check{D}	desired dimension of the feature matrix
E	blink energy
f	sampling frequency
F	blink frequency
\mathbf{F}	feature matrix
\mathbb{F}	feature space
$H(n)$	horizontal EOG signal
\mathcal{H}	a separating hyperplane in the features space
J	cost function
k_{CFS}	desired dimension of the feature matrix for CFS method
$K(\mathbf{x}_i, \mathbf{x}_j)$	kernel function
$L_{\text{maha}}(\mathbf{x}, \mathbf{y})$	Mahalanobis distance between \mathbf{x} and \mathbf{y}
$L_p(\mathbf{x}, \mathbf{y})$	metric of Minkowski for \mathbf{x} and \mathbf{y}
L_{win}	window length of the STFT
\mathcal{L}	Lagrangian function
\mathcal{L}_d	dual form of the Lagrangian function
m	number of classes
$M_{i,j}$	element of the i -th row and j -th column of the confusion matrix
\mathbf{M}	confusion matrix
$M_{\mathcal{R}}$	evaluation metric of the CFS method
n	sample
N	number of observations or samples in the training set
N_+	number of samples of the majority class
N_-	number of samples of the minority class
N_h	number of neurons in a hidden layer
N_μ	window size of the EWMA
N_{σ^2}	window size of the EWMVAR
N_{SMOTE}	number of samples to be added to the minority class in SMOTE
r	curve radius
\mathcal{R}	subset of features with k_{CFS} number of features
\mathcal{S}	total data set
$\mathcal{S}_{\text{validation}}$	validation set of the cross-validation
$\mathcal{S}_{\text{train}}$	training set of the cross-validation
s	number of subjects
\mathbf{S}	covariance matrix
t	time
t_0	statistic value of the paired-sample t-test
t_{baseline}	time interval for feature baselining
T	blink duration

T_c	closing duration of a blink
T_o	opening duration of a blink
T_{cl}	duration of a the closed phase of a blink
T_{ro}	delay of reopening of a blink
T_{50}	duration from 50% of the rise amplitude to 50% of the fall amplitude
T_{80}	duration from 80% of the rise amplitude to 80% of the fall amplitude
T_{90}	duration from 90% of the rise amplitude to 90% of the fall amplitude
$th_{k\text{-means}}$	amplitude threshold for distinguishing between clusters of the k -means method
th_{denosing}	amplitude threshold for noise removal by the DWT
th_{vel}	amplitude threshold regarding the amplitude of the blink velocity
th_s	amplitude threshold for distinguishing between saccades and long eye closures
$V(n)$	vertical EOG signal
$V'(n)$	first derivative of $V(n)$
$V_{\text{med}}(n)$	median filter processed $V(n)$
$\tilde{V}_1(n)$	denoised $V(n)$ by the DWT
$\check{V}(n)$	$V(n)$ after drift removal by the DWT
\mathbf{x}_i	i -th feature vector
\mathbf{w}	weight vector
w_{med}	window size of the median filter
z_0	statistic value of the Wilcoxon signed-rank test
α	confident level of a statistical test
α_i	Lagrange multiplier for \mathbf{x}_i
$\boldsymbol{\alpha}$	Lagrange multiplier
γ_i	geometrical margin of the i -th hyperplane in SVM
γ	parameter for a radial basis function kernel
γ_{opt}	optimal parameter for a radial basis function kernel
δ	angular displacement of the eyes
ζ	random value for adding a new sample in SMOTE
ϑ	sensitivity of the ANN
λ	number of vanishing moment for a wavelet
λ_μ	forgetting factor of the EWMA
λ_{σ^2}	forgetting factor of the EWMVAR
μ	mean value
ξ_i	slack variable associated with \mathbf{x}_i
ρ_p	Pearson product-moment correlation coefficient
ρ_s	Spearman's rank correlation coefficient
ε	residual
η	learning rate of the ANN
σ	standard deviation value
$\widetilde{\mu}_n$	exponentially weighted moving average at sample n
$\widetilde{\sigma}^2_n$	exponentially weighted moving variance at sample n
$\phi(t)$	wavelet scaling function
$\Phi(\mathbf{x})$	mapping function from \mathbf{x} to the feature space
$\psi(t)$	wavelet function
\mathcal{X}	DFT of signal x
$\mathcal{X}_\psi(a, b)$	continuous wavelet transform of x for scale a and translation b

Vehicle dynamics and sensors

Γ	displacement of the vehicle
r	curve radius
w	wheel speed
ψ	yaw angle
κ	road curvature
v	vehicle velocity
$\dot{\psi}$	yaw angle rate

Abbreviations

ACC	Accuracy
ACV	Average closing velocity
ADR	Average detection rate
ANN	Artificial neural network
ANOVA	Analysis of variance
AOV	Average opening velocity
ASR	Alpha spindle rate
CFS	Correlation based feature selection
CWT	Continuous wavelet transform
DAC	Driver alert control
DCA	Driver cursory attention
DDA	Driver diverted attention
DFT	Discrete Fourier transform
DMPA	Driver misprioritised attention
DNA	Driver neglected attention
DR	Detection rate
DRA	Driver restricted attention
DWT	Discrete wavelet transform
ECG	Electrocardiography
EEG	Electroencephalography
EOG	Electrooculography
ER	Error rate
EWMA	Exponentially weighted moving average
EWMVAR	Exponentially weighted moving variance
ESS	Epworth sleepiness scale
FDR	False detection rate
FNR	False negative rate
FPR	False positive rate
FWHM	Full width at half maximum
GPS	Global positioning system
GSRD	Generalization of the simulator data to real driving
HRV	Heart rate variability
IR	Infra red

KKT	Karush-Kuhn-Tucker theorem
<i>k</i> -NN	<i>k</i> -nearest neighbors
KSS	Karolinska sleepiness scale
MIA	Margin influence analysis
MCS	Monte Carlo sampling
MCV	Maximum closing velocity
MOV	Maximum opening velocity
NHTSA	National highway traffic safety administration
OKN	Optokinetic nystagmus
PERCLOS	Percentage of eye closure
PC	Precision
PSD	Power spectral density
RBF	Radial basis function
RC	Recall
REM	Rapid eye movement
SDLP	Standard deviation of lateral position
SEM	Slow eye movement
SMOTE	Synthetic minority oversampling technique
SNR	Signal-to-noise ratio
SBS	Sequential backward selection
SFFS	Sequential floating forward selection
SFS	Sequential forward selection
SSS	Stanford sleepiness scale
STFT	Short-time Fourier transform
SVM	Support vector machine
SWT	Stationary wavelet transform
TLC	Time-to-line crossing
TP	Tangent point
VEZ	Virtual edge zone
VOR	Vestibulo-ocular reflex
WT	Wavelet transform

Zusammenfassung

Die Zunahme müdigkeitsbedingter Verkehrsunfälle in den letzten Jahren verdeutlicht die Notwendigkeit anhand geeigneter Referenzmaße Assistenzsysteme zur Erkennung von Müdigkeit zu entwickeln. Das Ziel der vorliegenden Arbeit ist daher die Klassifikation des Fahrerzustandes basierend auf den Augenbewegungen anhand von Elektrookulographie (EOG).

Um einen Einblick in die Zustände des Fahrens zu geben, die zu sicherheitskritischen Verkehrssituationen führen, werden zunächst die Konzepte von Fahrermüdigkeit und -ablenkung sowie verschiedene damit zusammenhängende Terminologien beschrieben. Anschließend werden Techniken betrachtet, um den Fahrer wach zu halten und somit Autounfälle zu verhindern. Da diese Techniken keine lang anhaltende Wirkung auf die Wachheit des Fahrers zeigen, sind intelligente Systeme zur Erkennung der Müdigkeit des Fahrers notwendig. In der Vergangenheit sind derartige Systeme bereits entwickelt worden, von denen einige in dieser Arbeit vorgestellt werden.

Wie auch in früheren Studien festgestellt wurde, ist der Fahrerzustand durch objektive und subjektive Maße quantifizierbar. Zur Erfassung objektiver Maße wird der Fahrer entweder direkt oder indirekt überwacht. Bei der indirekten Überwachung des Fahrers werden Messgrößen verwendet, die die Fahrleistung des Fahrers widerspiegeln, wie zum Beispiel die Spurhaltung oder Lenkradbewegungen. Im Gegensatz dazu umfasst die direkte Überwachung hauptsächlich physiologische Messgrößen wie Hirnaktivität, Herzfrequenz und Augenbewegungen. Um diese objektiven Messgrößen beurteilen zu können, sind subjektive Messgrößen wie eine Eigenbewertung durch den Fahrer notwendig. Die vorliegende Arbeit stellt diese Messgrößen vor und diskutiert die Bedenken hinsichtlich ihrer Interpretation und Zuverlässigkeit.

Die auf dem Markt existierenden Müdigkeitsassistenzsysteme stützen sich alle auf fahrleistungs-basierte Maße. Diese setzen voraus, dass das Fahrzeug ausschließlich durch den Fahrer selbst gelenkt wird. Solange andere Assistenzsysteme mit dem Ziel das Fahrzeug in der Mitte der Fahrbahn zu halten aktiviert sind, würden Maßzahlen zur Fahrleistung falsche Entscheidungen hinsichtlich einer Warnung treffen. Der Grund dafür ist, dass die Sensoren eine Kombination aus dem Verhalten des Fahrers und des aktivierten Assistenzsystems messen. Das Müdigkeitswarnsystem kann den Beitrag des Fahrers an der Fahraufgabe nicht bestimmen. Dies unterstreicht die Notwendigkeit einer direkten Fahrerüberwachung.

Frühere Arbeiten haben als Indikator für Müdigkeit den Abfall der Alpha Spindelrate (ASR) eingeführt. Hierbei handelt es sich um ein Merkmal, das aus den Hirnaktivitätssignalen während der direkten Beobachtung des Fahrers extrahiert wird. Es wurde gezeigt, dass anhand ASR Ablenkung des Fahrers erfasst werden kann und insbesondere eine visuelle Ablenkung dabei einen entgegenwirkenden Effekt hat. Basierend auf den Augenbewegungen des Fahrers wurde ein Algorithmus entwickelt, um die negative Auswirkung der visuellen Ablenkung des Fahrers auf die ASR zu reduzieren. Der Zusammenhang von ASR und der Müdigkeit des Fahrers kann dadurch teilweise verbessert werden.

Da der Fokus dieser Arbeit auf den Augenbewegungen des Fahrers liegt, wird das visuelle System des Menschen vorgestellt und die Idee des „Was“ und „Wo“ beschrieben, um visuelle Aufmerk-

samkeit zu definieren. Desweiteren wird die Struktur des menschlichen Auges beschrieben und die relevanten Arten von Augenbewegungen während der Fahrt definiert. Außerdem werden Augenbewegungen in zwei Gruppen mit langsamen und schnellen Augenbewegungen kategorisiert. Es wird gezeigt, dass Lidschläge je nach Wachsamkeit des Fahrers zu beiden dieser Gruppen gehören können.

EOG als ein Werkzeug zur Messung der Augenbewegungen ermöglicht es uns zwischen müdigkeits- bzw. ablenkungsbedingten Augenbewegungen und Augenbewegungen aufgrund der Fahrsituation zu unterscheiden. Aufgrund dessen wurde im Rahmen einer Pilotstudie ein Experiment unter vollständig kontrollierten Bedingungen auf einem Testgelände durchgeführt, um den Zusammenhang zwischen den Augenbewegungen des Fahrers und verschiedenen realen Fahrszenarien zu untersuchen. In diesem Experiment sind unerwünschte Kopfschwingungen in den EOG-Signalen und das Sägezahnmuster der Augen (optokinetischer Nystagmus, OKN) als situationsabhängige Augenbewegungen erkannt worden. Kopfschwingungen treten aufgrund von Bodenanregungen auf, wohingegen OKN in Kurven mit kleinen Radien (50m) vorkam. Die statistische Untersuchung zeigt eine signifikante Veränderung in den EOG-Signalen durch unerwünschte Kopfschwingungen. Darüber hinaus wird ein analytisches Modell entwickelt, um den möglichen Zusammenhang zwischen OKN und den Tangentenpunkt der Kurve zu erklären. Das entwickelte Modell wird mit realen Daten einer Strecke mit hohen Krümmungen validiert.

Um alle relevanten Muster von Augenbewegungen während wacher Fahrten und Fahrten unter Müdigkeit zu erfassen, werden in dieser Arbeit verschiedene Experimente —inklusive Tag- und Nachtversuche— sowohl unter realen als auch simulierten Fahrbedingungen durchgeführt.

Basierend auf den in den Experimenten erhobenen Signalen werden verschiedene Ansätze zur Detektion der Augenbewegungen untersucht. Zunächst wird die Detektion von Lidschlägen basierend auf der Medianfilterung beleuchtet und ihre Nachteile bei der Erkennung langsamer Lidschläge und Sakkaden aufgezeigt. Danach wird ein adaptiver Erkennungsansatz basierend auf der Ableitung der EOG-Signale vorgeschlagen, welcher nicht nur Lidschläge, sondern auch andere fahrrelevante Augenbewegungen wie Sakkaden und Sekundenschlafereignisse erkennt. Der vorgeschlagene Algorithmus unterscheidet darüber hinaus, zwischen den häufig verwechselten fahrrelevanten Sakkaden und einer verringerten Lidschlagamplitude eines müden Fahrers, obwohl Müdigkeit die Augenbewegungsmuster beeinflusst. Die Auswertung der Ergebnisse zeigt, dass der dargestellte Algorithmus die bekannte Medianfilterungsmethode übertrifft, so dass schnelle Augenbewegungen während beider Phasen, wach und müde, korrekt erkannt werden.

Weiter befasst sich die vorliegende Arbeit mit der Erkennung der langsamen Lidschläge als typische Muster bei Müdigkeit durch die Anwendung einer kontinuierlichen Wavelet-Transformation auf EOG-Signale. In dem vorgeschlagenen Algorithmus werden die schnellen und langsamen Lidschläge durch die Einstellung der Parameter der Wavelet-Transformation gleichzeitig detektiert. Allerdings führt dieser Ansatz zu einer größeren Falscherkennungsrate im Vergleich zu der auf der Ableitung basierenden Methode. Deshalb wird in dieser Arbeit für die Lidschlagerkennung eine Kombination beider Verfahren angewandt. Um die Qualität der gesammelten EOG-Signale zu verbessern und Rauschen und Drift zu entfernen, wird die diskrete Wavelet-Transformation genutzt. Für die Rauschunterdrückung wird eine adaptive Schwellenstrategie für die diskrete Wavelet-Transformation vorgeschlagen.

Frühere Forschungsarbeiten haben gezeigt, dass die lidschlagsbasierte Merkmale des Fahrers (Lidschlagfrequenz, -dauer, usw.) zu einem gewissen Grad mit Müdigkeit korreliert sind. Daraus können diese —mit einer gewissen Unsicherheit— einen Beitrag zu Müdigkeitwarnsystemen leisten. Um diese Systeme zu verbessern, werden Eigenschaften der detektierten Lidschläge

bezüglich ihrer unterschiedlichen Herkunft untersucht. Im Rahmen eines Experimentes unter realen Straßenbedingungen zeigte sich, dass Lidschläge sowohl spontan als auch aufgrund von Blickwechseln auftreten. Die Blickwechsel zwischen festen Positionen, die aufgrund der visuomotorischen Nebenaufgabe eingetreten sind, induzierten und modulierten das Auftreten von Lidschlägen. Die Ergebnisse eines weiteren Fahrimulatorversuchs ohne Nebenaufgabe zeigen, dass die Menge der Blickwechsel (zwischen verschiedenen Positionen) mit der Wahrscheinlichkeit des Lidschlagsauftretens positiv korreliert. Aus diesem Grund wird für Müdigkeitwarnsysteme, die sich ausschließlich auf die Änderung der Lidschlagfrequenz stützen, empfohlen, durch Blickwechsel (z.B. während visueller Ablenkung) induzierte Lidschläge anders als spontan auftretende Lidschläge zu behandeln.

Nach der Analyse der Abhängigkeiten von Lidschlagfrequenz und Blickwechseln, wurden aus jedem erkannten Lidschlag von 43 Probanden, gesammelt unter simulierten und realen Fahrbedingungen, während 67 Stunden von Tages- und Nachtfahrten 19 Merkmale extrahiert. Dies entspricht der größten Anzahl von lidschlagsbasierten Merkmalen und der größten Anzahl an Probanden im Vergleich zu früheren Studien. Es werden zwei Ansätze zur Aggregation von Merkmalen vorgestellt, um ihren Zusammenhang mit der sich langsam entwickelnden Müdigkeit zu verbessern. Im ersten Ansatz wurden ausschließlich die Teile der gesammelten Daten untersucht, die am besten mit der subjektiven Selbstbewertung der Fahrer durch die Karolinska Sleepiness Scale korrelierten. Im zweiten Ansatz werden die gesamten Daten mit der maximalen Menge an Informationen zur Fahrermüdigkeit untersucht. Bei beiden Ansätzen wird die Abhängigkeit zwischen den einzelnen Merkmalen und der Müdigkeit statistisch unter Berechnung von Korrelationskoeffizienten betrachtet. Die Ergebnisse zeigen, dass sich die Müdigkeitsabhängigkeit der Merkmale in einem hohen Grad nicht-linear entwickelt. Darüber hinaus zeigte sich, dass für einige Merkmale bei unterschiedlichen Probanden verschiedene müdigkeitsabhängige Verläufe möglich sind. Daher stellen wir Warnsysteme in Frage, die sich nur auf ein einziges Merkmal für ihre Entscheidungsstrategie verlassen, und betonen, dass sie anfällig für hohe Falschalarmraten sind.

Um zu untersuchen, ob ein einzelnes Merkmal für die Vorhersage von sicherheitskritischen Ereignissen geeignet ist, untersuchen wir die Veränderung der Merkmale für alle Probanden kurz vor dem Auftreten des ersten unbeabsichtigten Verlassens der Fahrspur und des ersten unbeabsichtigten Sekundenschlafs im Vergleich zum Beginn der Fahrt. Basierend auf statistischen Tests ändern sich vor dem Verlassen der Fahrspur die meisten Merkmale signifikant. Daher rechtfertigen wir die Rolle der lidschlagsbasierten Merkmale für die frühe Fahrermüdigitserkennung. Dies gilt jedoch nicht für die Variation der Merkmale vor Sekundenschlaf.

Weiter wurden alle 19 augenbewegungsbasierte Merkmale gleichzeitig näher betrachtet. Der Fahrerzustand wurde dabei durch künstliche neuronale Netzwerke, Support-Vektor-Maschinen und k -nearest neighbour Klassifikatoren beurteilt, sowohl für binäre als auch Multi-Class-Fälle. Die binären Klassifikatoren sind sowohl fahrerunabhängig als auch fahrerabhängig trainiert worden, um die Generalisierungsaspekte der Ergebnisse für ungewohnte Daten zu adressieren. Die binäre Fahrerzustandsklassifikation (wach vs. müde) basierend auf Augenbewegungsmerkmalen ergab eine durchschnittliche Erkennungsrate von 83% für jeden Klassifikator. Für eine dreistufige Klassifizierung (wach vs. mittel vs. müde), betrug die Erkennungsrate lediglich 67%, möglicherweise aufgrund ungenauer Selbstbewertungen des Vigilanzzustandes. Darüber hinaus wurde die Frage der unausgeglichenen Daten mit klassifikatorabhängigen und -unabhängigen Ansätzen betrachtet. Wir zeigen, dass es für eine zuverlässige Fahrerzustandsklassifikation entscheidend ist, Ereignisse von beiden Phasen —wach und müde— in ausgeglichener Weise zu berücksichtigen. Grund hierfür ist, dass die vorgeschlagenen Lösungen früherer Untersuchungen zum Umgang

mit unausgeglichenen Datensätzen die Klassifikatoren nicht verallgemeinern, sondern zu ihrer Überanpassung führen.

Der Nachteil von Fahrversuchen in Fahrsimulatoren im Vergleich zu realen Fahrversuchen wird ebenfalls dargestellt. Zu diesem Zweck wird zunächst eine Reduktion der Daten vorgenommen. Weiter wenden wir die von uns trainierten Klassifikatoren auf nicht in den Trainingsdaten enthaltenen Daten müder Fahrer —gesammelt unter realen Fahrbedingungen— an, um zu untersuchen, ob die Müdigkeit in Fahrsimulatoren repräsentativ für die Müdigkeit unter realen Fahrbedingungen ist. Mit einer durchschnittlichen Erkennungsrate von über 68% für alle Klassifikatoren, kann von der Vergleichbarkeit beider Experimentalumgebungen ausgegangen werden.

Schließlich werden Ansätze zur Dimensionsreduzierung der Merkmale diskutiert, um die Fahrzeugtauglichkeit der extrahierten Merkmale zu bestimmen. Aus diesem Grund wurden Filter und Wrapper Ansätze eingeführt und miteinander verglichen. Unsere Ergebnisse zeigen, dass die Wrapper-Ansätze die Filter-basierten Methoden übertreffen.

Abstract

The increase in vehicle accidents due to the driver drowsiness over the last years highlights the need for developing reliable drowsiness assistant systems by a reference drowsiness measure. Therefore, the thesis at hand is aimed at classifying the driver vigilance state based on eye movements using electrooculography (EOG).

In order to give an insight into the states of driving, which lead to critical safety situations, first, driver drowsiness, distraction and different terminologies in this context are described. Afterwards, countermeasures as techniques for keeping a driver awake and consequently preventing car crashes are reviewed. Since countermeasures do not have a long-lasting effect on the driver vigilance, intelligent driver drowsiness detection systems are needed. In the recent past, such systems have been developed on the market, some of which are introduced in this study.

As also stated in previous studies, driver state is quantifiable by objective and subjective measures. The objective measures monitor the driver either directly or indirectly. For indirect monitoring of the driver, one uses the driving performance measures such as the lane keeping behavior or steering wheel movements. On the contrary, direct monitoring mainly comprises the driver's physiological measures such as the brain activities, heart rate and eye movements. In order to assess these objective measures, subjective measures such as self-rating scores are required. This study introduces these measures and discusses the concerns about their interpretation and reliability.

The developed drowsiness assistant systems on the market are all based on driving performance measures. These measures presuppose that the vehicle is steered solely by the driver himself. As long as other assistance systems with the concept to keep the vehicle in the middle of the lane are activated, driving performance measures would make wrong decisions about warnings. The reason is what the sensors measure is a combination of the driver's behavior and the activated assistance system. In fact, the drowsiness warning system cannot determine the contribution of the driver in the driving task. This underscores the need for the direct monitoring of the driver.

Previous works have introduced the drop of the alpha spindle rate (ASR) as a drowsiness indicator. This rate is a feature extracted out of the brain activity signals during the direct monitoring the driver. Additionally, ASR was shown to be sensitive to driver distraction, especially a visual one with an counteracting effect. We develop an algorithm based on eye movements to reduce the negative effect of the driver visual distraction on the ASR. This helps to partially improve the association of ASR with the driver drowsiness.

Since the focus of this study is on driver eye movements, we introduce the human visual system and describe the idea of *what* and *where* to define the visual attention. Further, the structure of the human eye and relevant types of eye movements during driving are defined. We also categorize eye movements into two groups of slow and fast eye movements. We show that blinks, in principle, can belong to both of these groups depending on the driver's vigilance state.

EOG as a tool to measure the driver eye movements allows us to distinguish between drowsiness- or distraction-related and driving situation dependent eye movements. Thus, in a pilot study,

an experiment under fully controlled conditions is carried out on a proving ground to investigate the relationship between driver eye movements and different real driving scenarios. In this experiment, unwanted head vibrations within EOG signals and the sawtooth pattern (optokinetic nystagmus, OKN) of eyes are realized as situation dependent eye movements. The former occurs due to ground excitation and the latter happens during small radius (50 m) curve negotiation. The statistical investigation expresses a significant variation of EOG due to unwanted head vibrations. Moreover, an analytical model is developed to explain the possible relationship of OKN and tangent point of the curve. The developed model is validated against the real data on a high curvature track.

In order to cover all relevant eye movement patterns during awake and drowsy driving, different experiments are conducted in this work including daytime and nighttime experiments under real road and simulated driving conditions.

Based on the measured signals in the experiments, we study different eye movement detection approaches. We, first, investigate the conventional blink detection method based on the median filtering and show its drawback in detecting slow blinks and saccades. Afterwards, an adaptive detection approach is proposed based on the derivative of the EOG signal to simultaneously detect not only the eye blinks, but also other driving-relevant eye movements such as saccades and microsleep events. Moreover, in spite of the fact that drowsiness influences eye movement patterns, the proposed algorithm distinguishes between the often confused driving-related saccades and decreased amplitude blinks of a drowsy driver. The evaluation of results shows that the presented detection algorithm outperforms the common method based on median filtering so that fast eye movements are detected correctly during both awake and drowsy phases.

Further, we address the detection of slower eye blinks, which are referred to as typical patterns of the drowsiness, by applying the continuous wavelet transform to EOG signals. In our proposed algorithm, by adjusting parameters of the wavelet transform, fast and slow blinks are detected simultaneously. However, this approach suffers from a larger false detection rate in comparison to the derivative-based method. As a result, for blink detection in this work, a combination of these two methods is applied. To improve the quality of the collected EOG signals, the discrete wavelet transform is benefited to remove noise and drift. For the noise removal, an adaptive thresholding strategy within the discrete wavelet transform is proposed which avoids sacrificing noise removal for saving blink amplitude or vice versa.

In previous research, driver eye blink features (blink frequency, duration, etc.) have shown to be correlated to some extent with drowsiness. Hence, within a level of uncertainty they can contribute to driver drowsiness warning systems. In order to improve such systems, we investigate characteristics of detected blinks with respect to their different origins. We observed that in a real road experiment, blinks occur both spontaneously or due to gaze shifts. Gaze shifts between fixed positions, which occurred due to secondary visuomotor task, induced and modulated the occurrence of blinks. Moreover, the direction of the gaze shifts affected the occurrence of such blinks. Based on the eye movements during another experiment in a driving simulator without a secondary task, we found that the amount of gaze shifts (between various positions) is positively correlated with the probability of the blink occurrence. Therefore, we recommend handling gaze shift-induced blinks (e.g. during visual distraction) differently from those occurring spontaneously in drowsiness warning systems that rely solely on the variation of blink frequency as a driver state indicator.

After studying dependencies between blink occurrence and gaze shifts, we extract 19 features out of each detected blink event of 43 subjects collected under both simulated and real driving

conditions during 67 hours of both daytime and nighttime driving. This corresponds to the largest number of extracted eye blink features and the largest number of subjects among previous studies. We propose two approaches for aggregating features to improve their association with the slowly evolving drowsiness. In the first approach, we solely investigate parts of the collected data which are best correlated with the subjective self-rating score, i.e. Karolinska Sleepiness Scale. In the second approach, however, the entire data set with the maximum amount of information regarding driver drowsiness is scrutinized. For both approaches, the dependency between single features and drowsiness is studied statistically using correlation coefficients. The results show that the drowsiness dependency to features evolves to a larger extent non-linearly rather than linearly. Moreover, we show that for some features, different trends with respect to drowsiness are possible among different subjects. Consequently, we challenge warning systems which rely only on a single feature for their decision strategy and underscore that they are prone to high false alarm rates.

In order to study whether a single feature is suitable for predicting safety-critical events, we study the overall variation of the features for all subjects shortly before the occurrence of the first unintentional lane departure and first unintentional microsleep in comparison to the beginning of the drive. Based on statistical tests, before the lane departure, most of the features change significantly. Therefore, we justify the role of blink features for the early driver drowsiness detection. However, this is not valid for the variation of features before the microsleep.

We also focus on all 19 blink-based features together as one set. We assess the driver state by artificial neural network, support vector machine and k -nearest neighbors classifiers for both binary and multi-class cases. There, binary classifiers are trained both subject-independent and subject-dependent to address the generalization aspects of the results for unseen data. For the binary driver state prediction (awake vs. drowsy) using blink features, we have attained an average detection rate of 83% for each classifier separately. For 3-class classification (awake vs. medium vs. drowsy), however, the result was only 67%, possibly due to inaccurate self-rated vigilance states. Moreover, the issue of imbalanced data is addressed using classifier-dependent and classifier-independent approaches. We show that for reliable driver state classification, it is crucial to have events of both awake and drowsy phases in the data set in a balanced manner. The reason is that the proposed solutions in previous researches to deal with imbalanced data sets do not generalize the classifiers, but lead to their overfitting.

The drawback of driving simulators in comparison to real driving is also discussed and to this end we perform a data reduction approach as a first remedy. As the second approach, we apply our trained classifiers to unseen drowsy data collected under real driving condition to investigate whether the drowsiness in driving simulators is representative of the drowsiness under real road conditions. With an average detection rate of about 68% for all classifiers, we conclude their similarity.

Finally, we discuss feature dimension reduction approaches to determine the applicability of extracted features for in-vehicle warning systems. On this account, filter and wrapper approaches are introduced and compared with each other. Our comparison results show that wrapper approaches outperform the filter-based methods.

1. Introduction

1.1. Problem statement and motivation

The number of injured and killed persons in traffic accidents of Germany based on the statistics provided by the Federal Statistical Office of Germany (DESTATIS, 2013a) has slightly decreased after 1970 (see Figure 1.1). This decrease is the result of introducing different safety regulations in the course of time. Despite the overall drop of both numbers after 1970, the fact of killing over 3000 persons along with 370000 person injuries in 2013 should not be neglected.

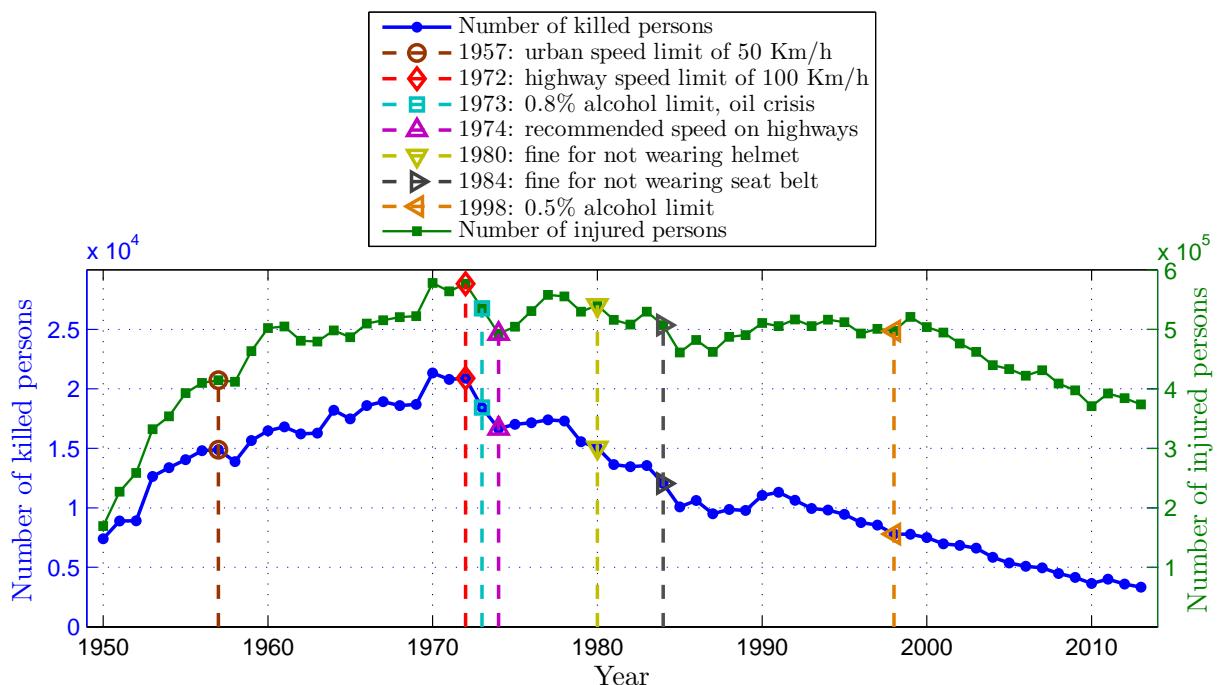


Figure 1.1.: The evolution of number of killed and injured persons in traffic accidents of Germany (DESTATIS, 2013a)

Car crashes, in general, occur because of different reasons such as driver drowsiness, distraction, bad weather condition, high speed, consuming alcohol, etc. Among these reasons, in Germany, sleepy drivers are responsible for about 25% of car crashes on highways (Zulley and Popp, 2012). In addition, one out of every six heavy road accidents with truck involvement is caused by a drowsy truck driver. Apart from that, an increase of 6% in the number of vehicle accidents involving person injuries due to driver drowsiness between 2008 and 2012 (see Figure 1.2) reveals that drowsiness has huge contribution in car accidents.

Based on 14268 crashes in the United States from 2009 to 2013 (Tefft, 2014), a drowsy driver was involved in the accident percentages with the following consequences:

- the vehicle was towed away from the scene: 6%

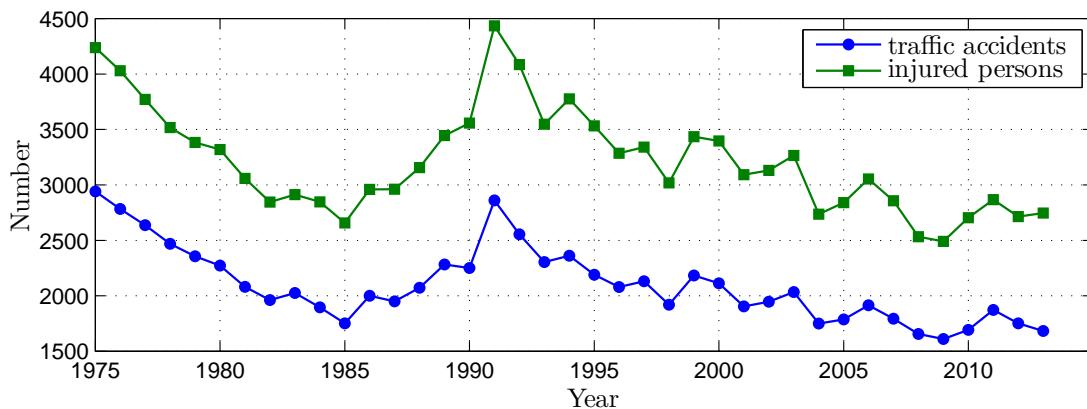


Figure 1.2.: The evolution of number of vehicle accidents due to driver drowsiness and the number of injured persons involved in them (DESTATIS, 2013b)

- a person was injured: 20%
- a person was killed: 21%.

It is also mentioned by Brown et al. (2014) that in the United States, overall, more than 80000 car crashes and 850 fatalities are the result of drowsy driving every year. In addition, based on the 100-car naturalistic driving study, 22–24% of crashes and near-crashes were caused by drowsy drivers (Klauer et al., 2006). According to the National Highway Traffic Safety Administration (NHTSA), young people, shift workers and people with sleep disorder diseases are at very high risk of car crashes due to the drowsiness (NHTSA, 1998). In the mentioned study, “*slower reaction time*”, “*reduced vigilance*” and “*impaired information process*” were noted as the consequences of drowsiness.

The aforementioned statistics underscore the necessity for developing drowsiness warning systems. In the ideal case, such assistance systems observe one or several drowsiness related measures and warn the driver in order to prevent accidents. This is done either based on driving performance measures such as the steering or lane-keeping behavior or based on the direct monitoring of the driver by means of physiological measures like eye blinks, yawning frequency, etc. Since adding some sensors to the vehicle for monitoring the lane keeping and steering wheel movement behaviors of the driver were affordable and technologically much easier to develop, initially, most of the commercial products by car manufacturers and supplier companies were all based on driving performance measures.

Along with the development of such driver drowsiness detection systems, the intelligent vehicle technologies have also grown in the field of driver assistance rapidly to prevent car crashes by means of other approaches. As an example, the *DISTRONIC PLUS with Steering Assist* of Mercedes-Benz (Daimler AG, 2014b) supports the driver to stay within the lane. In other words, the car *steers* to some extent together with the driver. Such a system, however, would not prevent driver drowsiness, and on the contrary, it may lead to insurmountable problem for the drowsiness warning system, if it is activated. The reason is that the resulting driving behavior is a combination of the driver actions and the supporting system. Since the warning system is unable to determine to what extent the driver is contributing, it fails to detect a drowsy driver. Therefore, the new assistance systems deteriorate the performance of the classical warning systems. To be emphasized again, many of the driver assistance systems help the driver to improve the driving performance and to avoid severe crashes, but they are unable to eliminate crashes to the full extent. This fact highlights the need of new driver drowsiness detection

systems even for cars equipped with a variety of assistance systems. A possibility to deal with this problem is to directly observe the driver. Hence, one can think of a driver observation camera.

The mentioned problem will even get worst, as soon as autonomous cars are developed in near future. Such cars *drive* (steer, brake or turn) independently and relieve the driver of the full concentration on the driving task. However, at the moment that the car is unable to interpret the surrounding information correctly and a crash deems unavoidable, the driver should undertake the driving task. Therefore, the car should inform the driver in a timely manner with respect to driver's level of attention. Clearly, a driver, who is distracted, should be informed earlier than a driver who observes the scene ahead carefully. Driver's level of vigilance and distraction cannot be determined based on the lane keeping and steering wheel movement behaviors, because the car itself is responsible for them. Therefore, an autonomous system should directly observe the driver the whole time as well, in order to assess driver's vigilance and attentional level. Again, one can think of a camera-based driver monitoring system.

A driver observation camera monitors one or several driver's physiological measures such as repeated yawning, slower reactions, difficulty in keeping eyes open, etc. (Dong et al., 2011). Prior to employing a camera or developing an eye tracking algorithm, it should, however, be ascertained to what extent features of the corresponding biological measures reflect the driver state, especially under real driving conditions which is the ultimate goal of all assistance systems. In other words, a reliable reference drowsiness measure is needed whose development is beneficial for evaluating any drowsiness detection system. On this account, this work concentrates on driver eye movements analysis based on a reference measuring system.

Before introducing different approaches for monitoring the driver and assessing the vigilance level, we must first define what exactly we are trying to measure and to quantify. Therefore, next section primarily concerns definition of drowsiness and inattention.

1.2. Definition of drowsiness and inattention

The states of driving, which lead to critical safety situations, have been described and distinguished by a variety of terminologies such as distraction, inattention, fatigue, exhaustion, sleepiness, and drowsiness. In addition, the proper states of driving are also referred to as awareness, vigilance and alertness. Therefore, prior to determining our terminology in this work, first we define and discuss the aforementioned terms.

Fatigue and sleepiness

Schmidt et al. (2011) stated that the terms fatigue and sleepiness are usually used as synonyms, although they are not identical. According to Hirshkowitz (2013), depending on the field of study, the term fatigue might have different meanings. In civil engineering, as an example, it is defined as "*a weakening or material breakdown over time produced by repeated exposure to stressors*". Human physical fatigue, however, refers to the "*weakness from repeated exertion or a decreased response of cells, tissues, or organs after excessive stimulation or activity*". Considering this definition, human physical fatigue is not necessarily associated with sleepiness, but improves after resting. As a result, it is the mental fatigue which is highly correlated with sleepiness and car crashes and is related to the context of driving (May, 2011). In general, Hirshkowitz (2013)

defined fatigue as “*a sense of tiredness, exhaustion, or lack of energy*” which is intensified due to stress load or in the course of time, also called *time-on-task*.

Based on Williamson et al. (2014), fatigue as a comprehensive terminology covers not only sleepiness and mental fatigue, but also fatigue due to illness. Therefore, they associated sleepiness only to effects such as *time since awaking* and *time of the day effects*, while fatigue might occur due to both duration and work-loading of a task and also sleepiness or illness factors. This is in agreement with Philip et al. (2005) who also defined sleepiness as “*difficulty in remaining awake, which disappears after sleep, but not after rest.*”

Driver fatigue, as defined by May (2011), is the demotivation of continuing the driving task and also sleepiness. She mentioned following causes for fatigue which are believed to affect each other as well:

- *task-related fatigue and environmental factors* (e.g. trip duration or weather/road condition)
- *sleep-related fatigue* (e.g. quality/quantity of sleep, circadian rhythm)

Moreover, May (2011) categorized task-related fatigue as either *active fatigue* or *passive fatigue*. The former, in general, is associated with the overload of “*attentional resource*” such as in the case of performing a secondary task during driving. The latter, however, interacts with the monotonicity or familiarity with the route. It is also emphasized that time-on-task deteriorates the driving performance, if it is combined with monotonous driving.

The circadian rhythm manages the timing of alertness and sleepiness. Based on the circadian rhythm, two peaks of sleepiness can be predicted (Čolić et al., 2014). For people, who sleep at night, these peaks occur before afternoon and at night. May (2011) referred to the circadian rhythm as the internal body clock with high contribution to sleepiness and driving performance degradation. As a result, for those people, who are not synchronous to the *circadian alerting process*, sleepiness is more probable (Hirshkowitz, 2013).

Čolić et al. (2014) used the general term drowsiness as a factor which threatens road traffic safety and related it to sleepiness reasons with the following subcategories: “*sleep restriction or loss, sleep fragmentation and circadian factors*”.

Distraction and inattention

Regan et al. (2011) suggested distinguishing between driver distraction and driver inattention for a better comparison of research findings. Oxford dictionary (Oxford, 2014) defines distraction as “*a thing that prevents someone from concentrating on something else*”. Regan et al. (2011) also summarized following points for defining distraction in the driving context:

- “*There is a diversion of attention away from driving or safe driving.*”
- “*Attention is diverted toward a competing activity, inside or outside the vehicle, which may or may not be driving-related.*”
- “*The competing activity may compel or induce the driver to divert attention toward it.*”
- “*There is an implicit, or explicit, assumption that safe driving is adversely effected.*”

Inattention, on the other hand, is defined as “*lack of attention; failure to attend to one's responsibilities; negligence*” (Oxford, 2014). Hoel et al. (2010) categorized attentional dysfunction as: “*inattention, attentional competition and distraction*”. Similar to distraction, inattentive

situations also involve the interference in the driving task. Thus, inattention in terms of the driving task occurs, while performing a secondary non-driving-related task such as text messaging. On the contrary, Hoel et al. (2010) linked distraction to *personal concerns* like daydreaming. Finally, performing secondary driving-related tasks in addition to the primary driving task is considered as attentional competition (e.g. driving and navigating). Unlike Hoel et al. (2010), Wallén Warner et al. (2008) (as cited by Regan et al., 2011) decomposed inattention as:

- “*driving-related distractors inside vehicle*”: e.g. navigation system
- “*driving-related distractors outside vehicle*”: e.g. road signs
- “*non driving-related distractors inside vehicle*”: e.g. speaking to a passenger
- “*non driving-related distractors outside vehicle*”: e.g. a passenger on the pavement
- “*thoughts/daydreaming*”: e.g. personal problems.

According to Pettitt et al. (2005) (cited by Regan et al., 2011), distraction leads to inattentive driving. Inattention, however, might have other motives than only the distraction.

Regan et al. (2011) defined driver inattention as “*insufficient, or no attention, to activities critical for safe driving*” with following sub-categories:

- **Driver restricted attention** (DRA) which is the result of a biological factor such as blinking. During blinking the driver is not able to perceive any visual information.
- **Driver misprioritised attention** (DMPA) which, as its name states, occurs, if multiple driving-related tasks are not correctly prioritized. As a result, a very relevant task to safe driving is thoroughly excluded such as looking over the shoulders while moving forward and not paying attention to the vehicle in front for a timely braking reaction. Regan et al. (2011) emphasized on different interpretations of such level of inattention with respect to driving experience.
- **Driver neglected attention** (DNA) which happens, if the driver neglects an important driving-related task. An example is a driver who does not expect a train at a railway level crossing and, therefore, does not observe properly. In fact, “*expectation and over-familiarity*” motivate this type of inattention.
- **Driver cursory attention** (DCA) which is the result of hastiness during driving. Consequently, driving-related tasks are not performed thoroughly.
- **Driver diverted attention** (DDA) which is similar to the definition of driver distraction and has been studied deeply for understanding of safety in comparison to other categories. In this category, the secondary task, which diverts driver attention, comprises not only internal or external activities, but also mental activities. Clearly, DDA is either driving-related, e.g. by navigation system, similar to attentional competition studied by Hoel et al. (2010) or non-driving-related, like by text messaging. In addition, it might occur both voluntarily/involuntarily and internally/externally. This category of driver inattention mainly covers unusual and unexpected tasks which drivers hardly ignore.

Despite the mentioned taxonomy of inattention, Regan et al. (2011) also stated issues such as whether tools and methods, in general, are able to collect data for all mentioned categories of driver inattention or whether it is possible to distinguish between the mentioned categories in a crash. It seems that except for the DDA, which has been studied systematically, other categories need new instrumentations and new algorithms for assessing the experiment data in future.

It is clear that in the taxonomy of driver inattention defined by Regan et al. (2011), the driver state itself is also included as a factor which affects the inattention level. For example, severe drowsiness results in longer eye closures and leads to DRA. In addition, they also believe that depending on the reaction of the driver to an event, different categories of inattention might occur.

Regan et al. (2011) interpreted the phenomenon of *looked but failed to see* as either resultant of drowsiness and, therefore, DRA or internal thoughts, i.e. DDA. Moreover, daydreaming has been categorized as DDA, since it is also a non-driving-related task. They discriminated daydreaming from internal unintentional thoughts from this aspect that daydreaming is more *fantasy-like*, while internal unintentional thoughts are linked to *current concerns*. In most cases, the driver recognizes the daydreaming only after it is finished.

Alertness and vigilance

According to Oxford dictionary (Oxford, 2014), vigilance is defined as “*the action or state of keeping careful watch for possible danger or difficulties*”. Moreover, alertness is defined as “*the state of being quick to notice any unusual and potentially dangerous or difficult circumstances*”.

Dukas (1998) defined vigilance as “*a general state of alertness that results in enhanced processing of the information by the brain*”. Thiffault and Bergeron (2003), however, categorized vigilance into two groups. The first one deals with “*information process and sustained attention*”, while the other one refers to “*physiological processes underlying alertness or wakefulness*”. Overall, during driving, the lack of both of them results in safety-critical situations as also mentioned by Schmidt et al. (2009).

Terminologies of this work

In this work, the term *drowsiness* will be used to refer to a broader scope including both sleepiness and fatigue. The term fatigue will not be used further to avoid ambiguity in its definition. Since in this work only performing a secondary task will be engaged in the primary driving task, we only use the term *distraction* or DDA in the taxonomy defined by Regan et al. (2011). The state that a driver does not threaten road safety and is aware of danger is called *awake*.

1.3. Countermeasures against drowsiness during driving

Countermeasures are techniques for keeping a driver awake and consequently preventing car crashes. Examples of such countermeasures are taking a nap (43%), opening a window (26%), drinking coffee (17%), pulling over or getting off the road (15%), turning on the radio (14%), taking a walk or stretching (9%), changing drivers (6%), eating (3%) and asking passengers to start a conversation or singing (3%), which are all “*driver-initiated*” (May, 2011). The percentages represent how often each countermeasure was selected by 4010 subjects based on the results provided by Royal (2003). Multiple countermeasures have been selected by half of the subjects.

Anund (2009) studied the interaction between countermeasures and various factors such as age, gender, driving experience, etc. and showed how often each countermeasure is used by drowsy subjects. She found taking a nap as a very typical countermeasure and showed that this countermeasure was used by those drivers who had experienced sleep-related crashes, professional drivers, males and drivers aged 46–64 years.

Schmidt et al. (2011) implicitly analyzed the effect of conversing as a countermeasure. They showed that a 1-min conversation for verbally assessing the drowsiness level led to an increase of the vigilance state estimated by physiological measures. However, the activating effect of this countermeasure lasted only up to 2 min in their experiment. Moreover, they emphasized in their study that their experiment does not declare any issue about the type of conversation and its generality as a countermeasure, since the conversation only comprised the driver self-estimation of the drowsiness level. All in all, the distracting effect of talking with a passenger or cell phone during driving should not be ignored.

Gershon et al. (2011) explored the drowsiness countermeasures based on their usage and perceived effectiveness aspects for professional and non-professional drivers regardless of the drivers' age. They considered a driver as a professional one, if driving was part of his work requirements e.g. a taxi or bus driver. However, non-professional drivers had other primary jobs than driving. The results revealed that listening to the radio and opening the window were the most frequently used and perceived as effective countermeasures by both groups. They believe that it is the accessibility of these two countermeasures which has made them so popular. Non-professional drivers used talking to passengers as the second most common countermeasure. Drinking coffee, however, was used by professional ones more often. In general, planning rest stops ahead, stopping for a short nap and drinking coffee were more often used by professional drivers in comparison to the non-professional ones. Gershon et al. (2011) justified this finding based on the level, at which each group counteracts drowsiness. They called it "*tactical/maneuvering level*" for non-professional drivers versus "*strategic/planning level*" for professional ones. The former, unlike the latter, tries to decrease the weariness and boredom without planning ahead.

Rumble strips located on the roadsides or on the middle line are another safety technique to prevent car accidents due to lane departure (May, 2011). As soon as the driver crosses or contacts the rumble strips with a wheel, vibrations together with a loud noise will be heard inside the car and notify the driver of the lane departure. Damousis and Tzovaras (2008) and Hu and Zheng (2009) used rumble strips in their experiments and defined events of hitting rumble strips as critical events for driver drowsiness detection systems whose occurrences should not be missed. Anund (2009), however, studied "*the effect of rumble strips on sleepy drivers*" and analyzed the 5 min windows preceding a rumble strip hit and shortly after the hit. The results revealed an increase in the studied sleepiness indicators such as blink duration and electroencephalography analysis (will be explained in Section 2.1.2). Interestingly, shortly after the hit the subjects were alert and enhanced performance was observed. Nevertheless, these effects did not last more than 5 min and signs of sleepiness returned. Therefore, as May (2011) also mentioned, rumble strips are only useful for highlighting the driver drowsiness level but clearly they do not eliminate drowsiness.

1.4. Driver drowsiness detection systems on the market

This section introduces driver drowsiness detection systems available on the market by car companies and explains the idea behind their detection methods based on the review provided by Čolić et al. (2014).

Mercedes-Benz In 2009, Daimler AG introduced the *Attention Assist* to warn drowsy drivers. This system is mainly based on steering wheel movements and their velocities (Daimler AG, 2008, 2014a). The idea behind the Attention Assist is that an alert driver steers with small corrections

and moderate movements. On the contrary, a drowsy driver does not steer for a short time followed by a fast large-amplitude steering activity. Hence, the transition between these two phases is monitored for vehicle speeds from 80 to 180 km/h. Since 2013, this range has been enhanced to 60 and 200 km/h in the new S and E-class series. The system determines whether the driver is drowsy by comparing the current steering behavior with that of the beginning of the driving session. If the difference between them exceeds a threshold, the driver is warned both audibly and visually, i.e. by an audible signal and displaying “Attention Assist: Break!” in the instrument cluster. In some new models, the *attention level* in form of a 5-level bargraph is also displayed during the drive.

Figure 1.3 shows an example of a typical steering event which is detected by the Attention Assist and its corresponding vehicle trajectory. In this figure, up to $t = 6$ s, there is no steering activity which leads to lane departure. However, shortly after $t = 6$ s, the driver corrects his lane departure by an abrupt steering wheel movement up to 10° .

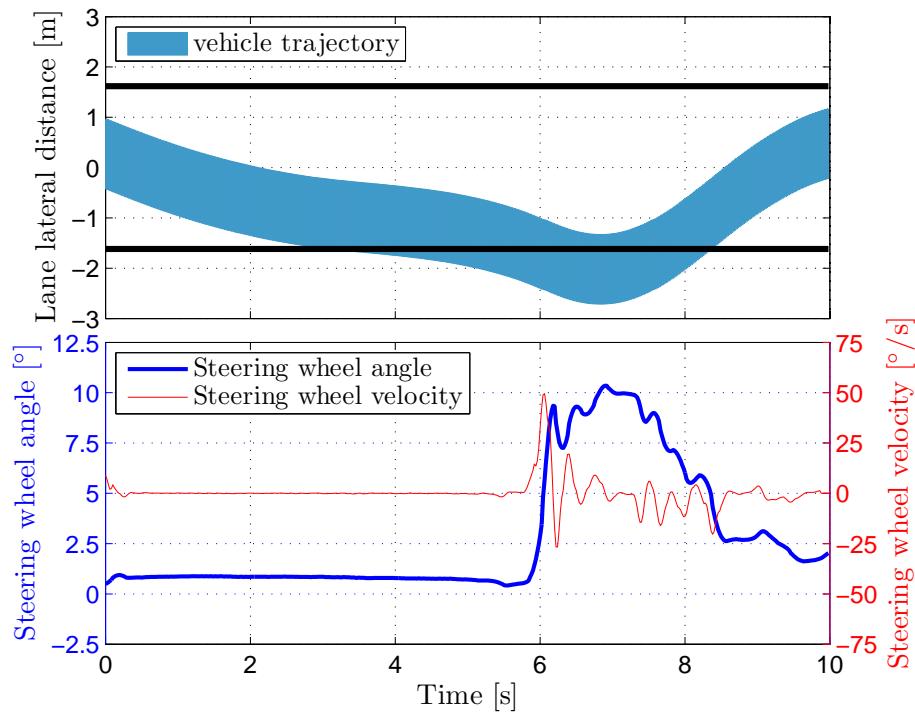


Figure 1.3.: A typical steering event detected by the Attention Assist as a drowsiness-related steering wheel movement

Attention Assist has several advantages. First of all, it is able to warn the driver early enough before the occurrence of microsleep events. Moreover, the parameters of the systems are set individually at the beginning of the drive. Therefore, regardless of the driving characteristics of a specific driver, the system adapts itself to the current driver. In addition, weighting factors such as time of day and driving time have been considered in the warning strategy. Longitudinal and transverse acceleration and vehicle speed also contribute to improve the system. Since not all steering wheel movements are necessarily drowsiness-related, the Attention Assist does not take all steering movements into account. Side wind, road bumps and operation of center console elements, e.g. navigation system or turn signal indicator, are examples of such non-in-vehicle and in-vehicle external influences on the steering wheel movements which are filtered out for assessing driver's level of drowsiness.

Volvo *Driver Alert Control* (DAC) introduced by Volvo Group in 2007 works based on a lane tracking camera which observes the lane keeping behavior (Volvo Group, 2014). It is located between the rear-view mirror and the windshield. Moreover, steering wheel movements are taken into consideration. The system is active, if the vehicle speed exceeds 65 km/h. During the drive, a bargraph displayed in the instrument cluster shows the drowsiness level to the driver. As soon as it decreases to one bar, the driver will be warned both visually, i.e. a message on the speedometer and acoustically, i.e. an audible warning.

Ford The Ford Motor Company has also developed a system called *Driver Alert*, which, similar to the system of Volvo, is based on the lane keeping behavior (Ford Motor Company, 2010). The lane tracking camera is located behind the rear-view mirror. Depending on the detected lanes, the position of the vehicle is predicted and then compared with the true one. If the difference exceeds a certain threshold, the driver is warned audibly and visually on the instrument cluster. If the system determines that the alertness level is still decreasing, a second warning is displayed which should be accepted by pressing a button. The system will be reset after turning the engine off or after opening the driver's door.

Volkswagen The *Fatigue Detection* system is developed by Volkswagen AG and is based on the steering movements and some other available signals (Volkswagen AG, 2014). It activates after a minimum driving time of 15 min and for a vehicle speed exceeding 65 km/h. In the case of the drowsiness detection, the driver is warned both visually and acoustically. After 15 min, the warning is repeated, if the driver keeps driving. Under some conditions the system might not work properly, such as “*sporty driving style, winding road and poor road surfaces.*”

In addition to the mentioned systems integrated into the vehicle, there are also some other systems available which can be installed as an additional equipment on a vehicle. More details about such systems are provided by Kircher et al. (2002), Barr et al. (2005) and Čolić et al. (2014).

1.5. Thesis outline

The outline of the thesis is pictorially shown in Figure 1.4. It starts with Chapter 2, in which we discuss and review the approaches for objective and subjective driver state measurements. In the context of objective measures, both driving performance and driver physiological measures are covered. Since this thesis concentrates on driver eye movements as a drowsiness indicator, the human visual system, its structure and relevant types of eye movements are introduced in Chapter 3. Further, in Chapter 4 the measurement system used in this work is investigated for in-vehicle applications. This chapter also describes the conducted experiments for collecting eye movement data which will be used later to achieve the goals of this study (data collection block in Figure 1.4). Our experiments are based on daytime and nighttime drives both on real roads and in the driving simulator. After data acquisition, Chapter 5 deals with the detection of different types of eye movements including both simple and complex approaches like the median filter-based and wavelet transform methods, respectively (event detection block in Figure 1.4). Afterwards, in Chapter 6, we answer the question how the occurrence of the eye movements is associated with each other, especially under distracted driving conditions. From the detected events in Chapter 5, we introduce extracted blink features in Chapter 7 where we extensively review the previous studies in order to compare them with our findings (feature extraction

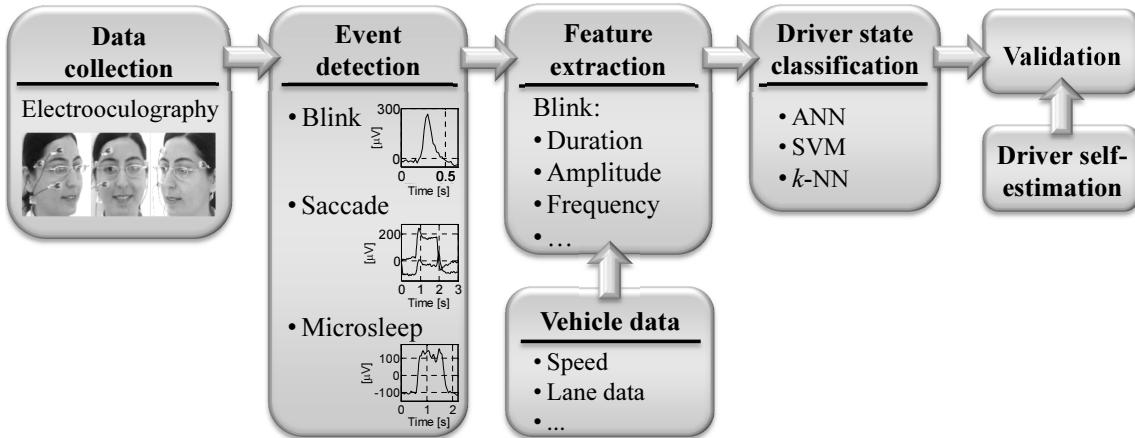


Figure 1.4.: Tool chain of this thesis

block in Figure 1.4). In addition, the relationship between extracted features and drowsiness is investigated individually by event-based and correlation-based analyses. In Chapter 8, a driver state classification is performed using the extracted features by applying three types of classifiers. The classification results are compared with each other and the optimal classifier is suggested (classification and validation blocks in Figure 1.4). This chapter ends by assessing feature dimension reduction approaches. Finally, in Chapter 9 we summarize and conclude the results of this work and give an outlook of this study which opens rooms for future work.

1.6. Goals and new contributions of the thesis

The main goal of this thesis is to provide a ground truth-based eye movement analysis for the future development of driver observation camera with concentration on driver drowsiness detection. In other words, this work targets the full coverage of requirements to be fulfilled during the development of the camera and the warning system for timely detection of the onset of drowsiness. Therefore, studying, implementation and evaluation of drowsiness-related eye movement features are on the central focus. To this account, well-known methods for event detection, feature extraction and classification need to be investigated along with providing new ideas and approaches. These goals can be achieved by designing daytime and nighttime experiments with representative driving scenarios.

In the following, the main contributions of this thesis towards driver drowsiness detection are summarized.

- Based on a thorough literature review on the terminologies related to drowsiness, a suitable term among many terminologies e.g. fatigue, sleepiness, etc. is targeted which best describes the driver state during driving. (Section 1.2)
- A new approach for enhancing the calculation of the alpha spindle rate is suggested and evaluated against the initial calculation method. This idea benefits from the fusion of eye movement activity information into the calculation of the alpha spindle rate. (Section 2.1.2)
- Most of the previous studies collected eye movement data with the eye movement measurement system used in this work (electrooculography) in laboratories or in fixed-base driving simulators. However, in this work, the reliability and robustness of this system

for in-vehicle measurements is evaluated on a proving ground and under fully controlled real road driving conditions. In addition, due to using vehicle sensors, road-dependent eye movements are thoroughly analyzed. (Section 4.1)

- The results and findings of this thesis are evaluated based on different experiments on both real roads and driving simulators with a total number of 43 subjects. Moreover, by designing both daytime and nighttime experiments, the collected data set contained all vigilance levels during driving. On the contrary, in previous studies, drowsiness detection has mostly been explored under simulated driving (see Table 7.3). In addition, almost all of the previous researches had a smaller number of participants than this study in their experiments. Moreover, in most of the previous works, sleep-deprived subjects were participated in the experiments. This leads to an imbalanced data set in terms of availability of information about different levels of driver drowsiness and vigilance. (Sections 4.2 and 4.3)
- In an experiment on the real road, where secondary tasks have been performed along with the primary driving task, eye blink behavior is analyzed. Based on the findings, a recommendation about task-induced blinks is made. (Chapter 6)
- In order to improve the performance of eye movement detection and the accuracy of the extracted features, two preprocessing steps, drift and noise removal, are proposed. The strength of our proposed noise removal approach is its flexibility in selecting noise removal threshold with respect to the amount of noise. (Section 5.3.3)
- For the detection of blink and distinguishing them from other eye movements, two novel algorithms are proposed. The first algorithm is based on derivative signal and is suitable for detection of fast eye movements. The second approach is based on continuous wavelet transform and covers the detection of both fast and slow eye movements. Therefore, in contrast to other studies, this work addresses the detection of all relevant eye movements to drowsiness. (Sections 5.2 and 5.3.2)
- This work is the most comprehensive study on eye blink features for in-vehicle applications and under real driving conditions. By considering all inconsistent definitions of features in previous studies, 19 features are well-defined and extracted per blink. Afterwards, their evolution due to drowsiness is studied individually. In addition, the findings are compared with previous studies which were mostly based on restricted conditions. (Section 7.2)
- Clearly, the quality of driver observation cameras in detecting eye blinks will not be as high as that of the measurement system used in this work. This issue is investigated and possible peak amplitude loss is evaluated. (Section 7.6)
- Two feature aggregation approaches are suggested for the investigation of the relationship between extracted eye movement-based features and drowsiness. Based on the first approach, features are analyzed shortly before safety-critical events. Moreover, the lane-keeping based and eye movement-based drowsiness detection methods are challenged. However, the second approach benefits from quick changes of drowsiness level in the course of time. On this account, in the second approach, feature values are extracted over time regardless of safety-critical events (Sections 7.1, 7.4 and 7.5)
- Since features based on physiological measures are highly individual and vary from one subject to the next, we propose two baselining methods to minimize this deteriorating effect. (Section 7.1.3)

- Apart from scrutinizing features as separate and independent source of information, they are fused and investigated by different state-of-the-art classifiers. The performance of the sophisticated classifiers are evaluated and compared with each other for different types of extracted features and different data division methods. (Chapter 8)
- One the one hand, data collection in driving simulators, either fixed or moving-base, and on real roads are very expensive. One the other hand, for driver drowsiness detection with high detection rate and low false alarm rate, representative data set with both awake and drowsy driving samples is essential. In this work, first, the imbalanced data sets are explored and their issue is addressed by artificially balancing the data sets. Moreover, it is investigated whether artificially balanced data sets can replace data collection of awake samples. (Chapter 8)
- This thesis provides new insights into the generalization of the data collected in the driving simulator to that of the real road driving on feature fusion level. To this end, two new approaches are studied. (Section 8.6)
- Finally, to address the concerns about computational resource in in-vehicle warning systems, feature dimension reduction approaches are applied. (Section 8.7)

2. Driver state measurement

This chapter introduces approaches for measuring the driver state either objectively or subjectively. The former deals with methods for developing an external measure to prevent car crashes due to driver drowsiness. The latter, however, serves as the reference for assessing the efficiency of an objective measure. In addition, previous studies, which have introduced these measures, are reviewed. A novel idea for improving one of the objective measures is also proposed and evaluated.

2.1. Objective driver state measures

Driver objective measures, as their name implies, are measures which are collected by a measurement technique such as sensors, electrodes, etc. with no deliberate interference of the driver in them. An objective measure is developed based on either driving performance measures, driver physiological measures or their fusion which is called *hybrid* measures.

2.1.1. Driving performance measures

Čolić et al. (2014) summarized car crashes due to drowsy driving with the following characteristics which were based on reports by the police or the driver himself:

- “*Higher speed with little or no breaking*” which means the combination of high speed with low reaction time due to drowsiness.
- “*A vehicle leaves the roadway*” which is also called *single-vehicle crash* due to lane departure.
- “*The crash occurs on a high-speed road*” which might be due to monotonicity of such roads.
- “*The driver does not attempt to avoid crashing*” which is the result of severe drowsiness and falling asleep.
- “*The driver is alone in the vehicle*”.

The common point in these characteristics is that they all lead to degraded driving performance. As a result, by quantifying them by means of sensors installed in the car, it is possible to develop a drowsiness indicator measure to prevent car crashes. These measures are called driving performance measures. A main specification of them is that they observe the driver indirectly. Moreover, they do not measure the drowsiness itself, but its consequences.

An advantage of such measures is that they can be measured even without having a direct contact with the driver, namely unobtrusively. These measures, which are all related to the vehicle, mainly contain steering wheel and lane keeping behaviors (Liu et al., 2009). In the following, both of these behaviors and studies on them are introduced and discussed.

Driver's lane keeping behavior

One of the measures reflecting driving performance is the lane keeping behavior which is extracted from the lane lateral distance. Lane lateral distance refers to the offset between the middle of the lane and middle of the vehicle. Analysis of this measure is mainly based on the assumption that an alert driver, unlike a drowsy one, stays in the middle of the lane. However, not staying in the middle of the lane is not necessarily a reason for a low vigilance state of the driver. A counterexample is a driver who keeps more to the left side of the lane for a better forward look, while another car is in front of him. Therefore, the *standard deviation of lateral position* (SDLP) is used instead to quantify to what extent the driver *swings* in the lane.

Johns et al. (2007) and Damousis and Tzovaras (2008), who studied the relationship between eyelid activity and drowsiness, used lane departure events as safety-critical phases of the drive for evaluating their eyelid-based drowsiness measure. Damousis and Tzovaras (2008) reported that the analysis of brain activity (will be explained in Section 2.1.2) did not correlate well with these events. Similarly, Sommer and Golz (2010) also relied on the SDLP as an objective reference measure which increased due to drowsiness in their experiment. They defined 13% deviation of it as the threshold between mild and strong drowsiness.

Verwey and Zaidel (2000) defined the following lane keeping behavior occurrences as driving errors:

- “*road departure error: leaving the pavement with all four wheels*”
- “*moderate lane crossing error: leaving the pavement with one or two wheels*”
- “*minor lane crossing error: crossing the solid lane markings with one or two wheels*”
- “*time-to-line crossing (TLC): crossing the solid lane marking within 0.5s, if no action is taken (TLC < 0.5s)*”

They acknowledged the TLC as a reliable measure reflecting poor driving performance. Skipper and Wierwille (1986) found a significant positive interaction between eyelid closure and SDLP whose variation improved distinguishing between alert and drowsy classes by a discriminant analysis model. Åkerstedt et al. (2005) also studied lane departure events defined as “*four wheels outside the left lane marking (accident) and two wheels outside the lane markings (incident)*”. In their study with shift workers, it was shown that the number of incidents increased three times due to drowsiness. SDLP also increased from 18 cm to 43 cm. Otmani et al. (2005), however, found no interaction between SDLP and sleep deprivation, although this measure increased during their experiment for both sleep-deprived and non-sleep-deprived subjects. In fact, it was the driving duration (time-on-task) which affected the SDLP. Ingre et al. (2006) analyzed the relationship between SDLP and a subjective measure for shift workers with both enough night sleep and no night sleep. They found that the significant relationship between them is *curvilinear*, i.e. looks like a curved line. Moreover, they emphasized the large between-subject differences in the values of the SDLP as an issue. Arnedt et al. (2005) reported a tendency to left side lane keeping for subjects with prolonged wakefulness.

Wigh (2007) and Ebrahim (2011) studied event-based lane keeping behavior. In order to define events, two zones, called *virtual edge zone* (VEZ), were defined on the right and left side of the vehicle near to each lane. The location and the width of the VEZ's were adjusted individually based on the lane keeping behavior of the driver during driving. Hence, the zones were adapted for a driver who tended to keep left or right in a lane without any problem. The entrance of the wheels to the zones were then weighted, i.e. the more the wheel entered the zone, the higher was

the corresponding weight. In the end, based on an incremental running mean of the weights and its comparison with a threshold, it was decided whether to warn the driver or not. In fact, the driver was considered drowsy with respect to the amount and the number of times he entered or got close to the defined zones which were not necessarily the road lane marking.

Driver's steering wheel movement behavior

Another very common method for an indirect observation of the driver is the analysis of steering wheel movements in terms of amplitude and velocity. According to Thiffault and Bergeron (2003), an alert driver recognizes changes in the road ahead early enough and consequently reacts with small steering wheel corrections from 1° to 5° . On the contrary, due to drowsiness, the driver's reaction time increases. As a result, he reacts to the trajectory changes by means of larger steering wheel movements from 6° to 10° or even larger than 10° . Clearly, this is linked to larger lane departures which are more dominant during drowsiness as mentioned before.

Thiffault and Bergeron (2003) explored the amplitude and the frequency of steering wheel movements during two conditions: a monotonous versus a non-monotonous driving scenario. Both measures increased in their study as a function of time-on-task. Bouchner et al. (2006) showed that for more than 80% of their sleep-deprived subjects, the ratio between fast and slow steering wheel corrections increased over time. Otmani et al. (2005), however, reported a significant dependency between age and small steering wheel movements (up to 5°), i.e. 2.26° for the middle-aged group versus 1.99° for the young group. Although this measure was not subject to changes due to sleep-deprivation, it increased during their experiment.

In addition to single feature analysis of the lane data or steering wheel movements, Friedrichs et al. (2010), Friedrichs and Yang (2010b) and Pimenta (2011) extracted a large number of lane-based and steering-based features and studied their relationship with drowsiness by applying classification methods. By considering the best 11 features (both lane and steering-based ones), Friedrichs and Yang (2010b) achieved an average detection rate of 73.5%. They defined driver's level of vigilance with three classes as awake, questionable and drowsy and evaluated them against a subjective self-rating measure.

Apart from the question whether steering wheel movements and lane keeping behavior are reliable measures for drowsiness detection, they are highly subject to external influences as studied by Friedrichs et al. (2011). In the mentioned study, in addition to driver distraction and vehicle operations, which might shortly affect steering movements, other factors such as weather (rain, fog, snow), light conditions (e.g. an illuminated tunnel), vehicle speed, road condition and curvature were also introduced and explored. They observed that the lack of vision during night drives affected steering and lane keeping behavior with no direct regard to the driver's vigilance state. Furthermore, vehicle speed was acknowledged as a factor with almost a direct impact on the steering velocity. Hence, effort should be made to compensate these effects while extracting steering movement-based features. Driving on roads with construction sites also seems to affect steering movements due to narrowed lane width. Higher road curvature led not only to a larger deviation of the lane lateral position from the middle, but also to more frequent fast steering velocities. Interestingly, they found the contribution of road curvature to the lateral lane deviation masking that of the drowsiness. As the last investigated external factor, the road condition was introduced which included road surface and bumps. Although road surface conditions were not found to be correlated with the steering velocity, the non-uniform delay between road bumps and steering jerks confirmed the destructive effect of road

bumps on driving performance measures. Friedrichs et al. (2011) also believed that the good lighting condition of a tunnel positively affects the driver vigilance.

Driver modeling is another approach which also addresses the driver state detection by means of steering movement and lane keeping behavior (Pilutti and Ulsoy, 1999). By applying system identification methods, a model is developed for predicting the steering wheel angle based on the lane lateral position. Changes in the model parameters or the deviation of the measured steering wheel angle from the predicted one are suggested as the objective measures for driver state identification by Pilutti and Ulsoy (1999). Hermannstädtter and Yang (2013) also distinguished between distracted and undistracted driving based on driver modeling.

Overall, driving performance measures have some drawbacks. First of all, the lane-based drowsiness detection is restricted to roads with lane marking. Moreover, monitoring steering wheel movements of a driver depends to some extent on the manufacturing technology of the vehicle. As a result, the parameters of the warning system should be tuned for each vehicle individually.

Another important issue in studying driver performance measures is the generalization of the findings to real road conditions from the experiments conducted in driving simulators. As an example, unevenness of real roads is not easy to implement in driving simulators which definitely affects the steering behavior. In a study by Hallvig et al. (2013), which compared real road driving with driving simulator, it was found that during night drives, the subjects tended more to the middle of the lane in comparison to the driving simulator. In addition, higher SDLP and shorter TLC were reported in the driving simulator. These behaviors were all interpreted to be highly related to “*threat against safety*” on a real road and the fact that higher level of drowsiness can be achieved in driving simulators. Similarly, Philip et al. (2005) also reported more frequent lane crossings with higher amplitude in a driving simulator in comparison to a real road scenario.

As mentioned in Section 1.1, the most critical drawback of driving performance measures is that they are valid measures as long as the driver plays the key role in performing the driving task. As soon as an assistance system supports the driver in steering and lane keeping, the introduced measures are unreliable. Therefore, for near future, other solutions of driver drowsiness detection in partly automated driving conditions are needed. In other words, an indirect monitoring of the driver should be replaced with the direct monitoring based on driver physiological measures.

2.1.2. Driver physiological measures

Driver physiological measures, in general, are measures based on the direct observation of the driver which can be either intrusive like *electrophysiological* ones or non-intrusive like cameras. Unlike the latter, the former requires a direct contact of the electrodes to the driver’s skin. These measurement techniques provide an objective measure to describe the driver state (Simon et al., 2011) and are believed to outperform other introduced measures from the point that they detect drowsiness at its early phase (Čolić et al., 2014).

In the following, these measurement techniques are introduced and reviewed.

Electroencephalography

Electroencephalography (EEG) is a physiological method for recording brain activity in terms of electrical potential by electrodes located on special positions on the head (Niedermeyer and da Silva, 2005; Kincses et al., 2008). This measurement method has the advantage of being a

continuous recording method together with a high temporal resolution (up to 512 Hz) (Kincses et al., 2008).

Figure 2.1 shows a 32-electrode arrangement (excluding 4 electrodes for eye movement data collection) of this measurement system which can also be used during driving by wearing a cap.

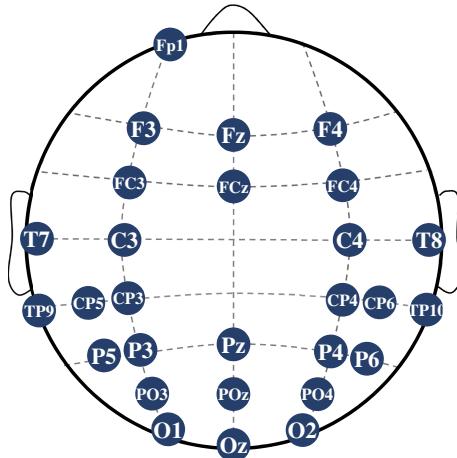


Figure 2.1.: 32-electrode arrangement of EEG (excluding 4 electrodes for eye movement data collection)

In order to improve the conductivity of the electrodes, a special paste should be used between the electrodes and the skin. However, depending on the number of electrodes being used, injecting the paste can be very time-consuming and the subjects should wash their hair after data collection. Figure 2.2 shows an electrode, an EEG-cap and the paste being injected from ActiCAP, Brain Products GmbH (2009).



Figure 2.2.: ActiCAP measurement system for EEG recording by Brain Products GmbH (2009)

The amplitude range of EEG waves is usually between 0 to $200\mu\text{V}$ which makes it difficult to distinguish them from noise and some artefacts (e.g. scratching the head) (Svensson, 2004; Damousis and Tzovaras, 2008). Moreover, EEG is very sensitive to movements and muscle artefacts. Even fast spontaneous eye blinks affect the waves easily. Therefore, if a suitable artefact removal is not applied, the collected data should not be analyzed further. Simon (2013) and Santillán-Guzmán (2014) studied different approaches for EEG artifact removal.

EEG waves can be either analyzed in the time domain or in the frequency domain. The former includes calculating some statistical values within an interval (Dong et al., 2011), whereas the latter covers the analysis within the following different frequency bands: δ (up to 3.5 Hz), θ (4–7 Hz), α (7–13 Hz), β (14–30 Hz) and γ (35–100 Hz) (Niedermeyer and da Silva, 2005). Among these bands, the α -band and especially α -bursts have shown to be the most drowsiness-related

bands for detecting early phases of drowsiness (O'Hanlon and Kelley, 1977; Kecklund and Åkerstedt, 1993; Eoh et al., 2005; Papadelis et al., 2007; Schmidt et al., 2009; Simon et al., 2011). Moreover, α -waves are mainly dominant during eye closure (Saroy and Craig, 2001). Figure 2.3(b) shows an EEG recording with closed eyes containing α -bursts. For a better comparison, EEG signals recorded with open eyes are also shown in Figure 2.3(a). The bursts occur with different amplitudes in different locations on the head and are very dominant in parieto-occipital electrodes, i.e. P3, Pz, P4, O1, O2 electrodes. This was also reported by Simon et al. (2011) who evaluated α -bursts with respect to the driver drowsiness and called them α -spindles.

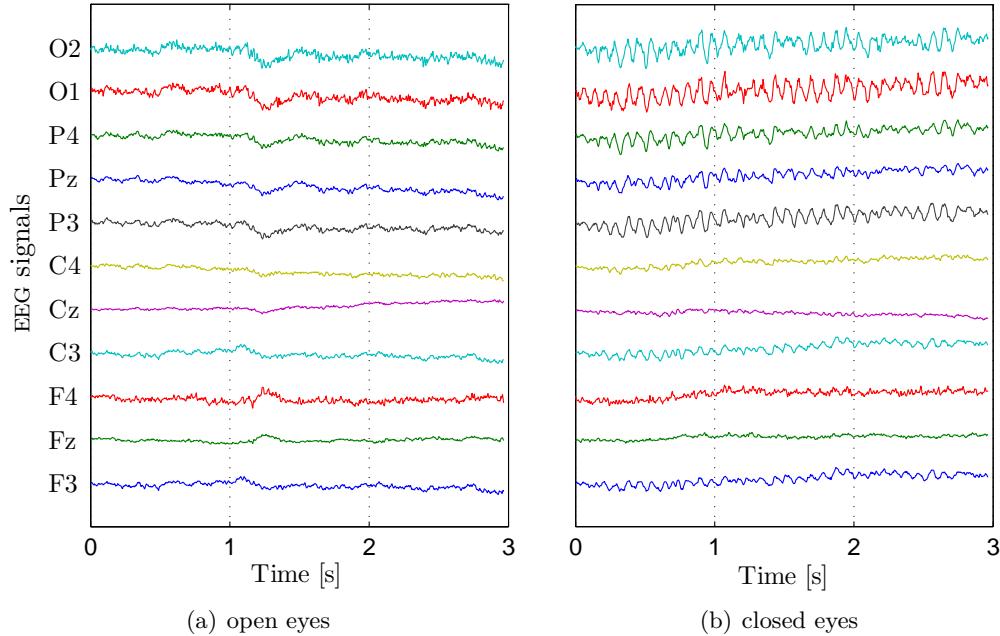


Figure 2.3.: EEG signals showing α -bursts with closed eyes versus open eyes

Figure 2.4 shows the Fourier transform of the electrode O2 for both cases with open and closed eyes. It can be seen that the bursts are within 8.8–12.5 Hz which corresponds to the frequency range of the α -band.

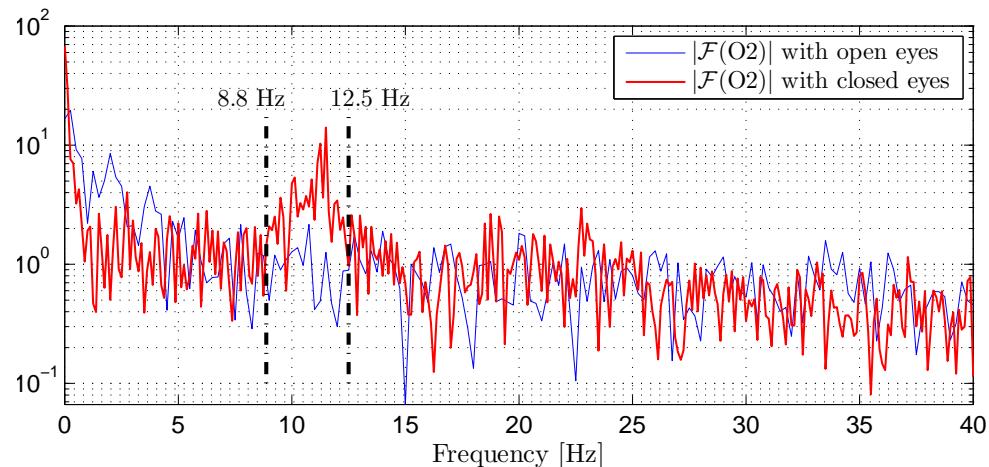


Figure 2.4.: Frequency components of the α -bursts by applying the Fourier transform to the wave of the O2 electrode shown in Figure 2.3

Simon et al. (2011) suggested a method for identification of the mentioned spindles as explained in the following. First, a 1-s zero-mean-made segment of the EEG recording with 75% overlap is multiplied with a Hamming window and the fast Fourier transform of it is calculated. If the maximum value of the calculated spectrum is located within the range of the α -band, then the full width at half maximum (FWHM) of the spectral peak is determined and compared with twice the bandwidth of the Hamming window ($BW_{Hamming}$). Depending on the result of the mentioned thresholding analysis (desired: $FWHM < 2BW_{Hamming}$), the segment is subject to further investigations. $BW_{Hamming}$ corresponds to the minimum bandwidth for an oscillatory activity. This procedure is repeated for all 1-s zero-mean-made segments. After calculating the *signal-to-noise ratio* (SNR) (for more details see Simon et al. (2011)), segments with acceptable SNR values and peak frequencies, whose deviation from each other is not more than 10%, are summed up as an α -spindle. Different features such as duration, spectral amplitude and peak frequency of the discrete α -spindle events are then calculated by moving average within a 1 or 5 min windows and 50% to 80% overlap. Simon et al. (2011) also introduced *alpha spindle rate* (ASR) as the number of α -spindle events occurring within the mentioned moving average time intervals. Based on the statistical analysis, they showed in their study that α -spindle parameters, averaged over all subjects, increased within the last 20 min of the drive in comparison to the first 20 min. In addition, ASR in their study outperformed the common α -power for the drowsiness detection.

In agreement with the mentioned study, Schmidt et al. (2011) explored the variation of ASR, blink duration, heart rate and reaction time in a monotonous daytime drive under real driving condition. The results showed a significant increase of ASR and a significant decrease of heart rate. Blink duration increased as well, however, it was not statistically significant. On the contrary, Anund (2009) showed that θ and α activity of the EEG do not necessarily detect phases shortly before a safety-critical event such as lane departures and hitting the rumble strips.

Unlike Schmidt et al. (2009) and Simon et al. (2011), who analyzed the long-term variation of the ASR, Sonnleitner et al. (2011, 2012) studied the short-time variation of it with respect to driver distraction, i.e. while the driver performed secondary tasks. These studies showed that under both real and simulated driving conditions, the ASR is subject to contrary variations due to the type of the secondary tasks, namely auditory and visuomotor, in comparison to the primary driving task. That means, performing the auditory secondary task leads to an increase of the ASR, while performing the visuomotor secondary task results in a drop. Their experiment under real road condition is explained in Section 4.2.2. Figure 2.5(a) shows the ASR of a participant in the experiment. ASR_{60} refers to the length of the moving average window, i.e. 60 s, for counting the number of α -spindles. The phases indicating secondary tasks are also shown. Sonnleitner et al. (2011) related these results to the “*visual information process*” whose increase and decrease directly influences the value of the ASR. Performing the visuomotor secondary task in their study is very similar to the data entering into the navigation system in daily routines. On the other hand, the auditory secondary task in this study can be considered as a cognitive distraction similar to a cell phone conversation during driving. Therefore, it can be concluded that the ASR and, in general, the EEG is a very dynamic and sensitive measure which makes its interpretation ambiguous. A high value of it can either be due to drowsiness or cognitive distraction and, similarly, its lower values show either alertness or visual distraction of the driver. To put it another way, not all α -spindles seem to be drowsiness-related.

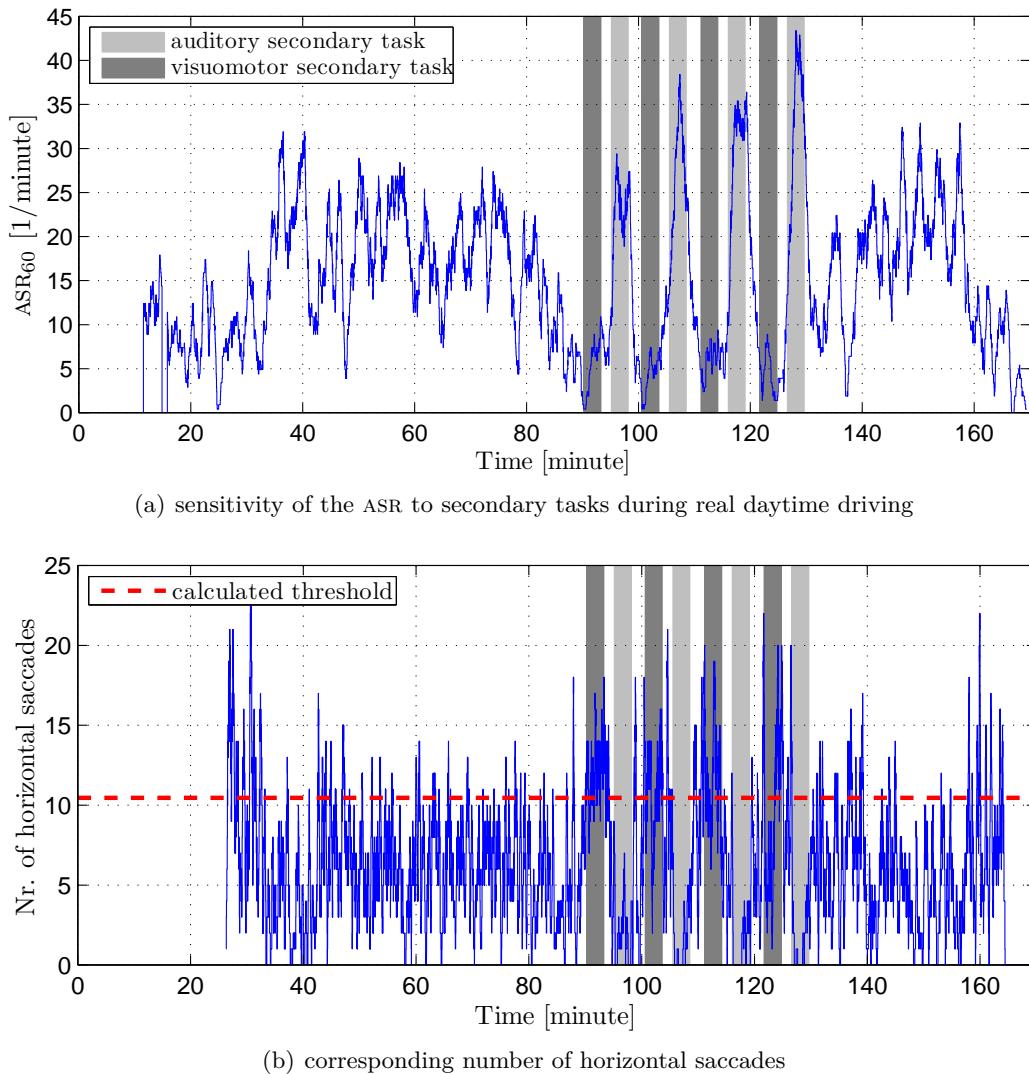


Figure 2.5.: Sensitivity of the ASR to auditory and visuomotor secondary tasks and the corresponding number of horizontal saccades

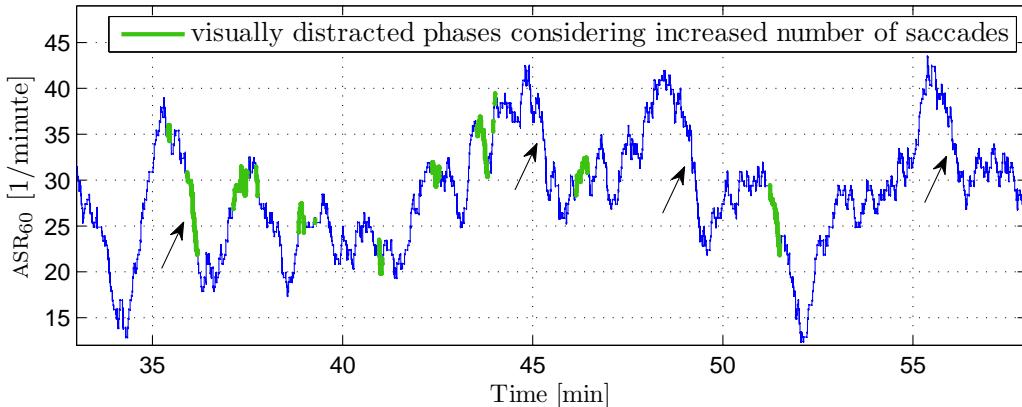
Our approach for enhancing the ASR

The new approach adopted in this work to resolve the aforementioned ambiguity in the interpretation of the ASR values and to consequently improve the correlation of the ASR with driver drowsiness is to remove phases of the drive where the driver is visually distracted. In fact, phases with a smaller number of spindles should be excluded in the calculation of the ASR, in order to prevent its decay. As a result, the increasing trend of the ASR due to drowsiness gets intensified. In the following, this approach is explained and applied.

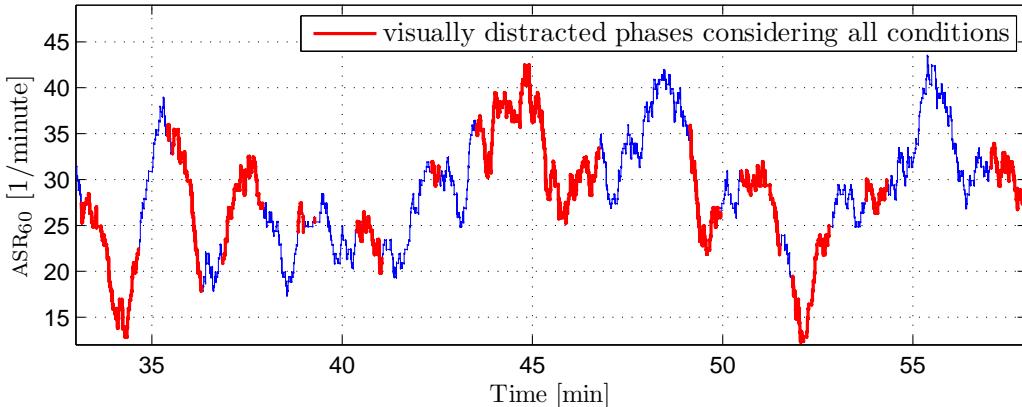
Since visual distraction is mostly linked with eye movements, i.e. saccades (will be defined in Section 3.3), a temporary increase in the number of eye movements can, as a result, be interpreted as an event with high visual information process. Figure 2.5(b) shows the number of horizontal eye movements for the same subject synchronous with the ASR. It is calculated by a moving sum window of 15 s which is shifted at 0.02 s steps. As a threshold for defining the increased number of eye movements, the average plus one standard deviation (std) of the found number is used, namely mean (nr. of eye movements) + std (nr. of eye movements) which is shown in

dashed line in Figure 2.5(b). This figure, which is representative of the behavior observed also for other participants of the mentioned experiment, shows that the visuomotor secondary task led to an increase of the eye movement occurrence. However, during the auditory secondary task, the number of eye movements has decreased and is even smaller than that of the driving parts without any secondary tasks.

Inspired by these findings, the ASR can be recalculated by excluding phases of the visual distraction as explained in the following. First, the number of eye movements is calculated and the mentioned threshold is applied to them. All parts of the drive, for which the number of eye movements exceeds the threshold, are then marked and excluded in the recalculation of the ASR. We call this signal, which controls the inclusion/exclusion of spindles, the *control* signal. Figure 2.6(a) shows ASR₆₀ of a nighttime experiment. The parts of the drive detected as visual distraction are marked.



(a) ASR with the control signal based on the increased number of eye movements



(b) ASR with the control signal based on all conditions

Figure 2.6.: ASR with different control signals

Around $t = 36$ s (shown with an arrow), as an example, the drop of the ASR is completely joint with the detected visual distraction. However, not all decreasing parts of the ASR are related to the increased number of eye movements, e.g. around $t = 45$ s or $t = 56$ s (shown with arrows). Obviously, some of the detected events are necessarily driving-task related, because a safe takeover maneuver also needs several looks in the side and rear-view mirrors. Furthermore, visual attention, which is not necessarily followed by an increased number of eye movements, should also be considered in the control signal. Therefore, we applied other criteria as well

in order to restrict the calculation of the ASR only to monotonous phases of drive with the least demanding visual attention. To this end, curves, construction sites, lane changes, braking moments can all be considered in the calculation of the control signal as shown in Figure 2.6(b).

Another possibility to make the ASR less sensitive to short-term variations is to calculate it within a longer window, e.g. 5 min (300 s) rather than 1 min (60 s). Figure 2.7 shows the ASR₃₀₀ before and after applying the control signal. The calculation of the ASR has started around $t = 10$ min and has an increasing trend up to 30 min. After that the overall trend is very weak, although the subject assessed himself as drowsy after $t = 81$ min until the end of the drive. By applying the control signal and recalculating the ASR, the overall trend did not change, on some parts it improved though, such as around $t = 70$ min.

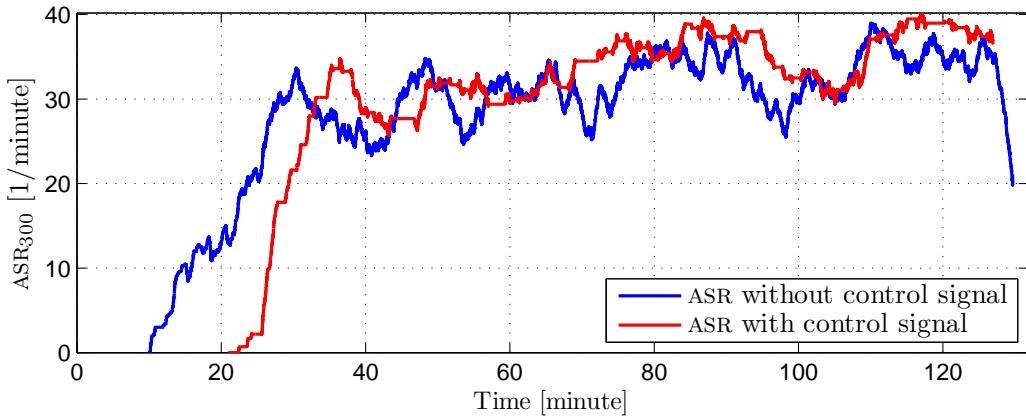


Figure 2.7.: ASR before and recalculated after applying the control signal

Since this finding is valid and representative for 30 subjects, we conclude that applying the control signal helped to improve the increasing trend of the ASR due to drowsiness for some parts of the drive. However, it could not improve the overall trend of the ASR versus drowsiness. One reason might be that we handled all non-relevant parts similarly which is not necessarily correct. A lane change event, as an example, might involve the subject and his visual attention differently depending on the situation. Apart from this, the control signal was responsible for only controlling the *visual load*. Sonnleitner et al. (2011) showed that in addition, cognitive distraction of the driver affects the ASR to a large extent as well. Unfortunately, due to their nature, the phases of the cognitive distraction are neither monitorable nor measurable for improving the control signal.

Heart activity and Respiration

By adding an extra electrode to the EEG measurement system, similar to *electrocardiography* (ECG), *heart rate* and *heart rate variability* (HRV) can be measured based on the so-called R-to-R (beat-to-beat) interval. There are some studies which believe that these measures are drowsiness indicators. During a prolonged night driving, Riemersma et al. (1977) showed the decrease of heart rate and the increase of the HRV in agreement with O'Hanlon (1972) and Lal. (2001). In the study by Lal. (2001), the drop of heart rate was quantified as up to 6 beat/min in the driving simulator. Schmidt et al. (2011) also reported a similar result in a monotonous daytime driving. Moreover, it seems that the regularity of the heart rate depends on the fact whether the subject is focusing on the task or not (Rosario et al., 2010).

Lal. (2001) and Tran et al. (2009) analyzed the power spectrum of the HRV based on its low (LF, 0.05–0.1 Hz) and high (HF, 0.1–0.35 Hz) frequency components and found correlation between these components or their ratio (LF/HF) with the drowsiness. Rosario et al. (2010) and Chua et al. (2012) also suggested the HRV for the detection of attentional failures, especially if it is fused with other measures for integration in safety systems.

In general, however, it is believed that the mentioned measures are also subject to variation due to other factors such as stress or relaxation. On the contrary to the mentioned studies, Papadelis et al. (2007) did not find any statistically significant change effects of the HRV in sleep-deprived subjects. This result was valid even by comparing the first and the last parts of the drive.

The correlation between respiration activity and drowsiness has also been studied. Rosario et al. (2010) reported 5% increase in the respiration amplitude during drowsy phases in comparison to the awake phase. Moreover, a decreased respiratory rate due to drowsiness was found by Dureman and Bodéen (1972).

Recently, new approaches are introduced for a contactless measurement of the heart rate and respiration which are all based on a camera or a radar signal, etc. (Bartula et al., 2013; Gault and Farag, 2013). Kranjec et al. (2014) reviewed non-contact heart rate measurement methods.

Eye movement analysis

Eye movement analysis is another approach for driver state detection which is the main topic of this work. Data collection can be performed by different measurement systems such as electrooculography (EOG), electromagnetic coil systems, also called search coil, and cameras (Holmqvist et al., 2011). In the following, these measurement systems are introduced based on Duchowski (2007).

■ Electromagnetic coil system

The electromagnetic coil system, the search coil and the scleral contact lens are the most precise methods for measuring eye movements, since they are directly attached to the eyes. As a contact lens, these measurement systems look like a ring which is placed over the cornea and sclera and, as a result, they are the most intrusive measurement systems. They are also believed to manipulate some eye movements.

Blinking behavior can also be measured by search coils, if they are placed around the eyes (e.g. above and below) (Hargutt, 2003). Depending on the distance between the coils, which corresponds to the distance between eyelids, electrical voltage is induced. Therefore, the eyelid gap can be measured in voltage unit and then converted into millimeter.

■ Electrooculography

Electrooculography is a popular measurement system from the EEG and ECG family which also comprises attaching electrodes directly around the eyes as shown in Figure 2.8.

According to Fairclough and Gilleade (2014), EOG benefits from the fact that eyes are considered as a dipole with its negative and positive poles at the retina and the cornea, respectively. Therefore, the eye has a static potential field under the assumption that the potential difference



Figure 2.8.: EOG electrodes attached around the eyes for collecting horizontal and vertical eye movement data

between its poles is fixed. As soon as the eyes move, the potential measured by the electrodes varies. This is exactly what EOG measures, the “*corneal-retinal potential*” (Stern et al., 2001).

As shown in Figure 2.8, for a bipolar electrode setup, in addition to reference and ground electrodes¹, four electrodes are needed. The two electrodes located at the right and left outer canthi² of the eyes collect the horizontal component of eye movements and the other two located about 2 cm above and below the eye collect eye blinks and the vertical component of the eye movements. In this arrangement of electrodes, it is assumed that the movements of both eyes are synchronous. Therefore, it is sufficient to locate the electrodes only around one eye. Since locating an electrode at the inner canthus of the eye is intrusive, the outer canthus of the other eye is usually used. This corresponds to a dipole with the negative pole on one eye and the positive pole on the other one. By moving the eyes, different poles get close to the electrodes which leads to the potential variation. The measured voltage is the difference between the potential measured at an active electrode and the reference electrode. The ground electrode, however, is used for common mode rejection (Nunez and Srinivasan, 2006). In addition to the vertical and horizontal eye movements, EOG can also record eye blinks. The reason is that during blinking, the eye ball rotates upward which also leads to the change of the dipole field. Consequently, blinks are only visible in the vertical component of the EOG. Since the occurrence of involuntary blinks (see Section 3.3) is inevitable, they can also be considered as artifacts in the capturing of voluntary eye movements.

An advantage of the EOG is its high sampling frequency (up to 1000 Hz) which makes it a very suitable system for extracting the velocity of very rapid eye movements. In addition, it provides a continuous recording with some artifacts, though. Its record is independent of almost all external factors such as wearing glasses, contact lenses or lighting conditions. Unlike cameras, it can be used in darkness.

According to Straube and Büttner (2007), EOG is subject to three types of noise:

- *inductive noise*: the residential power line or any electromagnetic field affects the recorded signal by induction and coupling into it. This noise, however, can be filtered out in the preprocessing step.
- *thermal noise*: the skin resistance and the electrode’s input resistance generate this type of noise which deteriorates the signal quality. Therefore, it is recommended to use conductive paste and to clean the electrodes and the skin before starting the measurement. Otherwise, drift might be visible in the recorded data as shown in Figure 2.9. We discuss about the methods for drift removal in Chapter 5.
- *capacitive noise*: in electrical circuits, capacitive noise refers to the noise due to the nearby

¹Reference and ground electrodes are located behind the ears.

²corner of the eye

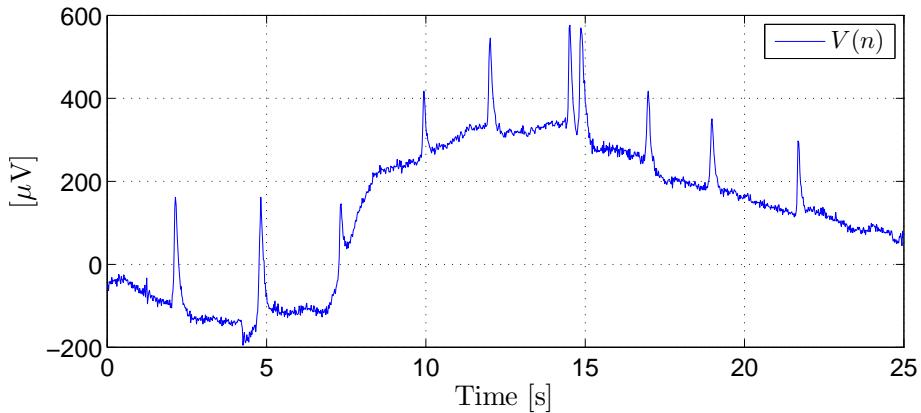


Figure 2.9.: An example of the drift in the collected EOG data (vertical component)

electronics. Similarly, in EOG, this noise corresponds to the nearby muscle artifacts like chewing.

Furthermore, another disadvantage of the EOG is its dependency to the location of electrodes. If the electrodes are placed very far from the eyes, the measured amplitudes will be smaller. Therefore, for a specific person, different eye movement characteristics might be measured at different recording times, if electrodes are located at other places than before.

Attaching electrodes directly around the eyes and injecting the paste also makes this measurement system an obtrusive one. Moreover, since EOG measures the potential resulting from eyeball movements, it can never measure the eyelid gap directly.

Since both EOG and the electromagnetic coil system measure relative eye movements, they cannot measure head movements.

By conducting a pilot study on a proving ground under fully controlled real road conditions, we investigated the robustness and reliability of the EOG measurement system for an in-vehicle eye movement data collection. Based on the achieved results and findings, which are discussed in Section 4.1, the EOG measurement system has been used in all other experiments conducted in this work for collecting eye movement data especially during real driving conditions.

■ Video-based eye-tracker

The technology of the video-based eye-tracker is based on the “*pupil-to-corneal reflection measurement*” (Holmqvist et al., 2011). It consists of an infra red (IR) illumination and a camera looking into the eye. This combination can be located either on a stable platform, e.g. on the dashboard, called *remote eye-tracker* or be mounted on the head like a helmet or glasses, called *head-mounted eye-tracker*. The disadvantage of a remote eye-tracker is its distance to the eyes which might affect the quality of the gaze tracking, if it is located very far from the eyes (Straube and Büttner, 2007). In comparison, a head-mounted eye-tracker benefits from its short distance to the eye which provides a larger number of pixels for the image processing task. It is also possible to add a scene camera to the mentioned eye-trackers, in order to know what a subject is looking at.

Independent of the technology and the manufacturer of the cameras, they typically suffer from problems listed below:

- *calibration step*: most of the cameras need to be calibrated prior to starting the measurement for each subject separately which is time-consuming. The EOG measurement system is, however, calibration-free in most of the applications.
- *off-road gazes shifts*: some cameras, especially the remote eye-trackers, cannot distinguish between real eye closures and off-road gaze shifts towards the dashboard. Unintended eye closures occur due to drowsiness, while off-road gaze shifts are a driving-related task. In both cases, the camera cannot detect the pupil and consequently false alarm rates are high in drowsiness detection applications. The mentioned issue does not exist in the EOG data, because the eyeball moves in different directions during the mentioned cases leading to distinguishable eye movements.
- *light and reflection*: as also mentioned by Friedrichs and Yang (2010a), light reflections on the glasses and unstable light conditions also deteriorate image quality and accordingly the eye movement detection quality.

On the contrary to the EOG, as an unobtrusive eye movement measurement method, cameras are able to measure the eyelid gap. Moreover, some of the cameras also measure head movement features like nodding frequency as the one used by Friedrichs and Yang (2010a).

Independent of the measurement system used for collecting eye movement data, previous research has shown that driver's eye blink features (such as blink amplitude, duration, etc.) are correlated to some extent with driver drowsiness (Hargutt, 2003; Åkerstedt et al., 2005; Schleicher et al., 2008). Hence, within a level of uncertainties they can contribute to driver drowsiness warning systems. In Chapter 7, these studies are reviewed.

Another approach for the camera-based drowsiness detection is pupillometry which studies the size variation of the pupil and its fluctuations (Rowland et al., 2005).

2.2. Subjective estimation of the drowsiness

Subjective estimation of the drowsiness, as its name says, is based on the rating of subjects about their vigilance or drowsiness level before, during and at the end of the experiment. This estimation can be done either by the subject himself or by an investigator.

Williamson et al. (2014) studied 90 drivers in a driving simulator to answer the question: “*Are drivers aware of sleepiness and increasing crash risk while driving?*”. According to this study, drivers are aware of their drowsiness level based on the access to their cognitive information. Nevertheless, they are poor in judgments about the risk of crashes due to drowsiness. This finding is in agreement with that of the Baranski (2007) with sleep-deprived subjects who showed that both subjective and objective measures were related to drowsiness. On the contrary, Moller et al. (2006) found no interaction between these two measures and concluded that subjects might suffer from lack of full insight into their degraded performance. Verwey and Zaïdel (2000) also reported disassociation between physiological and subjective measures.

Simon (2013) believes that since cognitive performance degrades in line with the increase of the drowsiness, a drowsy subject is also less able to estimate himself correctly. In fact, self-rating requires higher mental performance.

Clearly, due to its nature, subjective self-estimation of the drowsiness cannot be collected very frequently, because it affects the driver state at the time of recording, especially the monotonicity and drowsiness (Schmidt et al., 2011). It is possible to either collect it verbally, i.e. an

investigator asks the driver to rate his drowsiness level based on a pre-defined scale or by a touchscreen and pushing on the desired scale by the subject himself. Each of these variants has its own shortcomings. As an example, Schmidt et al. (2011) studied how verbal assessment of driver's state affects the vigilance during a monotonous daytime driving, i.e. task-related passive drowsiness. Implicitly, this study also discussed to what extent conversing with passengers can be considered as a drowsiness countermeasure. The results showed that verbal assessment of the drowsiness level led to an improved vigilance state which did not last longer than 2 min, though. Therefore, Schmidt et al. (2011) suggested 5-min interval data collection of subjective self-rating as an effective way to avoid contamination of drowsiness evolution. Using a touchscreen also involves some off-road gazes shifts and influences the driving performance similar to the visual distraction.

Another concern about the subjective measure is its interpretation due to its discreteness. In this context, the following issues raise:

- How should we compare the evolution of a continuous objective measure, which is collected with a higher frequency, with a less frequently collected subjective self-rating?
- Is it allowed to assume that the subjective self-rating between successive inputs remains constant or that it varies linearly/non-linearly?

Unfortunately, no consistent answers to these questions can be found in other studies. Our assumptions about the aforementioned questions will be discussed in Sections 7.1 and 8.1.1.

In the following, different scales for a subjective self-rating are introduced and discussed.

Karolinska Sleepiness Scale

The *Karolinska Sleepiness Scale* (KSS), as the most common self-rating scale (Dong et al., 2011), was first introduced by Åkerstedt and Gillberg (1990) and has 9 scales as shown in Table 2.1.

Table 2.1.: Karolinska Sleepiness Scale (KSS)

KSS	Description
1	Extremely alert
2	Very alert
3	Alert
4	Rather alert
5	Neither alert nor sleepy
6	Some signs of sleepiness
7	Sleepy, but no effort to keep alert
8	Sleepy, some effort to keep alert
9	Very sleepy, great effort to keep alert

According to Shahid et al. (2012b), this scale is sensitive to fluctuations and best reflects the psycho-physical state in the last 10 min of self-estimation. Ingre et al. (2006) and Anund (2009) as well as Sommer and Golz (2010) believed that parts of the drive with KSS ≥ 7 are mainly associated with safety-critical conditions.

There exist a lot of studies which relied on the KSS values as a drowsiness reference for evaluating other objective measures (Åkerstedt et al., 2005; Ingre et al., 2006; Fürsich, 2009; Sommer and Golz, 2010; Friedrichs and Yang, 2010a; Friedrichs et al., 2010; Friedrichs and Yang, 2010b; Pimenta, 2011). Belz et al. (2004), as an example, concluded that their studied metrics, such

as the minimum time to collision, were not drowsy indicators based on the correlation analysis with the KSS. Other studies, however, analyzed the correlation of objective measures with the KSS as an independent factor. Ingre et al. (2006) studied the relationship between the KSS and objective measures of the SDLP and blink duration. The results showed that both measures were significantly related to the KSS with a curvilinear effect. A similar result was found by Åkerstedt et al. (2005) for shift workers. Kaida et al. (2006) validated KSS against EEG features and achieved significant high correlations between them. Schmidt et al. (2009) studied physiological measures (e.g. EEG features and heart rate) under monotonous daytime driving while the subjects rated their drowsiness level based on the KSS. According to their finding, the evolution of all measures was consistent with each other, i.e. the ASR increased parallel with the increase of the subjective self-rating. Interestingly, at the last part of the drive, the KSS decreased, although all physiological measures kept their previous trends. They believed that this improved vigilance level may correlate with either the circadian effect, the intensified traffic density and even the joyful feeling that soon the experiment is over or a combination of all mentioned reasons. On the other side, however, by assuming that physiological measures correctly reflected drivers' state, they concluded that long monotonous driving (longer than 3 h) led to deterioration of the self-rating ability due to a declined vigilance level.

Different lengths of time intervals between successive KSS inputs have been reported in the mentioned studies which are listed in Table 2.2. They are varying from 2 to 30 min. Schleicher et al. (2008) used 30 min time interval, but suggested 15–20 min for future studies and interpolated the values linearly.

Table 2.2.: Literature review of the length of time intervals between successive KSS inputs

Author	time interval
Sommer and Golz (2010)	2 min
Ingre et al. (2006)	5 min
Åkerstedt et al. (2005)	5 min
Friedrichs and Yang (2010a)	15 min
Friedrichs and Yang (2010b)	15 min
Schmidt et al. (2009)	20 min
Schleicher et al. (2008)	30 min

According to Svensson (2004) and Sommer and Golz (2010), although each scale of the KSS is clearly defined, it is very probable that subjects interpret the scales inaccurately and relatively to the previous situations. This is indeed a disadvantage of the KSS. In other words, for each KSS input, it is very probable that the subjects compare their current state with the previous ones for a better self-rating. Anund (2009), as an example, even instructed the subjects to rate their drowsiness level with respect to the state in the last 5 min. Hence, depending on the preciseness of the first selected KSS value, there might be a bias shift on the other selected values until the end of the experiment. Furthermore, a subject, who due to the mentioned bias shift reaches KSS = 9 relative early, has no other choice to select during deeper phases of the drowsiness.

Stanford Sleepiness Scale

Stanford Sleepiness Scale (SSS) has 7 scales and only one of the scales should be selected at the time of query (Hoddes et al., 1973; Shahid et al., 2012c). SSS is very similar to the KSS and the description of its scales is listed in Table 2.3.

Table 2.3.: Stanford Sleepiness Scale (sss)

SSS	Description
1	Feeling active, vital, alert, or wide awake
2	Functioning at high levels, but not at peak; able to concentrate
3	Awake, but relaxed; responsive but not fully alert
4	Somewhat foggy, let down
5	Foggy; losing interest in remaining awake; slowed down
6	Sleepy, woozy, fighting sleep; prefer to lie down
7	No longer fighting sleep, sleep onset soon; having dream-like thoughts

Epworth Sleepiness Scale

Epworth Sleepiness Scale (ESS) introduced by Johns (1991) is another subjective measure which summarizes the likelihood of falling asleep during 8 different situations, such as watching TV, sitting and reading, as a passenger in the car for one hour without a break, etc. Each situation can be evaluated based on a 0–3 scale, i.e. from 0 for “*would never doze*” to 3 for “*high chance of dozing*” (Shahid et al., 2012a). The overall score of a subject is between 0 (0 for all situations) and 24 (3 for all 8 situations). Scores smaller than 10 and higher than 15 are interpreted as awake and sleepy, respectively (Čolić et al., 2014). This scale is not useful, if sleepiness should be measured repeatedly (Anund, 2009). The reason is that on the contrary to the other introduced scales which are situational, ESS gives an insight to the general tendency of the sleepiness.

In addition to the mentioned scales, which are collected based on the self-rating of the subject himself, it is also possible to rely on an expert-rating or video-labeling. Despite the fact that these are also subjective, they contain information about the subject’s state of which the subject might not be aware, e.g. a microsleep.

Expert-rating is performed either online, i.e. during the experiment or offline, i.e. based on the recorded video data. Both of the approaches mainly rely on the observable drowsiness symptoms such as yawning, heavy eyelids, improper lane keeping, etc. A more reliable expert rating is achieved, if more than one expert rate the experiment and its events and if the experts are trained using video examples to reach a common understanding (inter-rater reliability (Field, 2007)). In the end, the majority voting determines the final rating. In general, however, the quality of offline expert ratings depends highly on the quality of the recorded video data.

Schleicher et al. (2008), who studied blink behavior as a drowsiness indicator, used, in addition to a subjective self-rating measure, an offline video-rating based on symptoms such as facial gestures, blink frequency, scaring, etc. Damousis and Tzovaras (2008) also decided based on the video-labeling whether lane departures occurred simultaneously with microsleep events. In the experiment conducted by Rosario et al. (2010), an observer recorded body and face movements online.

In Chapters 5 and 7, we also rely on an offline expert-rating as the ground truth for evaluating the blink detection algorithm and the occurrence of safety-critical events, such as the lane departure and microsleep.

There exist other subjective measures such as *Visual Analog scales* (VAS) (Monk, 1989), *Crew Status Check Card and Sleep Survey Form* (Morris and Miller, 1996) and *Karolinska drowsiness score* (KDS) (Jammes et al., 2008; Hu and Zheng, 2009). Čolić et al. (2014) has reviewed many of them.

3. Human visual system

In this chapter, first, visual attention and the way the human eye operates is described. Afterwards, different types of relevant eye movements during driving are defined. Here, we concentrate on the drowsiness as the long-term distraction. Therefore, the information process in human visual system remains out of the scope of this work. This topic is of importance for the interpretation of the short-term distraction.

3.1. Visual attention

The material in this section is taken from Duchowski (2007). He explained visual attention based on “*where*” and “*what*”. The idea of “*where*” defines visual attention as roaming eyes in the space (von Helmholtz, 1925). On the contrary, with the definition of James (1981), visual attention means “*focus of attention*”, i.e. “*what*”. At the first glance, these two ideas seem to be independent. However, they support each other in a way that visual attention is only understandable, if both definitions are considered. The idea of “*where*” occurs *parafoveally* which means, at first, something roughly attracts our attention as a whole in the entire visual field. It is similar to a low resolution image. Then, the idea of “*what*” leads to the collection of more detailed information through the *foveal* vision¹. In fact, during the second step, the image will be perceived as a high resolution image. It is believed that in both steps, when the eyes move and are not fixating, the attention is turned off. There exist also some other ideas such as “*how*” which deals with the type of responses and the reaction of eyes to stimulus. Such ideas are out of our scope, though.

We discussed above about collecting full detailed information using foveal vision which is about 2° . The fovea is a narrow area with the sharpest image. Nevertheless, the information outside of it can be seen and perceived within a certain area. The size of this larger area, which is called the *functional visual field*, varies depending on the task being performed (Holmqvist et al., 2011). As an example, its size decreases in line with the increase of the cognitive load during driving.

3.2. Structure of the human eye

Figure 3.1 shows a very simple structure of the human eye. The human eye, which has a spherical shape, receives the light reflected from objects in the environment. The light rays are then bent by the cornea and refracted parallel towards the lens. Afterwards, the lens focuses the rays on the retina from where the image is then received by the rod and cone cells. They convert the image to electrical nervous stimuli and then sent them to the visual cortex for processing and interpretation.

¹Fovea is the central region of the retina with the most acute perception and the sharpest vision.

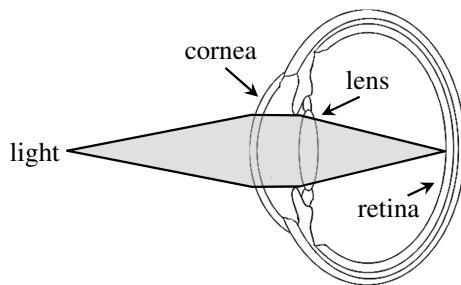


Figure 3.1.: Structure of the human eye while transmitting the ray of light

Six muscles in three pairs are responsible for moving the eye into different directions: the *medial* and *lateral recti* for sideways movements, the *superior* and *inferior recti* for up and down movements and the *superior* and *inferior obliques* for twisting. These muscles are all shown in Figure 3.2.

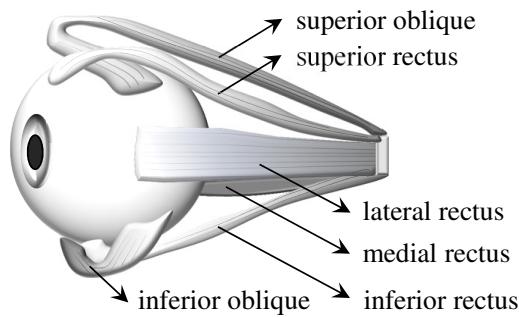


Figure 3.2.: Eye muscles

3.3. Types of eye movements

Eye movements can be categorized based on different characteristics. The first type of categorization is as follows: voluntary versus involuntary or reflexive. Another categorization is based on the velocity of the eye movements, i.e. slow versus fast eye movements as shown in Figure 3.3. Fast and slow eye movements are also called *rapid eye movements* (REM) and *slow eye movements* (SEM).

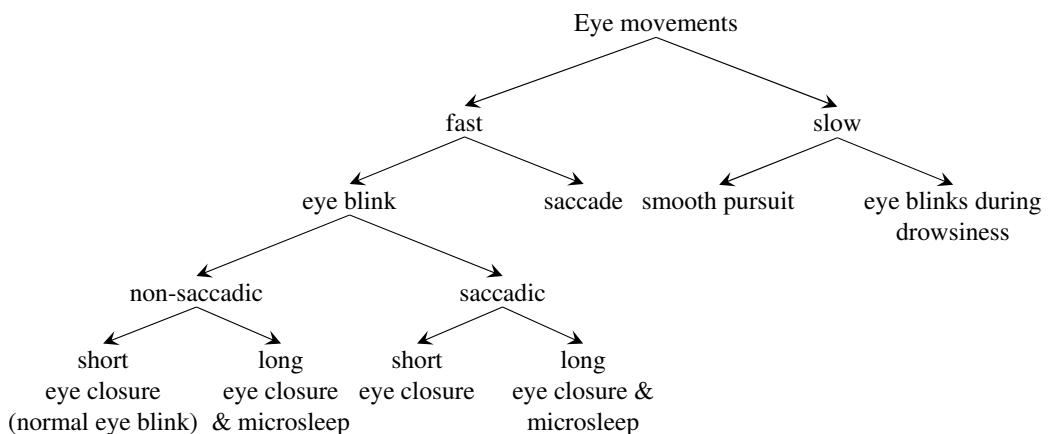


Figure 3.3.: Different categories of eye movements based on their velocity

In the following, blinks, saccades, fixation, smooth pursuit and optokinetic nystagmus as relevant eye movements during driving are defined.

Eye blinks

Regular and rapid closing and opening of the human eyes is called *eye blink* which consists of three stages: closing, closed and opening (Hammoud, 2008). Blinking occurs either voluntarily or involuntarily. Involuntary blinks can also be divided into spontaneous and reflex blinks. The former includes eye blinks which occur regularly to protect eyes against external particles. They also keep the eyes wet by spreading “*precornial tear film*” over the cornea (Records, 1979). Reflex blinks, however, are the result of an obvious, identifiable external stimulus like bright light or loud noise. Stern et al. (1984) also used similar taxonomy but under different terms: *endogenous* versus *exogenous* blinks. Exogenous blinks include reflex, voluntary blinks and long eye closures such as microsleeps. Thus, endogenous blinks are equivalent to the mentioned spontaneous blinks. In this work, however, long eye closures and microsleeps are also considered as spontaneous blinks.

Characteristics of spontaneous blinks like frequency can be influenced by factors such as vigilance, activity, emotion and tasks. Furthermore, air quality and cognitive process also affect the occurrence rate of such blinks (Stern et al., 1984). In general, performing tasks, which require visual attention, such as reading decreases the frequency of such blinks. According to Stern et al. (1984), the amount of the drop depends on the nature of the task and how demanding the task is. They have also mentioned that during performing tasks, blinks are liable to occur when the attention has decreased. Another moment, during which the occurrence of spontaneous blinks is also very probable, is the gaze shift. Gaze shifts are most often accompanied by spontaneous blinks, especially, while redirecting the attention to a new object (Records, 1979). In Chapter 6, it will be discussed how performing different secondary tasks and gaze shifts affect the occurrence of blinks. More information about the characteristics of blinks, e.g. duration, frequency, etc. will be provided in Chapter 7.

In EOG, blinking is only evident in the vertical component as shown in Figure 3.4(a). As this figure shows, during the awake phase, blinks are very sharp. Therefore, they are categorized as fast eye movements in Figure 3.3. During the drowsy phase, however, two characteristics were observed in our experiments. The first one, which is shown in Figure 3.4(b), contains eye blinks which are still fast in opening and closing motions but with longer closed duration. In fact, opening and closing phases are almost similar to the awake phase. In Figure 3.4(c), on the contrary, the blinks are much slower in opening and closing phases with slight changes in the closed duration. In Chapter 5, we will discuss different methods for detecting all types of blinks shown in Figure 3.4.

Saccades

We mentioned that the fovea consists of a very small area. Therefore, in order to see different objects sharply, their image should be projected on the fovea. This is possible by eyes in movements which are referred to as *saccades*. Saccades are fast movements of both eyes occurring due to the change of the looking direction in order to reposition the fovea from one image to another one. They can be characterized by their amplitude and duration (typically 10 to 100 ms (Duchowski, 2007)) which depend on the eyes rotation angle. The amplitude of saccades can

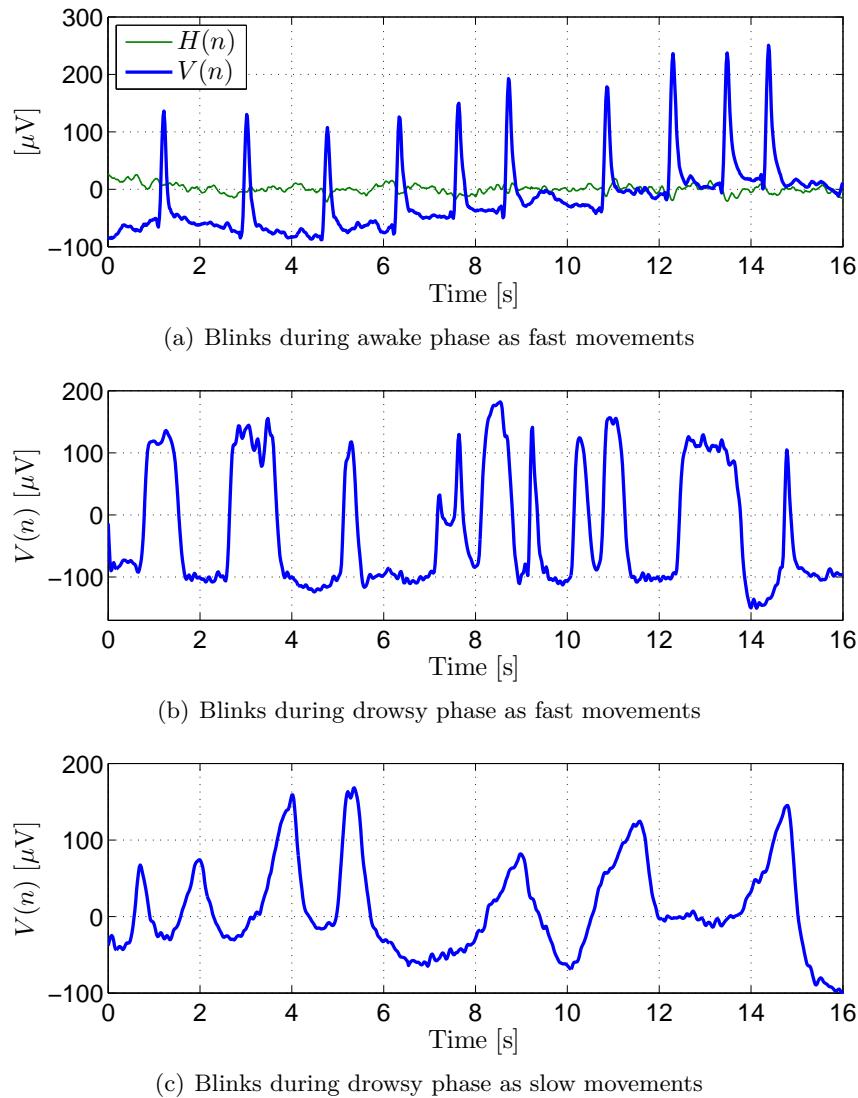


Figure 3.4.: Representative examples of blinks measured by the vertical ($V(n)$) and horizontal ($H(n)$) components of the EOG

be considered linear to the gaze angle up to $\pm 30^\circ$ (Young and Sheena, 1975; Kumar and Poole, 2002). A voluntary saccade is a saccade used for scanning the visual field. However, an involuntary saccade can be induced as a “*corrective optokinetic or vestibular measure*” (Duchowski, 2007). The very short duration of a saccade leads to a blurred image on the retina which cannot be perceived. In fact, during this period, we are *blind*. In addition, it is presumed that the distance to be traveled during saccadic movement is preprogrammed and consequently cannot be altered after being determined.

Figure 3.5 shows three examples of saccades occurring in different directions measured by the EOG. For a saccade occurring only in one direction, just one component of the EOG varies remarkably as shown in Figures 3.5(a) and 3.5(b). However, according to Figure 3.5(c), for diagonal saccades, both $H(n)$ and $V(n)$ signals are informative. Such saccades refer to the glance at the mirrors during driving. Saccades similar to Figure 3.5(b) occur while looking at the speedometer.

In Figure 3.5(c), the first saccade in both $H(n)$ and $V(n)$ is followed by a remarkable overshoot.

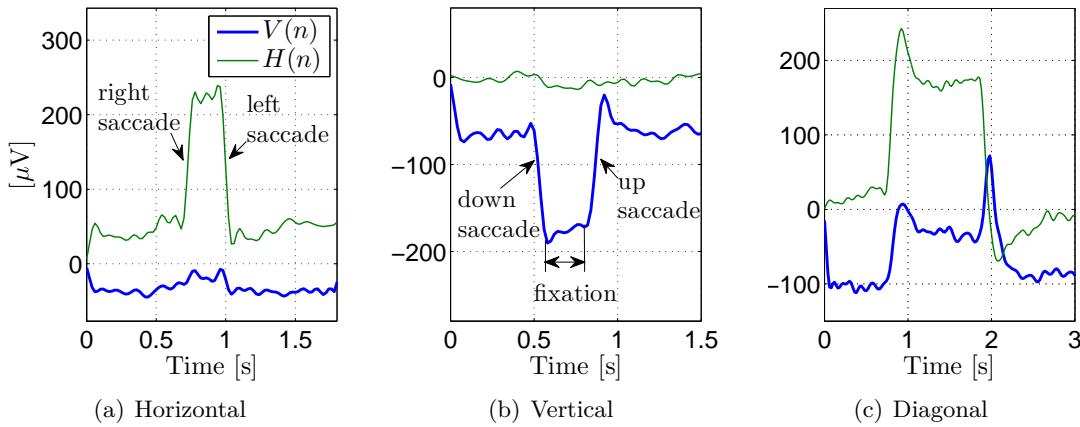


Figure 3.5.: $H(n)$ and $V(n)$ representing different types of saccades due to horizontal, vertical and diagonal eye movements

This happens, if the eye movements are time-locked to a head rotation and leads to the *vestibulo-ocular reflex* (VOR) (Sağlam et al., 2011). It is the result of a backward movement of the eyes after having reached the destination, while the head movement is not finished due to its slower velocity. The second saccade is also time-locked to an eye blink which is only present in $V(n)$. For the rest of this study, such eye blinks occurring simultaneously with a saccade will be called *saccadic eye blinks*. Other examples of such saccades are shown in Figure 5.7.

Figure 3.5 indicates that the amplitude of vertical saccades is smaller than that of the horizontal saccades. This is due to the fact that the horizontal space of the human eyes, i.e. from one corner to the other corner, is larger than that of the vertical one, i.e. from upper lid to the lower lid. Therefore, the eyes travel a larger distance in the horizontal direction.

Comparing the long eye closures of Figure 3.4(b) with saccades shown in Figure 3.5, it can be seen that both eye movements have similar shape and forms, although they occur in totally different situations. Unintended long eye closures are a drowsiness indicator, while saccades occur during scanning the visual field. In chapter 5, we suggest a method to distinguish them from each other.

Fixation

The time interval between two successive saccades, during which the eyes fixate on a new location, is called *fixation* (Figure 3.5(b)). Fixation is defined in ISO 15007 (2013) as the “*alignment of the eyes so that the image of the fixated area of interest falls on the fovea for a given time period*”. Fixation on an object can also be interpreted as focusing the attention on that object or *visual intake*. Nevertheless, this is not always the case, such as during the “looked but failed to see” phenomenon (Holmqvist et al., 2011). Fixation can also be considered as *miniature* or *micro* eye movements such as microsaccades.

During a fixation, following tasks are processed: analysis of the image on the fovea, i.e. processing available visual information, next fixation location and pre-programming of the following saccade (ISO 15007, 2013). These tasks might not be completed thoroughly during the fixation period which leads to some corrections by looking back to the previous location. As a result, a minimum duration of 100 or 150 ms has been assumed for the fixation.

At the first glance, it seems that fixation duration reflects the complexity of the task being performed and the depth of the cognitive process (Holmqvist et al., 2011). However, other factors like stress and daydreaming affect the fixation duration as well.

Smooth pursuit

Smooth pursuit describes the slow eye movements while tracking a moving object with the same velocity (up to $30^{\circ}/s$) as it moves (Leigh and Zee, 1999). During driving this eye movement occurs while fixating on any moving or non-moving object outside the vehicle.

Optokinetic nystagmus

The combination of a smooth pursuit followed by a saccade (without head movement) leads to a sawtooth-like pattern called *optokinetic nystagmus* (OKN). According to Young and Sheena (1975), OKN consists of phases with low and high velocities (slow and fast phases) in opposite directions. During the slow phase, the eyes fixate on a portion of a moving object while following it (smooth pursuit). However, during the fast phase, since that portion had moved out of the field of vision, by a correcting saccadic jump in the opposite direction, the eyes move back to the previous position to fixate on a new portion of the moving field. This type of eye movement also occurs during driving and will be studied in Section 4.1.5. An example of OKN is shown in Figure 4.8.

4. In-vehicle usage of electrooculography and conducted experiments

As mentioned in Section 2.1.2, EOG can be a suitable measurement system for collecting eye movement data. On the contrary to cameras, it does not need any calibration and is not affected by varying lighting conditions. However, it should be attached directly around the eyes which makes it impractical as an in-vehicle product for sale.

In this chapter, first, a pilot study is explained in Section 4.1 which concentrates on the application of the EOG measurement system in the field of automotive. This application is clearly different from using EOG in the laboratory or in fixed-base driving simulators. Hence, the pilot study evaluates the robustness of the EOG measurement system for in-vehicle applications by exploring road-dependent eye movements. Based on the achieved results of this study, EOG is used in other experiments of this work as well for collecting eye movement data. The conducted daytime and nighttime experiments for studying driver drowsiness detection are described in detail in Sections 4.2 and 4.3. These experiments have been designed such that they are representative of awake and drowsy driving scenarios in real life.

4.1. Eye movement measurement during driving - a pilot study

In this section, we introduce the conducted pilot study and discusses its results which are mostly based on Ebrahim et al. (2013b). The goal is to know whether the EOG measurement system is reliable and robust for collecting eye movement data during real driving conditions.

To explore the in-vehicle usage of the EOG measurement system, in a fully controlled experiment, the relationship between driver eye movements and different real driving scenarios is investigated by the EOG signals. In order to be able to reproduce similar conditions several times, all data of the pilot study was collected on the proving ground with a total of eight expert drivers (29–58 years, mean: 39.9 years, all male). All subjects were accompanied by an investigator during the experiment. Based on the collected data, it is explored, if and how driver eye movements can be influenced by the road structure, independent of the driver's drowsiness or distraction. In other words, we determine which eye movements during driving are road- or situation-dependent. Such results are, in general, of interest for any driver monitoring system with concentration on eye movements.

4.1.1. Material

Electrooculography measurement system

In the pilot study and all other experiments, which were conducted within the framework of this work, EOG signals were collected at 250 Hz by the measurement system ActiCAP, Brain

Products GmbH. Horizontal and vertical components H and V of the EOG were measured by four electrodes located around the eyes as shown in Figure 2.8. H and V were defined as follows

$$V = \text{electrode}_{\text{above}} - \text{electrode}_{\text{below}} \quad (4.1)$$

$$H = \text{electrode}_{\text{right}} - \text{electrode}_{\text{left}}. \quad (4.2)$$

Two electrodes were also used as reference and ground which were located on the bone behind each earlobe. $H(n)$ and $V(n)$ refer to the collected signals of both components at sample n . For further analysis, both $H(n)$ and $V(n)$ were sampled down to 50 Hz.

Vehicle

A Mercedes-Benz E-Class (W212) with an automatic gearbox equipped with the explained EOG measurement system was used for this experiment. In addition, a multitude of vehicle-related measures such as vehicle speed, lane lateral distance, steering wheel angle and global positioning system (GPS) were recorded synchronously at the sampling frequency of 50 Hz. During the entire experiment, four IR-cameras were installed in the car to record videos from the driver's face, the vehicle interior, the road ahead and behind synchronously. Therefore, it was always possible to analyze the driving sections offline.

4.1.2. Test tracks

Since driver eye movements and vehicle speed can highly be influenced by the traffic density during the experiment, we decided to collect our data under a fully controlled situation on Applus+ IDIADA proving ground (Applus+ IDIADA, 2014). This also helped to omit disturbing maneuvers, e.g. lane changes and takeover maneuvers. In addition, in a fully controlled experiment, it would be easier to investigate the reproducibility of the results by different subjects.

The selected tracks (see Figure 4.1), tasks and the concepts behind them are explained below. All tracks have been driven two times by each subject.

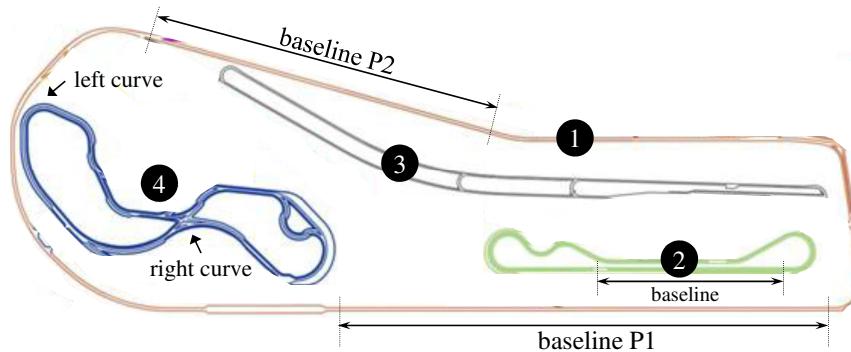


Figure 4.1.: Selected tracks of the Applus+ IDIADA proving ground (Applus+ IDIADA, 2014)

Track 1: General road

This track consists of two straight parts (P1, P2) which have been used for baseline driving in our experiment. During baselines, the subject was asked to look at the horizon and keep the head steady. Baselines were driven with the *adaptive cruise control* system adjusted at 100 km/h.

The parts of the track, which are not shown as baselines in Figure 4.1, were not driven according to the mentioned conditions. Measurements of this track are our references for assessing other tracks' data.

Track 2: Fatigue¹ track

Track 2 mimics a badly maintained road with many ground excitations leading to a lot of head movements during driving. This track was paved with setts and also contains a straight part which has been used as a baseline condition similar to track 1. The whole track 2 was driven at the maximum speed of 50 km/h. The experiment on such a track targets the answer to the question, if and to what extent the mentioned head movements cause EOG signal degradation.

Track 3: Straight line braking surface

On the contrary to track 2 with permanent ground excitations, this track contains temporary speed bumps with different sizes and shapes. Different categories of bumps were successively repeated 5 times on the track. The straight parts of track 3 were driven with the adaptive cruise control system at 80 km/h. Based on the data collected in this track, we study whether hitting a bump leads to unwanted vertical saccades or blinks.

Track 4: Dry handling track

We have also chosen a curved track, driven with the adaptive cruise control system at 80 km/h, to study the dependency of eye movements on the road curvature. The minimum curve radius of the track was roughly 50 m. It should be mentioned that track 4 is a wide track which makes turns at high speeds in larger radii than the curve's radii possible.

4.1.3. Comparing baselines of track 1 and track 2

We consider the two baselines of track 1 as the reference measurement which reflect the common behavior of the eyes and the intrinsic noise of our measurement system. Top and bottom plots of Figure 4.2 represent 30 s of $H(n)$ and $V(n)$ signals of subject S2 for the baseline P1 of track 1, respectively. The peaks in $V(n)$ imply the eye blinks. Figure 4.3 shows also 30 s of $H(n)$ and $V(n)$ related to the baseline of track 2 for the same subject. The high frequency variations of the EOG signal in Figures 4.2 and 4.3 are due to the mentioned intrinsic noise of the measurement system. Visual inspection of these two figures reveals that there exist low frequency changes both in $H(n)$ and $V(n)$ of track 2. They correspond to the slight compensatory movements of the eyes trying to keep the gaze direction concentrated on the horizon. In fact, the subject has experienced *unwanted* head vibrations, while fixating his eyes. In the EOG signals, however, it seems as if the eyes have had movements.

For the frequency analysis of baselines of track 1 and 2, we choose 20 s of $V(n)$ of subject S4 who had the longest eye blink-free time span among the subjects. The spectrograms and power spectral densities (PSD) of $V(n)$ of these baselines are shown in Figure 4.4. The difference between spectrograms/PSDs seems to be considerable within 0.5–2.5 Hz range in comparison to other frequencies which might correspond to the unwanted head vibrations of track 2. Moreover,

¹The term fatigue refers to its definition in the field of civil engineering and not to the driver drowsiness.

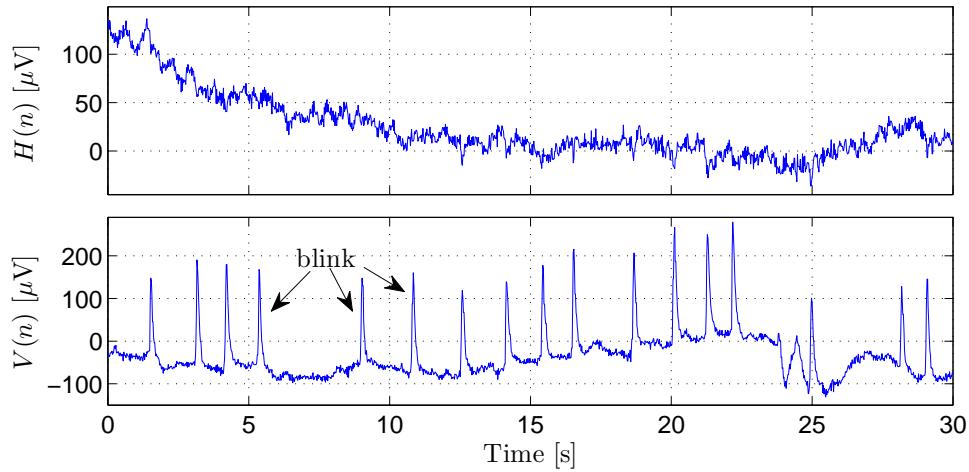


Figure 4.2.: $H(n)$ and $V(n)$ of subject S2 for baseline P1 of track 1

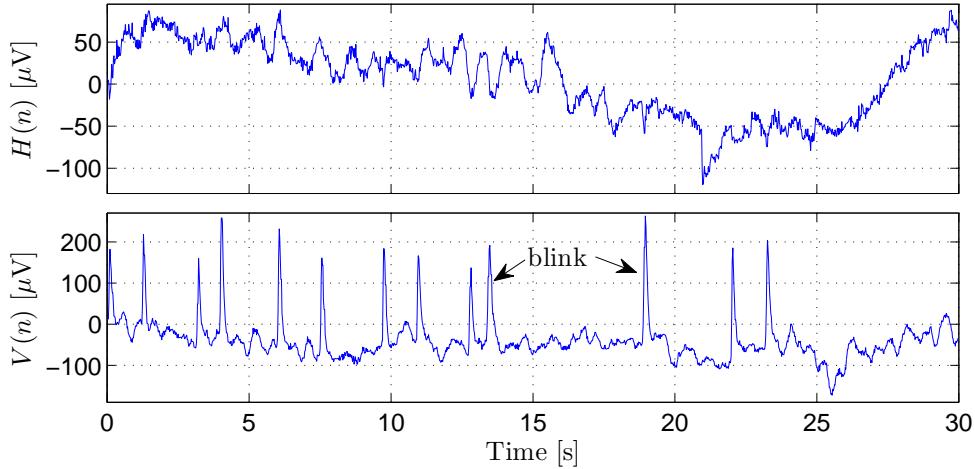


Figure 4.3.: $H(n)$ and $V(n)$ of subject S2 for baseline of track 2

similarity of the two PSDs in higher frequencies implies that ground excitation does not provide any artifacts in this range, i.e. electrodes do not move or vibrate despite the ground excitation.

In order to quantify the effect of ground excitation versus the normal track for all subjects, a moving standard deviation filter with the window length of 0.5 s and overlap length of 0.48 s was applied to $H(n)$ to quantify the variations in both baselines. In other words, at each time sample n , the standard deviation of the samples within the last 0.5 s (25 samples at 50 Hz) were calculated. We used moving standard deviation filter, because it describes best the local variation of the EOG due to ground excitation. Afterwards, the mean of the calculated values is used for further analysis of each baseline. The difference of the two mean values of the calculated standard deviations for all subjects represents the level of the contribution of head vibration in the eye movement signal. In order to avoid a false calculation, possible saccades were removed at first from $H(n)$. This was done by applying the saccade detection algorithm which will be explained in Section 5.2. All calculated means of standard deviations are listed in Table 4.1. Data of subject S3 has been excluded, as he did not follow the baselines' instructions. Figure 4.5 also shows the boxplots of the values listed in Table 4.1. Interestingly, most of the mean values of track 2 are larger in comparison to those of the track 1 regarding all subjects. Therefore, it can be concluded that ground excitation leads to unwanted head vibrations which can be measured by EOG signals, although the eyes have not moved. Consequently, EOG is

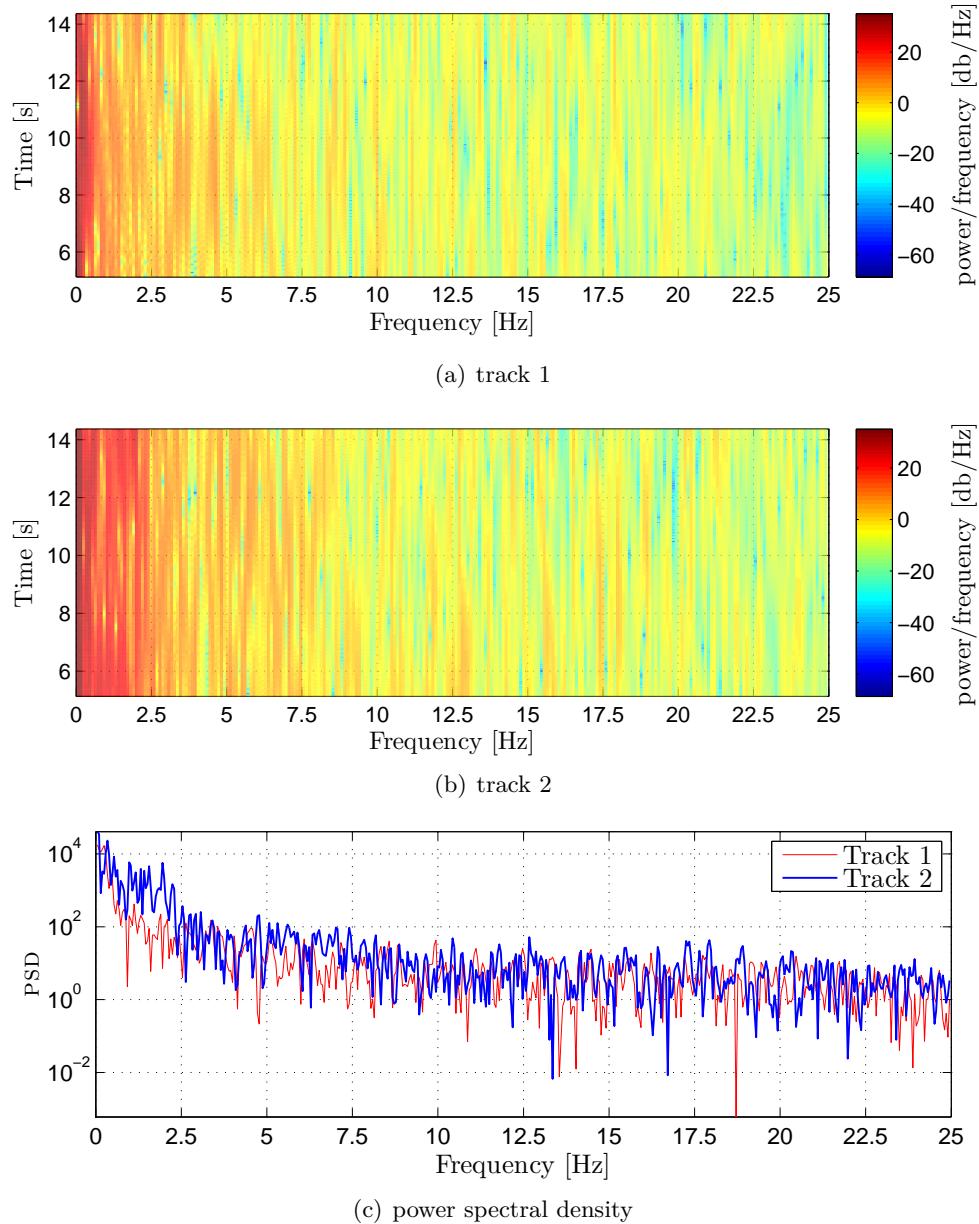


Figure 4.4.: Spectrogram and power spectral density (PSD) of 20 s of $V(n)$ of subject S4 for track 1 and track 2

not a suitable measurement system for the collection of eye movements on roads with ground excitation. This was, however, not the case for the highways used in other experiments of this work.

4.1.4. Bumps and eye movements: related or unrelated?

For assessing the impact of road bumps on eye movements, the exact locations of the bumps of track 3 should be extracted. In order to determine them, the wheel speed sensor data was used which was recorded synchronously with EOG signals. By calculating the *exponentially weighted moving variance* (EWVAR) $\sigma^2(n)$ (Friedrichs and Yang (2010a)) of the wheel speed $w(n)$ and

Table 4.1.: Means of moving standard deviations of $H(n)$ for all rounds (R) and parts (P) of track 1 and track 2, for all subjects (excluding subject S3)

Subject	Track 1				Track 2	
	R1,P1	R1,P2	R2,P1	R2,P2	R1	R2
1	7.3	8.7	9.5	9.2	10.8	8.9
2	6.3	7.4	6.5	6.6	9.8	9.9
4	5.5	5.8	6.8	5.3	7.0	6.4
5	6.3	7.9	6.8	5.2	7.3	7.4
6	6.2	6.3	5.8	6.3	7.9	7.2
7	5.7	7.6	5.5	6.6	14.1	11.7
8	8.3	11.2	7.4	8.5	14.3	12.0

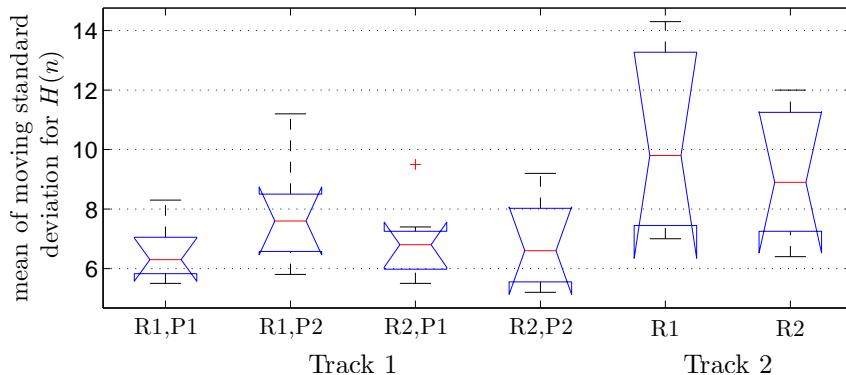


Figure 4.5.: Boxplot of the values listed in Table 4.1

applying a threshold (5 (rpm)^2) to it, we distinguished between bumps and even sections of the track. $\tilde{\sigma}^2(n)$ is calculated as follows

$$\tilde{\sigma}^2(n) = \lambda_{\sigma^2} \tilde{\sigma}^2(n-1) + (1 - \lambda_{\sigma^2}) (w(n) - \tilde{\mu}(n))^2 \quad (4.3)$$

$$\tilde{\mu}(n) = \lambda_{\mu} \tilde{\mu}(n-1) + (1 - \lambda_{\mu}) w(n) \quad (4.4)$$

$$\lambda_{\mu} = \frac{N_{\mu} - 1}{N_{\mu}} , \quad \lambda_{\sigma^2} = \frac{N_{\sigma^2} - 1}{N_{\sigma^2}} , \quad (4.5)$$

where $\tilde{\mu}(n)$ is the *exponentially weighted moving average* (EWMA) of $w(n)$. λ_{σ^2} and λ_{μ} are the forgetting factors adjusted by window size $N_{\sigma^2} = N_{\mu} = 3$ samples. The initial values of $\tilde{\sigma}^2(0)$ and $\tilde{\mu}(0)$ were set to the average of $w(n)$. Figure 4.6 shows five detected large amplitude bumps with the height of about 8 cm based on $\tilde{\sigma}^2(n)$.

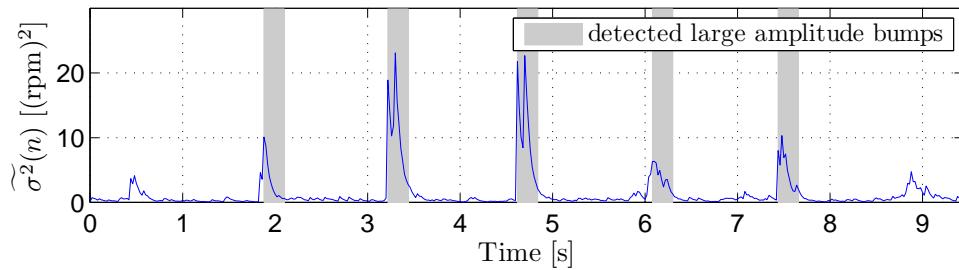


Figure 4.6.: Detected large amplitude bumps based on the EWVAR of wheel speed sensor data

Figure 4.7(a) shows the detected time intervals of multiple single small amplitude bumps (with

the height of about 5 cm) over $V(n)$ for subject S2. Interestingly, as shown in this figure and valid for all subjects, we did not find any considerable distortion in the EOG signal. This can be justified by two reasons. One reason might be that in this experiment, we used a vehicle with above normal damping and ride comfort and consequently the disturbing effect of small amplitude road bumps was filtered out by the vehicle to a large extent. Moreover, our body acts as a low pass filter (damper) to compensate the excitement of such small and short-duration bumps. However, as Figure 4.7(b) shows the influence of bumps with larger amplitudes and longer durations on EOG is very similar to that of track 2, i.e. occurrence of low-frequency components. This is due to successive repetition of each bump category (each 5 times).

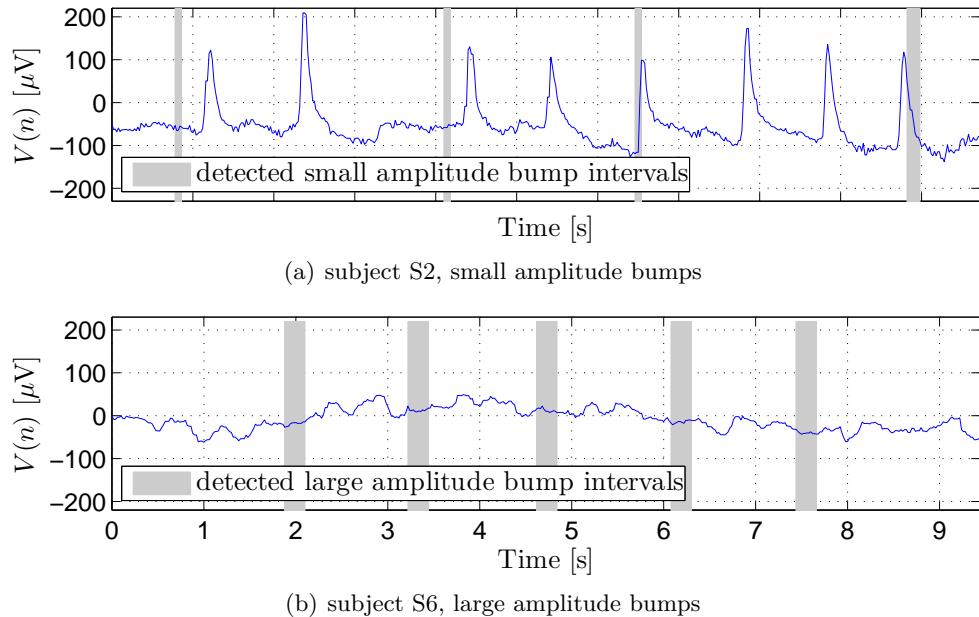


Figure 4.7.: $V(n)$ of subject S2 and S6 for small (top plot) and large (bottom plot) amplitude bumps of track 3

4.1.5. Patterns of eye movements due to road curvature

During the investigation of EOG signals in the curves of track 4, we observed a peculiar pattern of the EOG signal (see Figure 4.8) which is similar to the sawtooth pattern of the OKN described in Section 3.3. Figure 4.8 shows this pattern for two right and left curves of track 4 (top and bottom plots, respectively). These two curves are labeled in Figure 4.1, too. According to Figure 4.8, it can be concluded that the sawtooth pattern direction is related to the curve direction. Moreover, the occurrence frequencies of the sawtooths and the amplitudes of the fast phases (saccades) in the right curve are smaller than those of the left curve. These results, which are valid for all subjects, might be related to different vehicle speeds, different positions of the subjects in the vehicle and the radii of both curves. The vehicle speed for all subjects was almost the same on this track due to the used adaptive cruise control. In addition, we assume that the difference between the positions of the subjects in the car is negligible. The radii of the curves, however, are to a larger extent different from each other and are roughly 150 m and 50 m for the right and left curve, respectively. We observed that the sawtooth pattern was more pronounced for curves with higher curvature.

Many research studies were devoted to interpreting the pattern of eye movements in curve

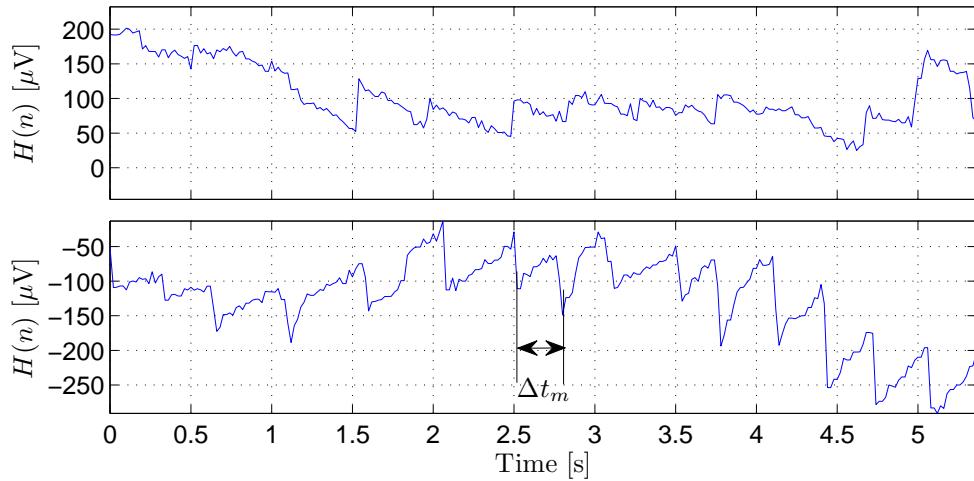


Figure 4.8.: $H(n)$ of subject S8 for track 4, top/bottom: right/left curve

negotiation. Jürgensohn et al. (1991) stated that similar to OKN, during curve negotiation, the subject fixates on a point in front of the car for a certain period (slow phase). As soon as the point is close to the car, the subject chooses another fixation point (fast phase). Based on Land and Lee (1994), this moving point can be referred to the tangent point (TP) of the curve or a point very close to it. As stated by Kandil et al. (2010) and Authié and Mestre (2011), this point describes the intersection of a tangent through the vehicle and the inner lane marking. They also believed that subjects rely heavily on the TP and suggested it as a useful source of information for correct steering during curve negotiation. According to Authié and Mestre (2011), the location of this point is dependent of the lateral position of the vehicle towards the lanes and the vehicle's orientation. Considering all of our results and observations, the sawtooth pattern rarely occurred for curves with larger radii, meaning that the subject does not necessarily follows the TP or any other point, while negotiating curves with lower curvature.

Since our experiment was not equipped with a head-mounted camera, it is not clear where the subjects were looking at during curve negotiation and whether the OKN pattern of Figure 4.8 is the result of reliance on the TP as the informative point. This can be clarified, if the characteristics of sawtooth pattern, e.g. the average time interval of its occurrence Δt , can be described as a function of the curve radius r , under the assumption that TP has been observed continuously during curve negotiation. In Appendix A, this relationship is investigated analytically for the left curve shown in Figure 4.1 with $r = 50$ m. We denote the analytically calculated values of Δt with Δt_c . Furthermore, the Δt_m value is available from our measured data as shown in Figure 4.8. By evaluating the calculated Δt_c against the measured Δt_m , we can show whether subjects relied on the region of the TP during the curve negotiation.

According to Appendix A, Δt_c is a function of curve radius r and the angular displacement of the eyes δ , i.e. $\Delta t_c = f(r, \delta)$. Therefore, for the calculation of Δt_c , the value of δ is needed. In order to find δ , we conducted an experiment in a stationary position, i.e. no driving, to investigate the relationship between δ and the saccade amplitudes of OKN pattern. There, the subject was asked to look at points located horizontally between $\pm 45^\circ$ with 5° spacing at the height of the eyes (without head movement). 0° was calibrated with respect to subject's position while looking straight ahead. The results agree with findings of Young and Sheena (1975) and Kumar and Poole (2002) who stated that the amplitude of saccades is considered linear to the gaze angle up to $\pm 30^\circ$. We measured that the saccades within the amplitude range of the sawtooths of Figure 4.8 (roughly $50 \mu\text{V}$) have happened by eye movements between 5° to 10° , i.e. $5^\circ \leq \delta \leq 10^\circ$.

Given $\delta = 4^\circ, 6^\circ, 10^\circ$ and $50 \text{ m} \leq r \leq 150 \text{ m}$, possible values of Δt_c are shown in Figure 4.9. As mentioned before, to evaluate Δt_c , we compared it with Δt_m . According to Figure 4.8, $\Delta t_m \approx 0.34 \text{ s}$, which agrees with the calculated values of Δt_c , namely $0.3 \text{ s} < \Delta t_c < 0.4 \text{ s}$ for $r = 50 \text{ m}$ and $4^\circ \leq \delta \leq 10^\circ$. This implies that relying on the region of the TP was the case during the curve negotiation of $r = 50 \text{ m}$ at $v = 80 \text{ km/h}$ in our experiment. Moreover, this result states that tracking the TP is also a possible reason for the existence of OKN in our data.

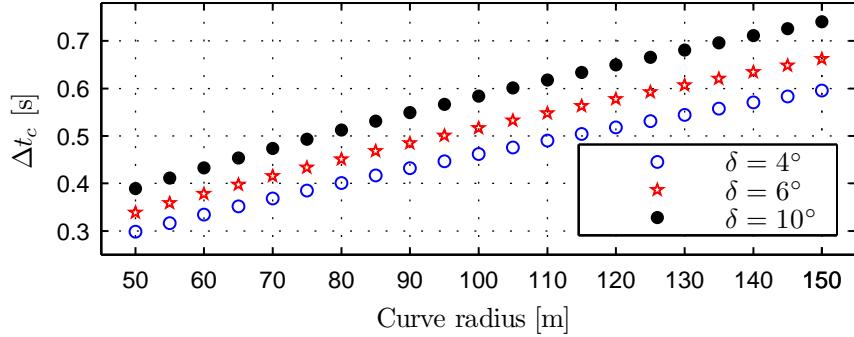


Figure 4.9.: Calculated sawtooth occurrence time period Δt_c versus curve radius r for $\delta = 4^\circ, 6^\circ, 10^\circ$

By analyzing the $V(n)$ signal, we realized that the occurrence of OKN was not recognized within the vertical eye movements for all subjects. For example, during the occurrence of the thoroughly apparent sawtooth pattern in $H(n)$ of subject S8 (Figure 4.8), no vertical eye movements were observed in $V(n)$. On the contrary, for subject S4, the vertical and horizontal components of EOG revealed similar patterns.

It is clear that the observed inevitable sawtooth pattern due to curve negotiation is not related to the driver's inattention or drowsiness. Therefore, we suggest the exclusion of tortuous road sections for further investigation of driver eye movements. Whether features of the OKN depend on the curve parameters and vehicle velocity is, however, out of scope of this work and opens rooms for future work.

Conclusion

We studied the relationship between driver eye movements and different real driving scenarios by EOG signals in a fully controlled experiment. In this pilot study, we explored whether and how driver eye movements are influenced by the road properties, independent of driver's drowsiness or distraction. In addition, the usage of EOG measurement system for in-vehicle applications was studied.

It can be concluded that ground excitation and large amplitude bumps add an extra pattern to the EOG signals which was characterized as a low frequency component. On the other hand, monitoring driver eye movements seems to be undisturbed by a single small amplitude bump. Moreover, the sawtooth patterns of OKN during curve negotiation are not drowsiness-related. Therefore, considering all results provided in the pilot study, we conclude that EOG is a robust and reliable measurement system for the collection of eye movement data on real roads and highways. Therefore, in other experiments explained in the following sections a similar measurement system has been used, because the parts of the highways selected for conducting the real road experiments were all free of ground excitation, large amplitude bumps and very high curvature.

4.2. Real road experiments

The real road experiments are the experiments conducted on real roads and cover the collection of data sets related to the awake and drowsy driving in this work.

4.2.1. Daytime driving with no secondary tasks

Two daytime drives with no secondary tasks have been conducted at different times. The first one was conducted in May 2012 and the second one in September 2013. Since, except for some additional measurement instruments, both experiment systems and procedures were the same, we combined the collected data and call them as the *daytime drive with no secondary tasks*.

Subjects

In total, 18 voluntary subjects, 3 females and 15 males, with an average age of 41.1 ± 10.7 years (27–62 years) participated in the experiment who were all the employees of Daimler AG. All of the subjects were additionally trained for driving the experiment vehicle with a lot of measurement systems. These subjects are labeled S26 to S43 in Chapters 7 and 8.

Material and experiment procedure

The same Mercedes-Benz E-Class used for the pilot study was also used for this experiment. It was equipped with an EOG and ECG measurement system and a head tracker. For 9 subjects a vital camera for measuring the heart and respiration rate and for the other 9 subjects a driver observation camera were additionally installed in the car. Only EOG data has been analyzed in this work. In addition, lots of vehicle-related measures such as vehicle speed, lane lateral distance, steering wheel angle and GPS were also recorded at the sampling frequency of 50 Hz. The KSS was also collected every 15 min via a touchscreen which was prompted automatically by a beep tone. After rating the drowsiness level, the subjects should have also answered a question about the acceptance of a drowsiness warning with either *correct*, *acceptable* or *false*. During the whole experiment, four IR-cameras were installed in the car to record videos from the driver's face, the vehicle interior, the road ahead and behind synchronously. Therefore, an offline analysis of all driving sections was always possible.

The A81 highway route in Germany was selected for this experiment as shown in Figure 4.10. Since our goal was to collect EOG data related to alert and less drowsy phases of the drive, the experiment was conducted at 9 a.m. or at 1 p.m. to have more fit subjects and low highway traffic. On average each subject drove about 260 km unaccompanied. All subjects were asked not to perform any secondary tasks such as listening to the radio, operating the navigation system or talking with mobile phone and to obey all traffic rules, while not driving faster than 130 km/h. They were also allowed to use the adaptive cruise control.

4.2.2. Daytime driving with secondary tasks

This experiment was initially designed for studying the variation of the ASR during visuomotor and auditory secondary tasks under real driving conditions by Sonnleitner et al. (2011). Since the EOG data was collected as well, we used it in this work for studying gaze-shift induced blinks

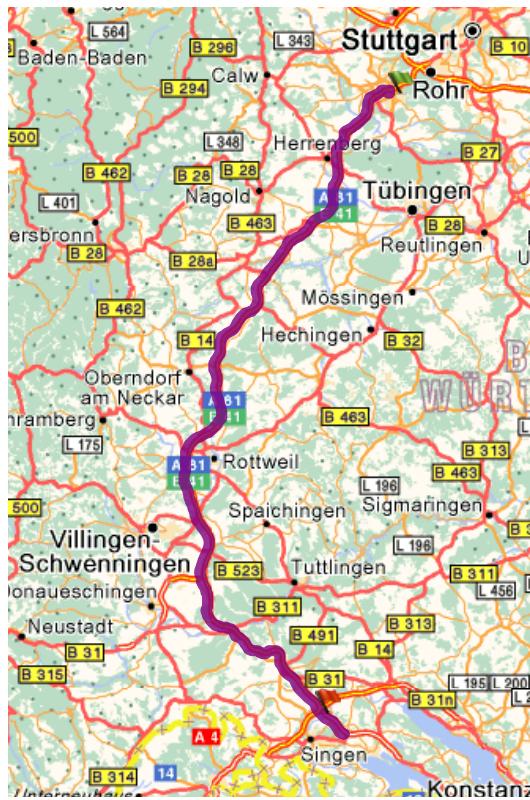


Figure 4.10.: Daytime drive experiment's route (ViaMichelin, 2014), about 130 km

as will be discussed in Chapter 6. The following explanations are derived from Sonnleitner et al. (2011) and Ebrahim et al. (2013c).

Subjects

A total of 26 voluntary employees of Daimler AG, 7 females and 19 males, participated in the experiment with the average age of 43.7 ± 8.7 years (25–56 years). All of the subjects were additionally trained for driving the experiment vehicle. The participants of this experiment were partially different from those of the previous experiment. These subjects are labeled S1 to S26 in Chapter 6.

Vehicles

Two Mercedes-Benz S-Class vehicles (W221) and one E-Class vehicle (W212) were used in this experiment equipped with an extra brake and gas pedal on the passenger seat similar to driving school vehicles. This was done due to safety reasons during performing the secondary tasks. All vehicles were also equipped with the EEG (16-electrode-cap) and EOG measurement systems similar to the previous experiment. IR-cameras were installed as well. Vehicle-related measures were also collected as explained before.

Experiment procedure

During the experiment, the subjects performed the primary driving task together with four blocks of secondary tasks lasting 40 min on the same highway route as in the previous experiment (see Figure 4.10). The secondary tasks contained both visuomotor (representative of navigation system demands) and auditory (comparable with the mobile phone conversation) tasks. All subjects were instructed to always prioritize the primary task and to drive under official traffic regulations. The maximum speed allowed was 130 km/h. During the tasks, no overtaking maneuver was allowed for safety reasons. Moreover, a trained investigator accompanied subjects during the experiment to intervene in case of safety-critical situations using the extra pedals.

Each block contained 3 min of visuomotor task, 1.5 min of driving with no secondary task, 3 min of auditory task and finally 1.5 min of driving with no secondary task as shown in Figure 4.11. Start and end markers of each block were recorded automatically. This means that the time gaps before and after the beginning and end of blocks were discarded.

visuomotor	driving	auditory	driving
3 min	1.5 min	3 min	1.5 min

Figure 4.11.: One block of daytime real road experiment with secondary tasks

Visuomotor secondary task

For the visuomotor task, a 2×2 matrix of four Landolt rings was shown on a display located at the central console on the right side of the navigation system (see Figure 4.12). The subject had to determine which side of the screen (right or left) contains the ring with different direction of opening by pushing on two adjacent buttons of an external number keypad (4: left, 6: right) located within driver's reach on the lower central console. In this work, the number of correctly identified rings is not evaluated, because just the gaze shifts between the road and the screen are of interest as will be discussed in Chapter 6.

Auditory secondary task

During this task, the subjects listened to an audio book, in order to detect the German definite article “die” by pressing a button fitted to their left index finger. At the end of each block, subjects answered a question about the content of the presented audio book. Again, the answers are not evaluated here.

Figure 4.12 shows the experimental setup of the mentioned secondary tasks.

4.2.3. Nighttime driving with no secondary tasks

This experiment was conducted in March 2010 to collect drowsiness related data under real driving conditions for studying EEG-based features. Since EOG electrodes were used as well, the experiment is studied in this work.

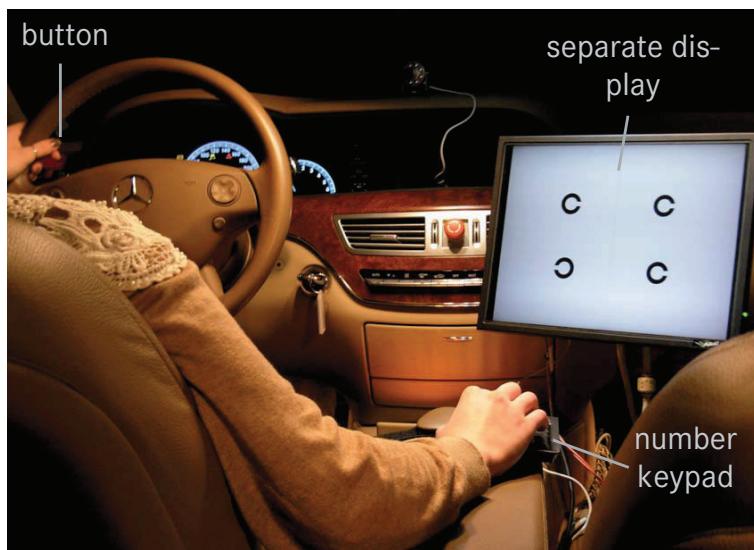


Figure 4.12.: In-vehicle setup of the daytime experiment with secondary tasks (taken from Sonnleitner et al. (2011))

Subjects

In total, 46 voluntary subjects, who were all employees of Daimler AG, participated in this experiment. Data of 16 subjects was removed due to technical problems for collecting EEG or other sensor's data. Out of the 30 remaining subjects, 14 of them aborted the experiment due to severe drowsiness which can be considered as an objective measure reflecting subjects' deep level of drowsiness. Only the data of these 14 subjects is of interest in this work. Due to data quality problems of EOG signals, however, in the end only data of 10 subjects out of 14 (1 female and 9 males) are studied with the average age of 35.9 ± 10.1 years (24–57 years). These subjects did not participate in other experiments of this work.

Material and experiment procedure

The vehicles used in this experiment were all equipped similarly as in the previous experiments. The selected route, shown in Figure 4.13, was from Stuttgart to Würzburg via Ulm and driven directly back to Stuttgart again (not on the same road as before).

The driving task started around 10 p.m. The Mercedes-Benz E-Class vehicle used in the previous experiment together with an S-class vehicle equipped similarly were used for collecting the data. In addition, the secondary pedals were installed in the vehicle. The reason was that all subjects were accompanied by an investigator who was responsible for intervening and controlling the vehicle in case of safety-critical events. All subjects were told to abort the experiment, if they felt drowsy. Every 15 min, KSS data was collected as well. As a rule, as soon as a subject estimated his drowsiness level at $KSS = 9$ or two times at 8 successively, the experiment was aborted. On average, a subject aborted the experiment after driving 244 km. Similar driving regulations as explained in Section 4.2.1 were held in this experiment.



Figure 4.13.: Nighttime drive experiment's route (ViaMichelin, 2014), about 450 km

4.3. Nighttime driving experiment in the driving simulator

Since severe drowsiness phases and occurrence of microsleeps cannot be induced in real road experiments due to safety concerns, drowsy data was collected at the Mercedes-Benz moving-base driving simulator. With a 360° projection screen, it is to a large extent comparable with real driving (Zeeb, 2010). This experiment was conducted for collecting the most relevant eye movements to drowsiness.

Subjects

25 employee of Daimler AG, 11 females and 14 males, with an average age of 33.9 ± 8.0 years (25–56 years) drove at night starting either at 6 p.m. or 10 p.m. after a usual working day. No driving simulator sickness was reported. These subjects are labeled S1 to S25 in Chapters 7 and 8 and did not participate in other experiments of this work¹.

Material and experiment procedure

An S-Class Mercedes-Benz cabin and a highly monotonous, low traffic highway driving condition during the night with two lanes were selected for this experiment. In addition to the EOG and ECG data, KSS and the acceptance of warning were also collected every 15 min prompted by a

¹The subjects of the daytime driving with secondary tasks in Section 4.2.2, who are also labeled S1 to S26, were not the same as those who participated in the driving simulator experiment.

dong tone sound. On the contrary to the real road experiments, subjective self-rating of the drowsiness level was collected verbally. Similar to other experiments, an IR-camera was installed in the car for recording the subject's face during driving. 14 subjects were additionally equipped with a head tracker and a head mounted eye tracker. These subjects also performed a speech test right after alternate KSS events. They were asked to repeat some sentences for about 4 min. These parts of the experiment, which were collected for other purposes, were removed for further analysis, since talking leads to noisy EOG data.

The very first minutes of the experiment were intended for getting familiar with the simulator and accommodation of the eyes. Unlike the daytime experiments, where the subject should have driven the whole route, in this experiment the subjects were asked to drive as long as they could and even give effort to fight against drowsiness, if it was possible. The length of the circular route was 200 km and it was repeated after completing one round of it. On average, each subject drove 335 km with the maximum speed of 130 km/h. Two construction sites were also included through the 200 km route at 62nd and 88th km to make the driving scenario more realistic. In addition, some takeover maneuvers were also designed. The subjects were allowed to activate the adaptive cruise control. They were asked not to talk to the investigators in the control room who were responsible for observing the subjects during the experiment and for documenting the experiment. In general, subjects aborted the experiment due to sever drowsiness, either by themselves or suggested by the investigators. We emphasize that the subjects were not necessarily drowsy during the entire drive in this experiment.

Table 4.2 summarizes all conducted experiments.

Table 4.2.: Summary of experiments studied in this work

	daytime experiment			nighttime experiment	
	pilot study	without secondary tasks	with secondary tasks	without secondary tasks	
real or simulated driving	real	real	real	real	simulated
number of subjects	8	18	26	10	25
driven distance	–	260 km	260 km	244 km	355 km
driving duration [hh:mm]	–	2:30	2:30	2:10	2:40
starting time	–	9 a.m. or 1 p.m.	9 a.m.	10 p.m.	6 p.m. or 10 p.m.
EOG data collection	yes	yes	yes	yes	yes
KSS data collection	no	yes	no	yes	yes
accompanied by an investigator?	yes	no	yes	yes	no

5. Eye movement event detection methods

This chapter discusses eye movement detection methods. First, it is necessary to clarify, why the precise detection of eye movements provides the foundation for successful driver drowsiness detection. If eye movement events are not detected properly, it is obvious that the features being extracted from them lack useful information for further analysis such as the classification. Improper event detection is the one with a high rate of missing events or false detections. Under these circumstances, the relationship between features and driver drowsiness cannot be determined correctly. As a result, detection of eye movements should be done with care and high precision, since it directly affects the results of next analysis steps.

It should also be mentioned that correct detection of an eye movement event, e.g. a blink, also contains the detection of its start and end points precisely. Therefore, if a blink is only partly detected, it cannot be counted as an acceptable detected event.

In this chapter, first, a well-known method of blink detection based on median filtering is explained. After discussing the shortcomings of this method, our developed detection algorithms based on derivative signal and continuous wavelet transform will be explained as alternatives to the median filter-based method. Moreover, it will be shown how our new approaches complete each other in terms of detection of different eye moments. The proposed derivative-based detection algorithm is responsible for the detection of rapid eye movements, while the suggested wavelet transform-based algorithm detects slow eye movements. At the end of this chapter, all studied event detection methods are discussed and compared with each other.

Parts of the material in this chapter are drawn from Ebrahim et al. (2013a).

5.1. Eye movement detection using the median filter-based method

Some examples of blinks during the awake and drowsy phases were shown in Figure 3.4. During the awake phase, in which a person does not suffer from sleep deprivation, blinks often follow similar characteristics, i.e. their amplitude and duration do not change remarkably. Some examples of such blinks are shown in the top plot of Figure 5.1. This implies that blink detection can easily be performed by applying some fixed criteria, e.g. by comparing the $V(n)$ signal with a fixed threshold. However, as mentioned before, drift, which is not related to any type of eye movements, is inevitable in the EOG signal (see top plot of Figure 5.1). Therefore, a fixed threshold does not lead to the correct blink detection.

A conventional method for eliminating the drift in the EOG signal and improving the blink detection is to apply a median filter to $V(n)$ and then subtract the result from the original $V(n)$ signal. This method has been applied in several studies for blink detection, such as in Hargutt (2003), Martínez et al. (2008), Krupiński and Mazurek (2010) and Huang et al. (2012). As a result, we have

$$\hat{V}(n) = V(n) - V_{\text{med}}(n), \quad (5.1)$$

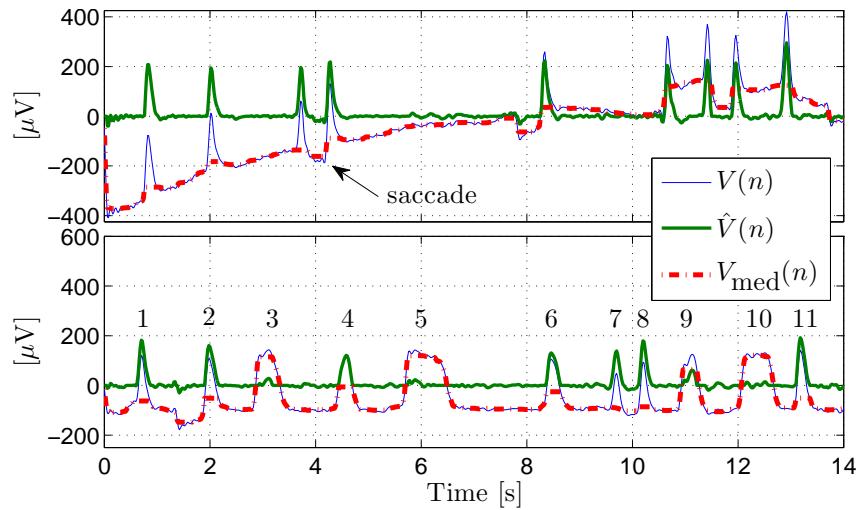


Figure 5.1.: Drift removal by applying a median filter to $V(n)$ to improve blink detection. Top: awake phase, bottom: drowsy phase

where $V_{\text{med}}(n)$ refers to the median filter processed $V(n)$ with empirically chosen window size of $w_{\text{med}} = 2(\frac{f}{4}) + 1$ samples ($f = 50$ Hz) in this study. At sample n , $V_{\text{med}}(n)$ represents the median of $V(n)$ calculated within the interval $[n - w_{\text{med}} + 1, n]$. The top plot of Figure 5.1 shows the result of applying this method. It is clear that due to the eliminated drift in $\hat{V}(n)$, the blinks can now be easily detected by applying an amplitude threshold like $th_{\text{amp}} = 100 \mu\text{V}$.

The bottom plot of Figure 5.1 shows $V(n)$ signal of the same subject during the drowsy phase. Subject's level of drowsiness was assessed based on the video analysis and collected KSS values. As shown in this plot and Figure 3.4, blinks in the drowsy phase differ widely in amplitude and duration from those of the awake phase. The same median filter has been applied to process $V(n)$ of the drowsy phase as well. In this case, not all blinks were detected correctly in $\hat{V}(n)$ signal. The half of the fourth blink, as an example, has been removed by applying the median filter as magnified in Figure 5.2. In addition to the amplitude, the duration of this blink is different in $V(n)$ and $\hat{V}(n)$.

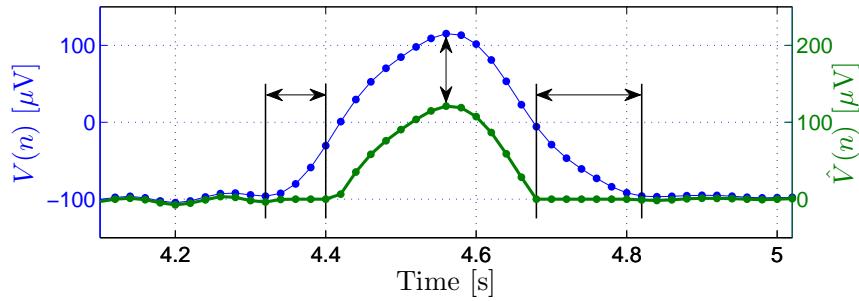


Figure 5.2.: Information loss of slow blinks by median filter method

Moreover, the blinks with longer durations (3rd, 5th, 9th and 10th events in Figure 5.1) almost disappeared after median filtering. Setting a small value for th_{amp} also does not help, as saccades or noise might be incorrectly detected as blinks. The problem is that the efficiency of the median filter-based method is highly dependent on the chosen w_{med} . The more w_{med} of the median filter matches the blink duration, the less blink information is lost in $\hat{V}(n)$. Since the blink duration not only varies inter-individually, but also for an individual according to the level of drowsiness, applying a fixed window size median filter would not lead to successful blink detection for the drowsy phases. In addition, another deficiency of this method is that saccade detection becomes

impossible, as all saccades shown in Figure 5.1 have been filtered out as drift. Therefore, a median filter is not suitable for our application, because, here, all fast eye movements (blinks and saccades) are of interest and a median filter removes both slowly varying drift as well as some blinks and saccades.

5.2. Eye movement detection using the derivative-based method

In the previous section, we saw that the median filter-based method is a powerful blink detection method for detecting sharp and short blinks of the awake phase. However, during the drowsy phase, most of the events are missed or detected with low precision. To overcome the mentioned problem, this section describes a method which benefits from the derivative of the EOG signal for blink detection. Some previous studies also used derivative signal for detecting blinks such as Jammes et al. (2008), Hu and Zheng (2009) and Wei and Lu (2012). However, their proposed algorithms had some weaknesses. Jammes et al. (2008) mentioned that their detection algorithm is unable to distinguish between longer eye closures and vertical saccades occurred during looking at the dashboard, because they look very similar to each other. Distinguishing between these eye movements has been mentioned as an important issue for driver drowsiness detection in other studies as well (Svensson, 2004). In addition, Jammes et al. (2008) showed that their algorithm did not detect slow blinks. Wei and Lu (2012) also applied the blink detection method suggested by Jammes et al. (2008) and used frequency-based methods (Fourier and wavelet transform) additionally to extract the number of slow eye movements in the horizontal EOG signal.

A clear disadvantage of the mentioned studies is that they only concentrated on the detection of blinks in $V(n)$. However, as mentioned in Sections 3.1 and 3.3, observing the scene ahead by saccadic eye movements is essential for the safe driving. Unfortunately, distinguishing between saccades and blinks (either fast blinks or micro sleep events) in $V(n)$ of EOG signals have not been addressed in previous derivative-based algorithms. Here, we introduce a novel derivative-based approach which takes saccade detection into consideration as well, such that blinks and saccades are detected simultaneously as fast eye movements and are distinguished from each other. This property of our algorithm tackles the existing problem in the conventional derivative-based blink detection algorithms.

The derivative of the $V(n)$ signal, $V'(n)$, calculated by the Savitzky-Golay filter (Savitzky and Golay, 1964) with the polynomial order 1 and the frame size 7, is shown in Figure 5.3. In the Savitzky-Golay filter approach, a polynomial of degree 1 is fitted to 7 samples of $V(n)$ successively in the least-squares sense. $V'(n)$ at the midpoint of the 7 samples is obtained by performing the differentiation on the fitted polynomial rather than on the original $V(n)$. The mentioned parameters of the filter are only used for the detection of blink events¹.

In the following, our blink detection algorithm based on $V'(n)$ is explained.

1. Detecting potential blinks

According to Figure 5.3, potential blink events can be detected by setting an amplitude threshold th_{vel} regarding the blink velocity $V'(n)$ to consider all of its peaks, namely

$$|V'(n)| > th_{vel}. \quad (5.2)$$

Afterwards, around all accepted peaks, three successive sign changes are searched for. These points define *start*, *middle* and *end* points of a blink as *a*, *b* and *c*, respectively as

¹For the sake of accuracy, another parameter set is used for the extraction of blink features.

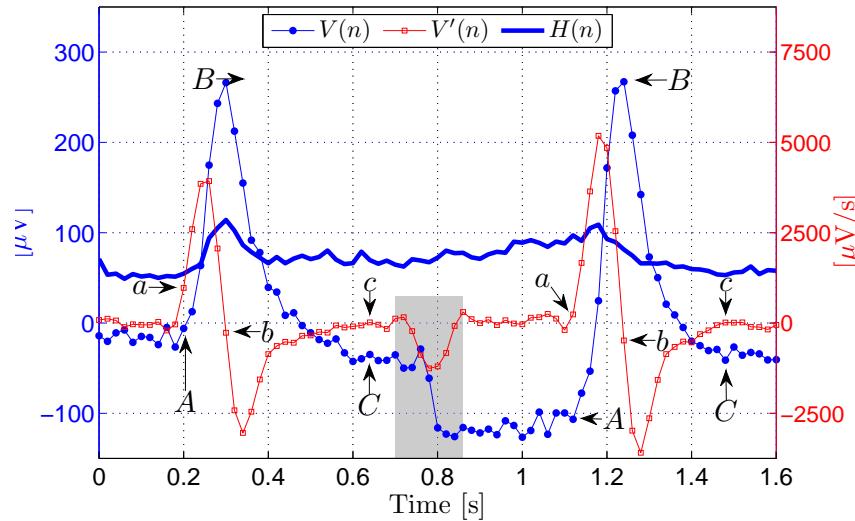


Figure 5.3.: $V(n)$ and its derivative $V'(n)$ representing eye blinks during the awake phase

shown in Figure 5.3. We have considered the first point after a sign change. Negative to positive transitions of $V'(n)$ are defined by a and c , while positive to negative transitions by b . These three respective points of $V(n)$ are also marked as A , B and C in Figure 5.3. a to b (A to B) and b to c (B to C) transitions describe closing and opening of eyes during a blink event. At the end of this step, all potential blinks are detected.

2. Calculating the blink amplitude of potential blinks

After detecting all potential blinks, the blink amplitude is extracted. This comprises both closing and opening amplitudes, namely $B - A$ and $B - C$. The difference between these amplitudes is negligible for normal blinks. Usually, the amplitudes measured at the beginning (A) and at the end (C) of a blink remain unchanged, as for the first blink in Figure 5.3. For saccadic blinks, however, this difference is non-zero and is equivalent to the amplitude of the saccade time-locked to the blink. Therefore, in order not to consider the amplitude of the saccade in the blink amplitude, the blink amplitude for the i -th blink is defined as

$$amp_i = \min(B_i - A_i, B_i - C_i) . \quad (5.3)$$

Wei and Lu (2012), however, used $\frac{(B_i - A_i) + (B_i - C_i)}{2}$ as the amplitude of the blink. Their definition ignores the difference between the amplitude of saccadic and non-saccadic blinks.

3. Categorizing potential blinks with respect to their amplitude

Now, the question is whether all detected patterns are true eye blinks. In order to assess this, the histogram of the amplitude in (5.3) was analyzed as shown in Figure 5.4 for 11 subjects. These subjects participated in the driving simulator experiment (see Section 4.3). The histograms are normalized with respect to the maximum number of occurrences for each subject separately. According to Figure 5.4, for almost all subjects, two clusters of amplitude are distinguishable. The question is what these clusters refer to. The cluster with the smaller amplitude describes the vertical saccades and microsleep events. This means that although the focus of detection was on blinks, other eye movements have been detected as well. The reason is that the chosen th_{vel} was small enough to take other fast eye movements besides blinks into consideration. Figure 5.5 and the highlighted area in Figure 5.3 show saccades detected by the explained detection algorithm.

As explained in Section 3.3, saccades and blinks with long eye closures have similar shape

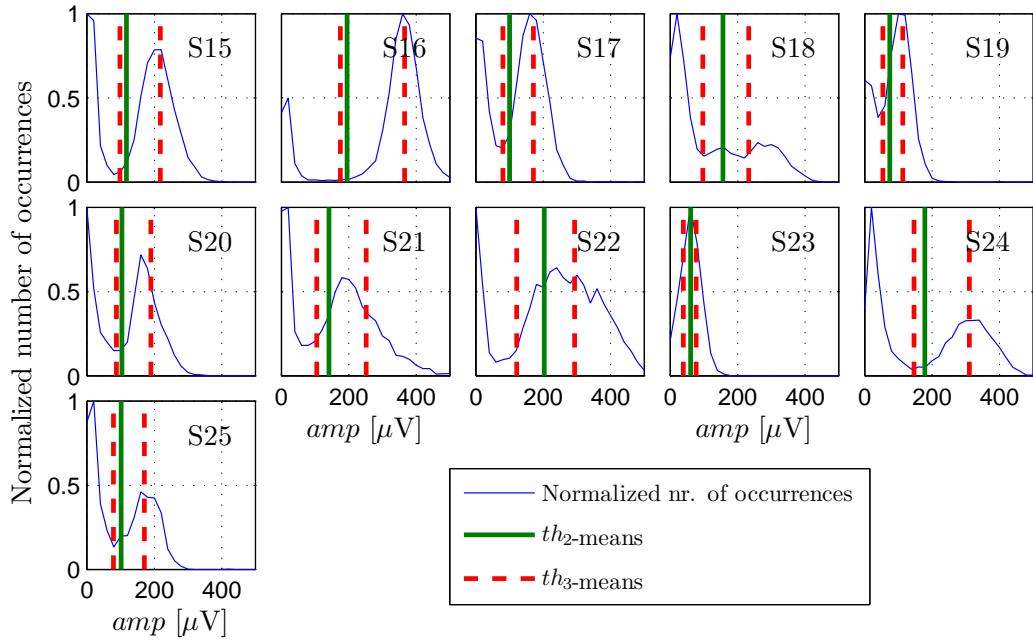


Figure 5.4.: Normalized histogram of all detected potential blinks and their clustering thresholds by the k -means clustering method for 11 subjects

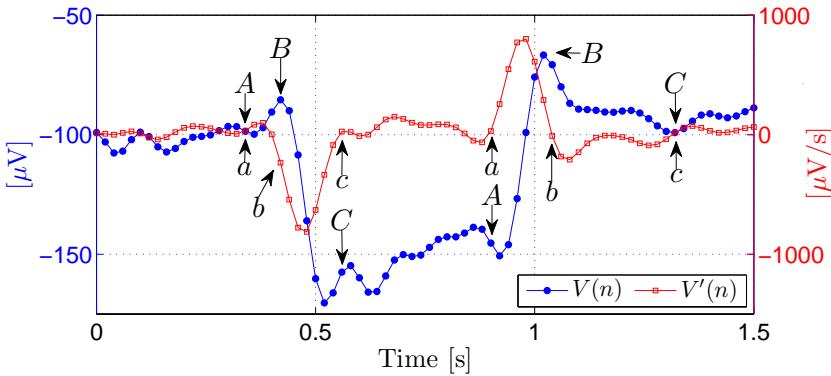


Figure 5.5.: Simultaneous detection of saccades by eye blink detection algorithm

and form (see Figures 3.4(b) and 3.5). Therefore, analogous to saccades, opening and closing stages of such blinks are detected by the algorithm and considered in the group with smaller amplitudes in Figure 5.4. After identifying different clusters in Figure 5.4 (blinks versus saccades/microsleeps), a clustering method such as the k -means clustering (see Appendix C) is required to find the exact border between them. At first sight, applying a 2-class clustering seems to be sufficient. However, in addition to saccades and microsleep events, the data includes blinks from both awake and drowsy phases. In fact, three clusters are available: 1) saccades and microsleeps (similar to Figure 3.4(b)), 2) blinks during the drowsy phase or with longer eye closure and smaller amplitude due to drowsiness (similar to Figure 3.4(c)) and 3) blinks during the awake phase or with short eye closure (similar to Figure 3.4(a)). Therefore, applying a 3-class clustering algorithm is recommended. The thresholds of both 3-means, th_3 -means, and 2-means, th_2 -means, are shown in dashed and solid lines, respectively, in Figure 5.4. th_3 -means on the left side refers to the threshold between saccades/microsleeps (cluster 1) and blinks of the drowsy phase (cluster 2). The threshold on the right side discriminates the second and the third clusters. Obviously, by

applying the 2-class clustering, many decreased amplitude blinks due to drowsiness will be incorrectly clustered as saccades or vice versa as for subjects S18 and S22. Finally, all events fulfilling both $B - A > th_{3\text{-means},\text{left}}$ and $B - C > th_{3\text{-means},\text{left}}$ are accepted as blinks. It is noticeable that the thresholds differ from person to person which implies that no fixed threshold should be applied for distinguishing between the clusters.

4. Distinguishing between vertical saccades and blinks with longer eye closure

The goal of this step is to distinguish between saccades and other eye movements which are all clustered in a common group in the previous step. There, only the minimum amplitude of $B - A$ and $B - C$ was of interest, while now the actual amplitude of these events are studied. The amplitude of the i -th eye movement of this group, $amp_{em,i}$, is defined as

$$amp_{em,i} = |C_i - A_i|. \quad (5.4)$$

In fact, the relative variation of the amplitude is considered, overcoming the overshoots in saccadic eye movements (see Figures 3.5(c) and 5.3). Similar to the previous step, the histograms of the amplitudes calculated by (5.4) are analyzed as shown in Figure 5.6 for subjects S15, S19 and S22.

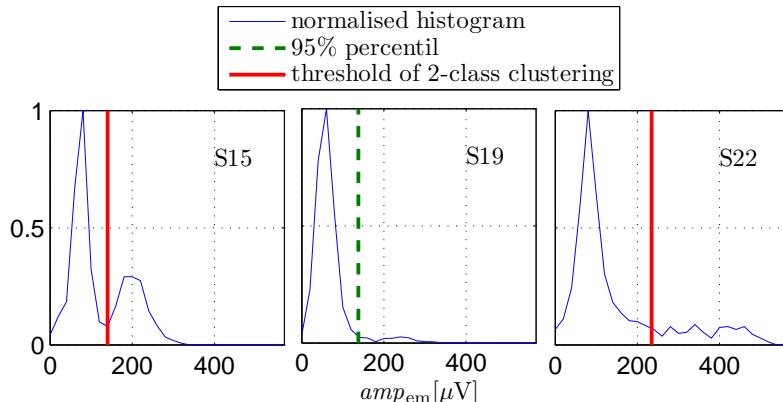


Figure 5.6.: Normalized histogram of all detected potential saccades and blinks with long eye closure. Their clustering thresholds are also shown.

For subject S15, two classes are again distinguishable. The group with smaller amplitudes refers to vertical saccades, while the other group describes blinks with long eye closure. Due to smaller vertical than horizontal space of human eyes, the amplitude of vertical saccades are limited. As a result, vertical saccades with large amplitudes comparable to the microsleep events shown in Figure 3.4(b) do not occur during common driving tasks. On the other hand, it is clear that if microsleep events similar to the saccadic pattern shown in Figure 3.4(b) do not occur often, the clusters will not be as distinguishable as for subject S15. This is the case for subject S22, while for subject S19, the number of such events was so small that only one cluster can be distinguished. Therefore, for histograms similar to subjects S15 and S22, the k -means ($k = 2$) algorithm is applied and for histograms without distinct clusters, the 95th percentile is used as the border between vertical saccades and microsleep events. The thresholds are shown in Figure 5.6.

5. Plausibility check of detected fast eye movements

Since EOG signals are very sensitive to any muscle artifact around the electrodes, it might be possible that some artifacts are confused with eye movements. A possible method for overcoming this problem is to ascertain which eye movements are related to each other and

to exclude unrelated ones. This is logical during driving, as it is assumed that the driver looks straight ahead most of the time. Therefore, a *main looking direction* can be defined. As a result, for all saccades representing looking away from the main looking direction, another saccade in the opposite direction should be present as shown in Figure 3.5. All detected saccades, which do not fulfill the mentioned criterion, are discarded. For saccades occurring as saccadic blinks, a threshold, th_s , is additionally required to avoid confusing them with non-saccadic (normal) blinks. In fact, the value of $|A - C|$ in Figure 5.3 is crucial. Based on a similar argument, microsleep events can be checked as well, because eye closures should be followed by an eye opening during driving. It is obviously supposed that none of the subjects falls thoroughly asleep until end of the experiment. Finally, all detected eye movements, which are not assigned to adjacent movements, are considered as false detections and are removed from the detected events' data set. In order to find th_s , the histogram of the $|A - C|$ amplitude of the blinks occurring both before and after detected saccades are analyzed. The reason is that if a saccade is adjacent to a saccadic blink, four different combinations of them may occur:

- up-going saccade, down-going saccadic blink
- down-going saccade, up-going saccadic blink
- down-going saccadic blink, up-going saccade
- up-going saccadic blink, down-going saccade.

These pairs are shown in the first four plots from the left of Figure 5.7. The th_s is found by applying the k -means algorithm ($k = 2$) to the histogram of $|A - C|$ amplitude of these blinks to distinguish between two clusters: saccadic versus non-saccadic blinks. After finding the threshold, all blinks fulfilling $|A - C| > th_s$ are candidates to be considered as a pair together with a detected saccade. Moreover, it is also possible that two saccadic blinks are joined together as shown in the first two plots from the right of Figure 5.7. Finally, only the saccades of pairs similar to those shown in Figure 5.7 are considered as correct eye movement detection.

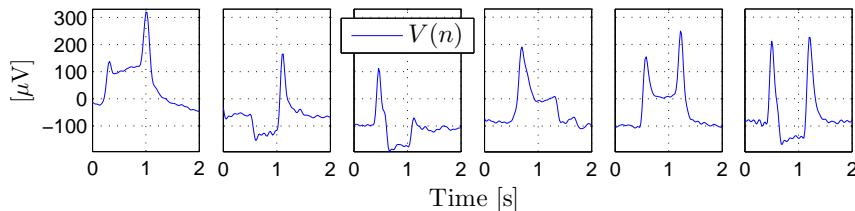


Figure 5.7.: Possible combinations of two vertical saccades in $V(n)$

6. Horizontal saccade detection

Similar to the vertical saccades, horizontal saccades are detected by comparing $|H'(n)|$ with a threshold as explained in Step 1. As blinks are not available in $H(n)$, detected patterns are either saccades or artifacts. Therefore, only the saccades, which pass the mentioned plausibility check, are considered as horizontal saccades in the end.

Figure 5.8 summarizes the described algorithm for blink detection.

A clear drawback of the suggested algorithm is that slow blinks might not be detected correctly based on this method. If both closing and opening phases of the blinks are very slow, then the derivative-based method definitely misses the detection of both of them. If th_{vel} in (5.2) is selected to be very low for detecting these slower blinks, noise or a lot of non-relevant eye movements will be detected wrongly as well and removing them would be very cumbersome.

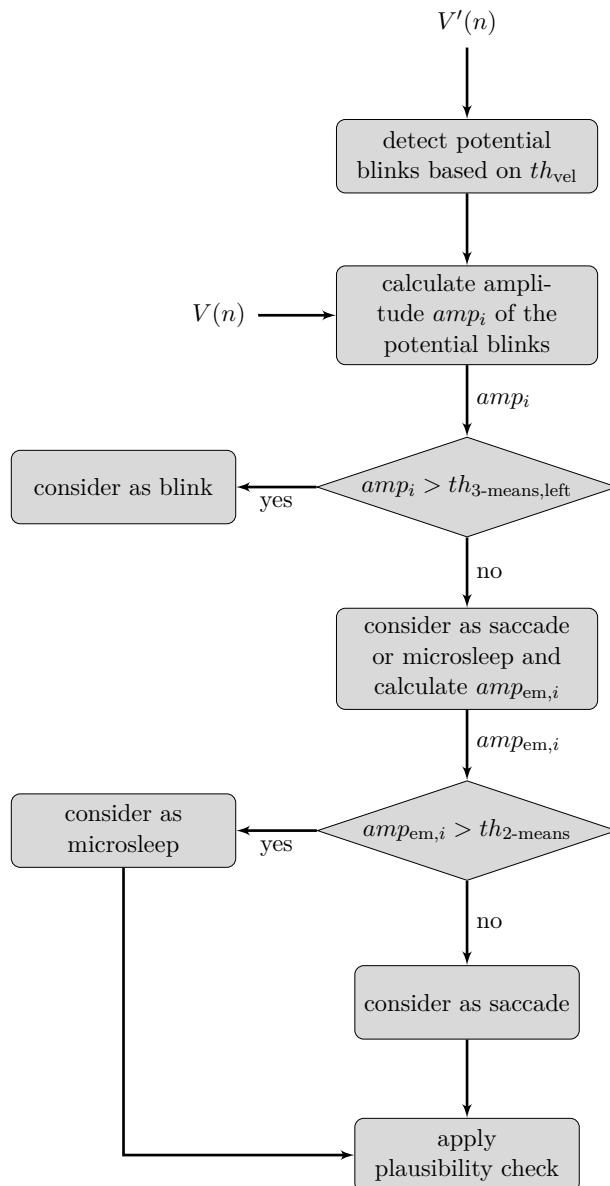


Figure 5.8.: Flow chart of the derivate-based method for blink detection

For blinks, which are fast in one phase and slow in the other phase, either the opening or the closing phase is detectable based on our proposed algorithm, not both of them. We call such events *incomplete events*. The problem is that according to step 5 of the explained detection algorithm, incomplete events do not pass the plausibility check and are consequently discarded. We observed that it is the closing phase which gets slower due to drowsiness in comparison to the opening phase. A solution to detect incomplete events as well is to *adapt* the threshold applied to $V'(n)$ in (5.2) for event detection. This approach was also applied in some cases in this study. Therefore, as soon as an incomplete event is detected, the threshold in (5.2) is reduced in order to find the missing pair of it, namely either opening or closing phase of it. However, for slow velocity blinks in both closing and opening phases another approach is needed which is explained in the next section.

5.3. Eye movement detection using the wavelet transform-based method

This section studies another method of eye movement detection which is based on the *wavelet transform* (WT). The wavelet transform has been applied for detection of fast blinks and saccades in brain computer interaction and activity recognition applications (Reddy et al., 2010; Bulling et al., 2011; Barea et al., 2012). Obviously, relevant eye movements to such applications are fully controlled and are different from spontaneous eye movements relevant to driver drowsiness. Barea et al. (2012) only targeted the detection of saccadic eye movements without discussing blink detection. In the studies by Reddy et al. (2010) and Bulling et al. (2011), only the detection of fast blinks was addressed. Therefore, their proposed algorithms are inapplicable to the detection of slow blinks due to drowsiness. Analysis of slower movements has been investigated in Magosso et al. (2006) and Wei and Lu (2012). The former concentrated on very slow eye movements which occur while the eyes are completely closed for a long time. Such eye movements are interesting for sleep disorder-related research studies and are out of the scope of driver drowsiness detection. The method suggested by Wei and Lu (2012), however, was used for extracting a feature based on slow horizontal eye movements. It did not target the detection of slow blinks in $V(n)$.

Unfortunately, the introduced algorithms based on wavelet transform in previous studies are not applicable to driver drowsiness detection, since they have only aimed to detect sharp blinks similar to those occurring during the awake phase. In this section, however, after highlighting the advantage of the wavelet transform over the Fourier transform, we will propose two new algorithms. The first algorithm targets the detection of both slow and fast blinks by applying *continuous wavelet transform* in Section 5.3.2. On the one hand, this algorithm can be considered as a supplementary to the previous algorithm explained in Section 5.2 for additionally covering the detection of slow blinks. On the other hand, it can also be applied independently to detect all fast and slow blinks and saccades. Therefore, events are never incomplete, if this method is applied. In fact, at the end of this section, all of the relevant eye movements for driver drowsiness detection are investigated. The second algorithm is a preprocessing step which benefits from the properties of the *discrete wavelet transform*. First, it will be shown in Section 5.3.3 how this transform can be used to remove noise and drift in the EOG signals. Then, the second new algorithm *adaptively* removes noise in the collected data in order to avoid information loss. We applied both noise and drift removal approaches based on wavelet transform to all EOG signals in this work before applying eye movement detection algorithms. This helped to improve the performance of the event detection.

The background theories in this section are taken from Burrus et al. (1998), Niemann (2003), Keller (2004), Mallat (2009), Poulikas (2009) and Soman et al. (2010).

5.3.1. Discrete Fourier transform

Fast and slow eye movements can be characterized based on their frequency components. Therefore, frequency analysis of the EOG signal is another approach for detecting different types of eye movements.

A discrete time signal $x(n)$ is analyzed in the frequency domain by applying the *discrete Fourier*

transform (DFT) $\mathcal{F}\{\cdot\}$ to it, defined as

$$\mathcal{F}\{x(n)\} \equiv \mathcal{X}(\Omega) = \sum_{n=0}^{N-1} x(n) e^{-i\Omega n}, \quad (5.5)$$

where $\Omega = \frac{2\pi k}{N}$ with $k = 0, 1, 2, \dots, N - 1$ and number of time samples N .

Since the DFT does not provide time localization information, the *short time Fourier transform* (STFT) was introduced as the solution to this problem. Unlike the DFT, which considers the whole signal from the first to the last sample at once, in STFT, the signal is first multiplied by a window function, which is non-zero for a short time, and then the DFT is calculated. Therefore, the resulting DFT only represents the frequency components of the windowed part of the original signal and consequently provides the time localization information. As an example, a cosine signal $x(n)$ with the frequency of 8 Hz (low frequency component) and one discontinuity from $n = 1.5$ s to 1.502 s (high frequency component) is analyzed as shown in Figure 5.9 (top left plot). The sampling frequency is 1000 Hz. The STFT was applied by choosing the Hamming window as the window function by considering different window lengths L_{win} , namely $L_{\text{win}} = 0.02$ s, 0.2 s and 2 s. It is clear that depending on the L_{win} value, STFT provides different frequency component information. The high frequency component of $x(n)$ is apparent with the best time localization for shorter L_{win} , e.g. $L_{\text{win}} = 0.02$ s (top right plot). In this case, the 8 Hz frequency of the cosine as the low frequency component expands up to 100 Hz. By choosing a longer window, e.g. $L_{\text{win}} = 2$ s, however, the frequency resolution for cosine wave improves at a cost of less time localization for the discontinuity (bottom right plot). For $L_{\text{win}} = 2$ s, the high frequency component is completely lost in the spectrogram.

All in all, although the STFT is a good solution for adding time localization information to the DFT, its efficiency in terms of eye movement detection depends to a large extent on the chosen window length. This fact is even highlighted, if non-periodic and stochastic signals like EOG signals are being analyzed.

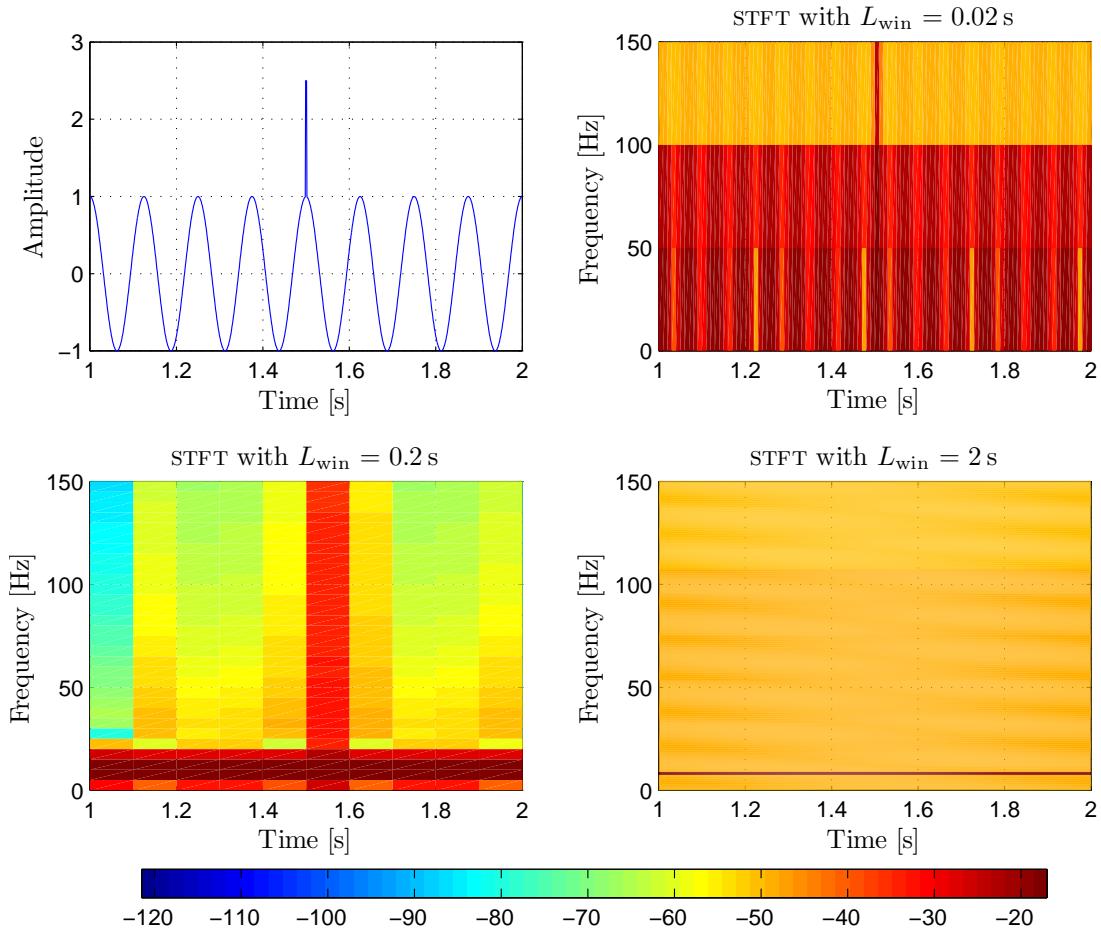
5.3.2. Continuous wavelet transform

Another possibility for providing time localization information in the frequency domain is the wavelet transform. Similar to DFT, which is based on a trigonometric function, this transform also uses special functions which are called *wavelets*. According to Young (1993), a small wave $\psi(t)$, which is oscillatory (zero average) and decays quickly, with the following characteristic

$$c_\psi = \int_{-\infty}^{+\infty} \frac{|\Psi(\Omega)|^2}{|\Omega|} d\Omega < \infty \quad (5.6)$$

is called a *wavelet*, where $\Psi(\Omega)$ is the Fourier transform of $\psi(t)$. (5.6) is an important condition for the existence of the inverse wavelet transform (Niemann, 2003). $\psi(t)$ is also referred to the *mother wavelet*. Additionally, each mother wavelet has λ *vanishing moments* which fulfill the following condition

$$\int_{-\infty}^{+\infty} t^k \psi(t) dt = 0, \quad (5.7)$$

Figure 5.9.: The impact of window length L_{win} on the efficiency of the STFT

where $0 \leq k \leq \lambda - 1$ and λ is a positive integer. Figure 5.10 shows examples of *Haar*, *Daubechies*, *Coiflets*, *Symlets*, *Mexican hat* and *Morelet* mother wavelets. db2, coif2 and sym2 denote the second vanishing moments of the corresponding mother wavelets. The Haar wavelet has only one vanishing moment at $\lambda = 0$.

The WT can be performed either continuously or discretely. The continuous wavelet transform (CWT) of $x(t)$ is defined as

$$\mathcal{W}\{x(t)\} \equiv \mathcal{X}_\psi(a, b) = \int_{-\infty}^{+\infty} x(t) \frac{1}{\sqrt{a}} \psi^*\left(\frac{t-b}{a}\right) dt = \int_{-\infty}^{+\infty} x(t) \psi_{a,b}^*(t) dt, \quad (5.8)$$

where $\psi_{a,b}(t)$ ¹ is a set of *scaled* ($a \in \mathbb{R} > 0$) and *translated* ($b \in \mathbb{R}$) wavelets originating from the mother wavelet $\psi(t)$. a is referred to as *scale* and the factor $\frac{1}{\sqrt{a}}$ normalizes the energy of the scaled wavelets. Obviously, in CWT, continuous values of a and b can be selected. Thus, by this transform, a time-scale (two dimensional) representation of the original signal $x(t)$ is possible, since the variation of a and b results in the multiplication of all scaled and translated variants of $\psi(t)$ with $x(t)$ (see Figure 5.11). The resulting time-scale plane is called the *scalogram*. In MATLAB, scalograms represent the absolute value of the calculated $\mathcal{X}_\psi(a, b)$ and are scaled between 0 and 240. This representation is used in this work as well.

¹The asterisk * denotes the complex conjugate.

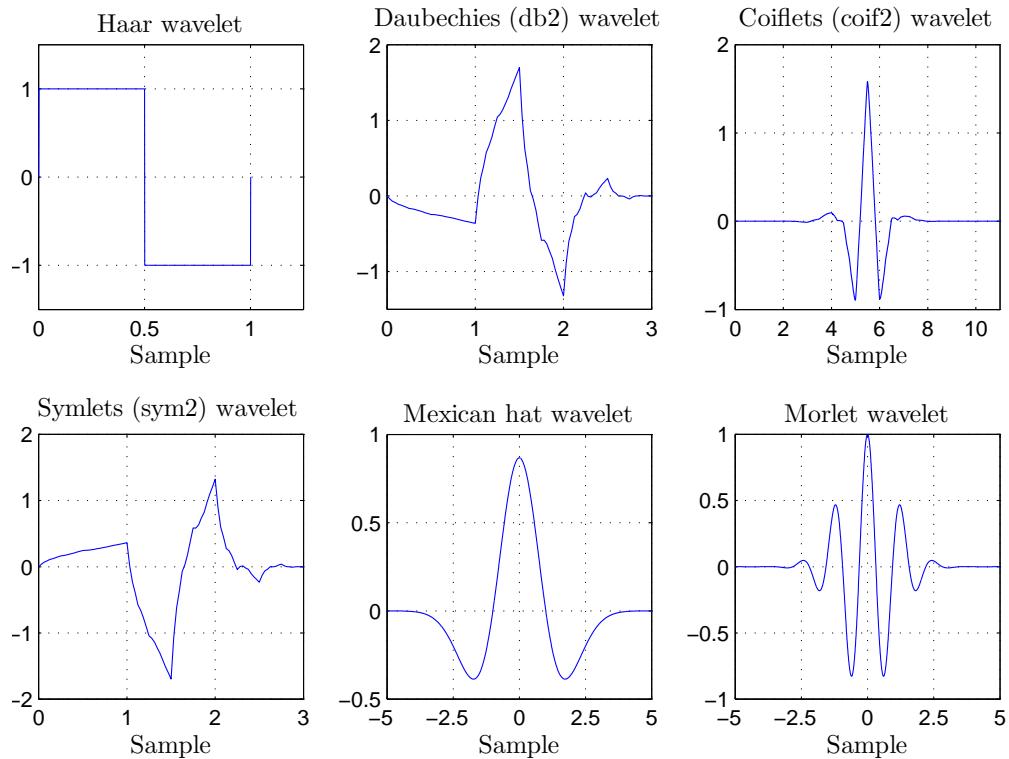
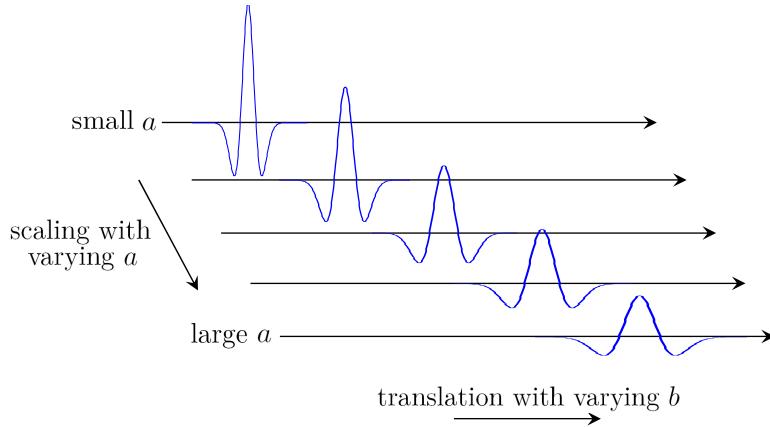


Figure 5.10.: Examples of typical mother wavelets

Figure 5.11.: Scaling and translation of the mother wavelet with varying a and b

A large value of a yields a stretched $\psi(t)$ and underscores slow changes of $x(t)$, namely low frequency components. On the contrary, a small scale a results in a compressed $\psi(t)$ which is proper for highlighting rapid changes of $x(t)$, namely high frequency components. The relationship between scale a and frequency is shown in Figure 5.11.

Soman et al. (2010) relates the WT to the correlation analysis such that large transform values represent a well match between parts of the signal under investigation and wavelets. In fact, CWT measures the similarity between $x(t)$ and the wavelet set. Poulikas (2009) defines the WT as the decomposition of $x(t)$ into sets of basic functions $\psi_{a,b}(t)$.

In order to calculate the CWT numerically, (5.8) should be discretized. Therefore, discrete values of a , b and t are needed. In addition, the integral is replaced by the summation. The upper and lower limits of the integral are also substituted by the upper and lower limits of the domain of

$x(t)$ and selected $\psi(t)$.

The scalogram of Figure 5.12 shows the CWT of the $x(t)$ signal in Figure 5.9 for $1 \leq a \leq 256$ with Haar as the mother wavelet (see Figure 5.10). On the contrary to the STFT, here, both frequency components are visible in the scalogram. Larger scales, e.g. $a = 120$, represent the low frequency component of $x(t)$ (8 Hz), while the high frequency component (the discontinuity) is shown by lower scales. Therefore, WT is a powerful method for analyzing signals such as EOG with different frequency components at different time samples.

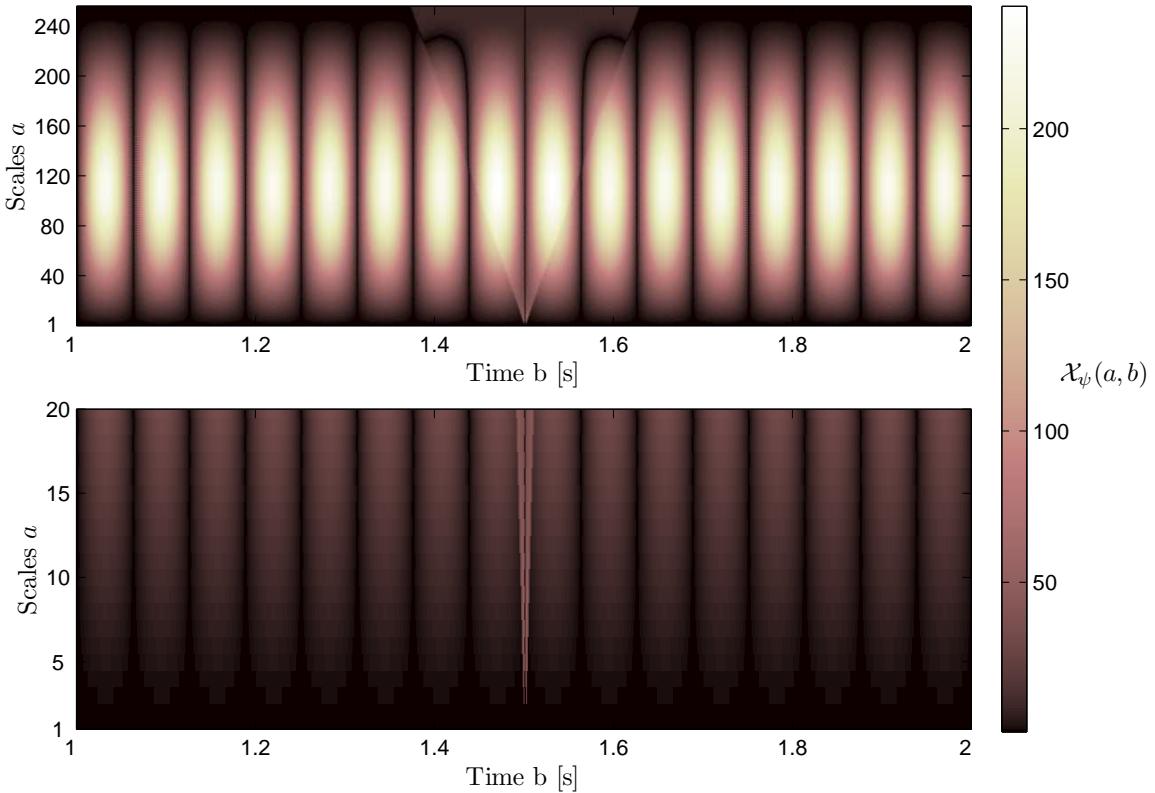


Figure 5.12.: Scalogram of the CWT for $x(t)$ signal shown in Figure 5.9, top plot: $1 \leq a \leq 256$, bottom plot: $1 \leq a \leq 20$

For the EOG signals, the goal is to detect blinks in all phases of the drive and to distinguish between saccades and blinks. To this end, Figure 5.13 shows the scalograms of the CWT with Haar, Morlet, Daubechies and Coiflet mother wavelets for 20 s of the awake and drowsy phases of the drive. The scalogram was calculated based on (5.8). The scale a has been varied from 1 to 50 ($1 \leq a \leq 50$). Comparison of the scalograms considering all mother wavelets indicates that time localization quality of blinks are the same for all of the mother wavelets. Sharp blinks of the awake phase are representative of abrupt changes which are highlighted in low scales. For the drowsy phase, however, the blinks are more visible for larger values of a in comparison to the awake phase due to their lower frequency and slow changes. This corresponds to the smaller values of $\mathcal{X}_\psi(a, b)$ (darker color of the scalogram) at lower scales in Figure 5.13. Overall, the scalograms show that $a > 15$ does not provide accurate time localization information.

Figure 5.14 shows $\mathcal{X}_\psi(a, b)$ of Figure 5.13 at $a = 5, 10$ and 15 . It can be seen that for higher scales, the amplitude of the CWT is higher at a cost of worse time localization for all mother wavelets. In addition, detection of peaks in the CWT by the Haar wavelet seems to be easier than other mother wavelets. This is due to the similarity between pattern of blinks and Haar

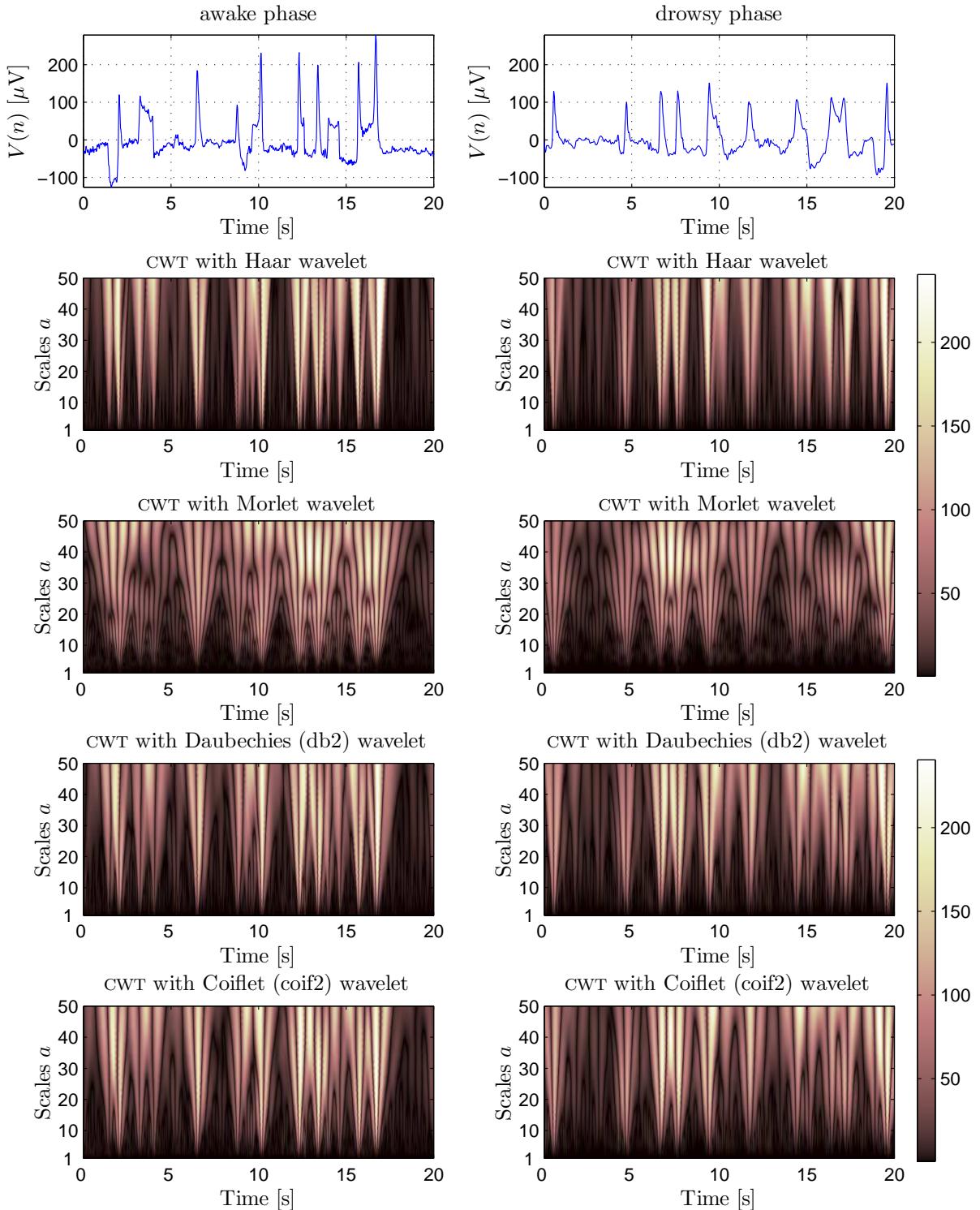


Figure 5.13.: Scalograms of the CWT with different mother wavelets for $V(n)$ signal of the awake phase (left plots) and the drowsy phase (right plots) of the drive

wavelet. As an example, considering $a = 10$ for the Morlet wavelet (third row in Figure 5.14), it is clearly difficult to identify the start and end of an eye blink. For the drowsy phase, as mentioned before, $a = 5$ seems not to be a suitable scale due to the low amplitude of $\mathcal{X}_\psi(a, b)$. Overall, the detection of blinks can be performed by defining a threshold regarding $\mathcal{X}_\psi(a, b)$ with

specific values of a , because $\mathcal{X}_\psi(a, b)$ does not suffer from the drift problem introduced in Section 2.1.2. In fact, similar to STFT, due to varying characteristics of eye movements, for detection of all events more than one scale is needed.

The results with the Haar wavelet are very similar to the negative of the derivative of the EOG signal $V'(n)$ (Mallat, 2009; Barea et al., 2012). Mallat (2009, Chapter 6.1.2) proved mathematically that applying the WT with a mother wavelet with λ vanishing moments can be interpreted as the λ -th order derivative of the original signal. Hence, applying the Haar mother wavelet to a signal in the WT yields the first derivation of that signal at low scales.

Figure 5.15 demonstrates $-V'(n)$ together with $\mathcal{X}_\psi(a, b)$ with the Haar mother wavelet at $a = 5, 10, 15, 30$ and 100 for the same EOG signals shown in Figure 5.13. Except for the fact that at low scales $\mathcal{X}_\psi(a, b)$ is much smoother than $V'(n)$, for the detection of fast blinks both signals provide recognizable peaks at low scales. However, slower blinks, e.g. around $t = 10$ s in the right plots of Figure 5.15 (shown with an arrow), are difficult to detect in $\mathcal{X}_\psi(a, b)$ with $a = 5, 10$ and $V'(n)$, because the amplitudes are not significantly high. Interestingly, the amplitude of $\mathcal{X}_\psi(100, b)$ is very large around $t = 10$ s which makes the wavelet transform at large scales suitable for detection of slow eye movements. This fact is highlighted in Figure 5.16. In the top plot of this figure, the first and the last 20 s represent fast (high frequency) and slow (low frequency) blinks, respectively. The bottom plot shows $\mathcal{X}_\psi(a, b)$ with $a = 10, 30$ and 100 . At $a = 10$, only fast blinks are recognizable in the $\mathcal{X}_\psi(a, b)$ with accurate time localization information. At $a = 30$, some of the slower eye movements are also evident, such as at $t \approx 28$ s. However, in $\mathcal{X}_\psi(100, b)$, all of the slower movements can be detected due to their large amplitudes. In fact, the bottom plot of Figure 5.16 clarifies that larger amplitudes at higher scales represent low frequency components of the EOG signals. Although at $a = 100$ high frequency movements are also recognizable, extracting the exact location of these blinks is not as easy as at lower scales.

Our approach for detection of fast and slow eye movements by CWT

If only detection of fast eye movements based on the CWT signals is of interest, the event detection algorithm explained in Section 5.2 can be applied to e.g. $\mathcal{X}_\psi(10, b)$ due to the similarity between the result of CWT and $V'(n)$ signals. For the detection of both fast and slow blinks, however, the following algorithm has been applied.

Similar to (5.2) in the previous algorithm, $\mathcal{X}_\psi(a, b)$ signal is compared with a threshold to detect relevant peaks. We have empirically selected $a = 10, 30$ and 100 to cover the following. $a = 10$ and 100 are suitable for the detection of fast and slow blinks, respectively. We have used $a = 30$ to improve the time localization of slower blinks detected in $\mathcal{X}_\psi(100, b)$ as explained in the next steps.

Figures 5.15 and 5.16 show that the amplitude of the CWT signals have different ranges depending on the value of a . Hence, different thresholds should be set for event detection at each scale. The asterisks in Figure 5.17 show all detected peaks of $\mathcal{X}_\psi(a, b)$ at each a by applying different thresholds, separately.

According to the bottom plot of Figure 5.16, an event might be detected at several scales in $\mathcal{X}_\psi(a, b)$ signals depending on its velocity. First, we analyze the lowest scale, namely $a = 10$. For each detected peak of $\mathcal{X}_\psi(10, b)$, we consider a $\pm\Delta t$ time offset around its time index $t_{\text{peak},10}$. Empirically, we selected $\Delta t = 0.3$ s. If other peaks at other scales, namely $a = 30$ and 100 , are also detected in the time interval $[t_{\text{peak},10} - \Delta t, t_{\text{peak},10} + \Delta t]$, they will be merged, since they are referring to the same event which is already detected at a lower scale. Otherwise, that peak

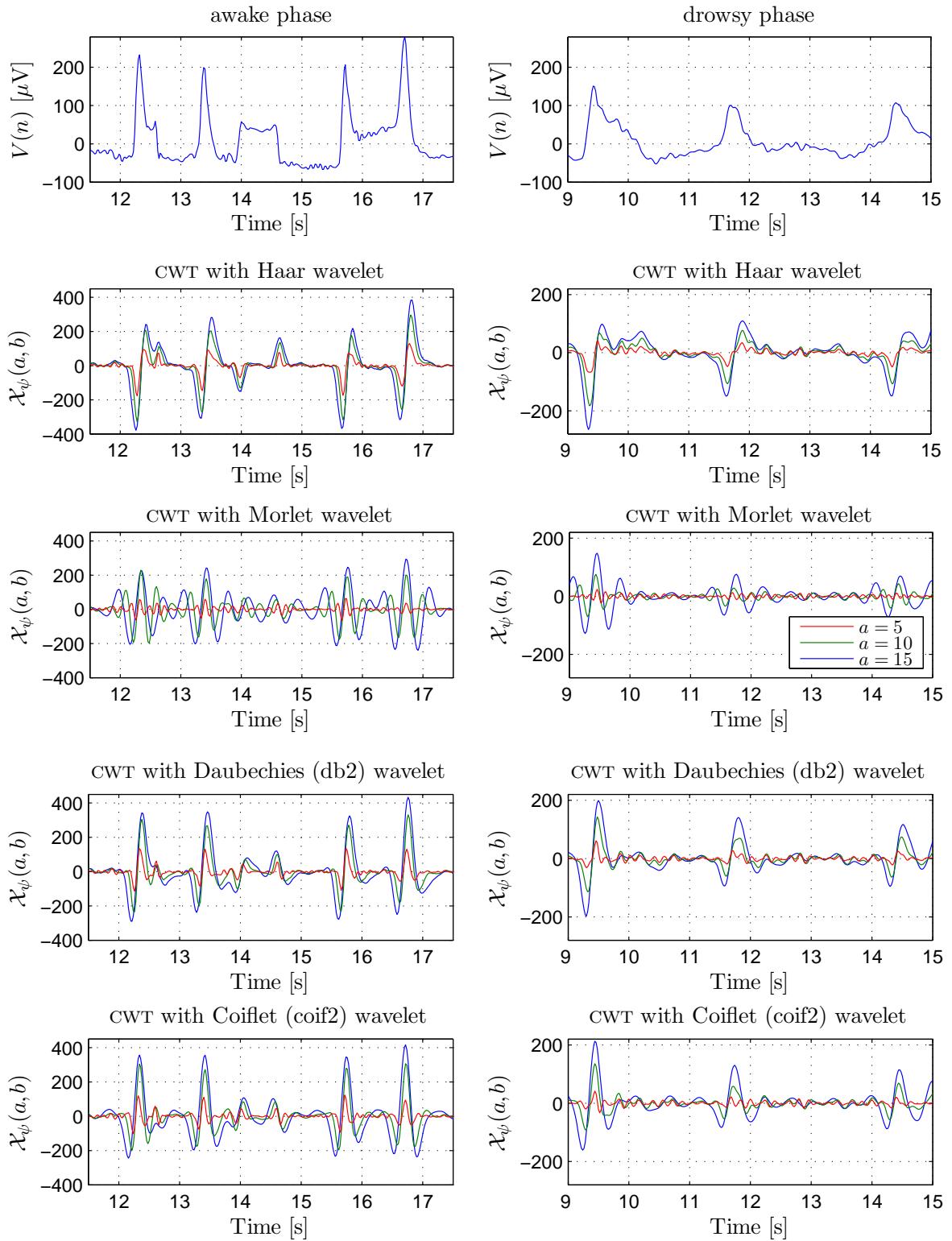


Figure 5.14.: CWT with different mother wavelets for $V(n)$ signal of the awake phase (left plots) and the drowsy phase (right plots) of the drive with $a = 5, 10$ and 15

will be analyzed further. The same procedure is applied to a peak which is detected only at $a = 30$ and 100 . Mathematically, we have

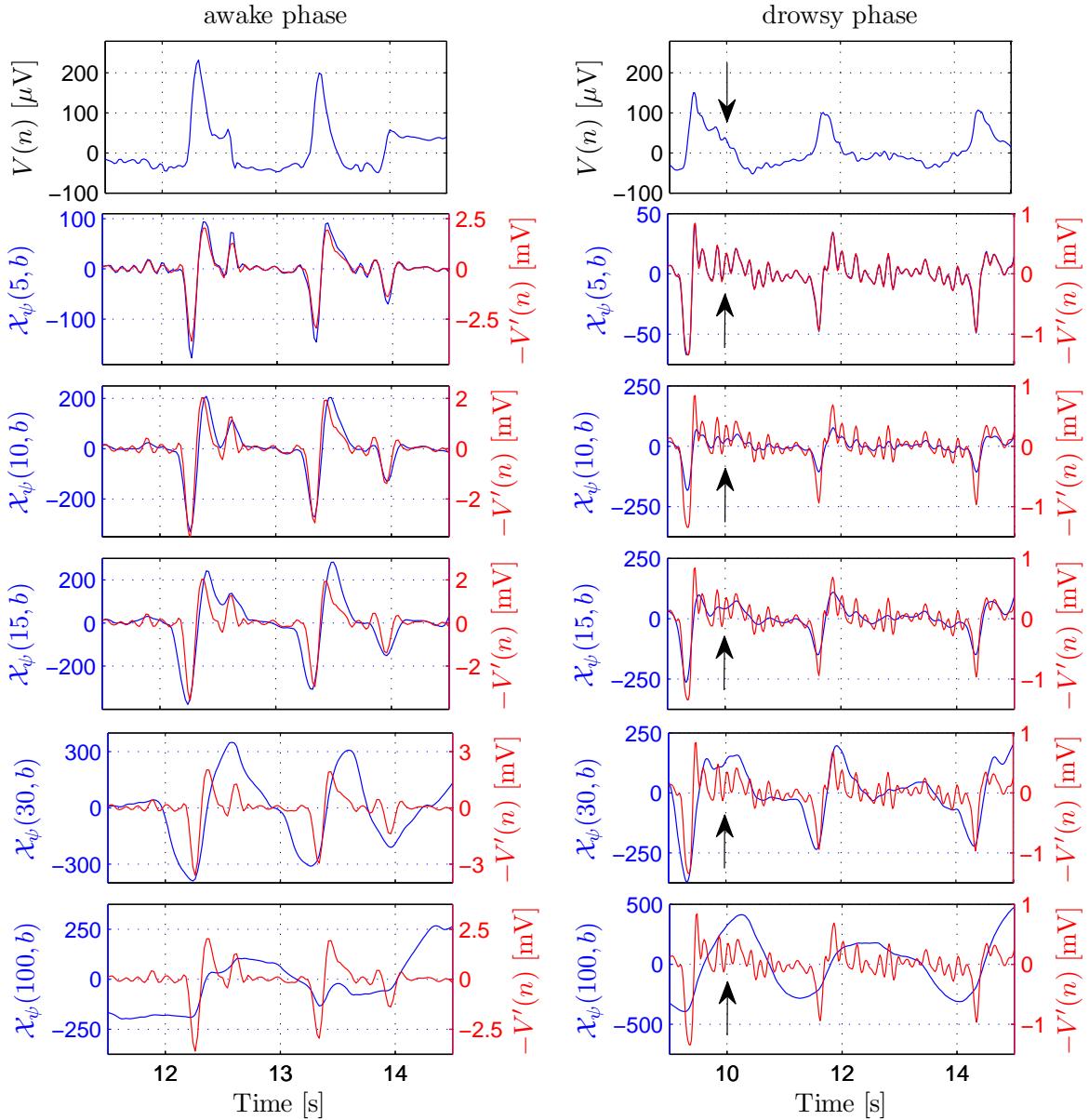


Figure 5.15.: Comparison of $X_\psi(a, b)$ with the Haar wavelet at $a = 5, 10, 15, 30$ and 100 with the negative of the derivative of the EOG signal $-V'(n)$ for the awake and drowsy phases of the drive

- $a = 10$: if $t_{\text{peak},a \neq 10} \in [t_{\text{peak},10} - \Delta t, t_{\text{peak},10} + \Delta t]$, then merge $t_{\text{peak},10}$ and $t_{\text{peak},a \neq 10}$. Otherwise, accept $t_{\text{peak},a \neq 10}$ as a new peak.
- $a = 30$: if $t_{\text{peak},a \neq 30} \in [t_{\text{peak},30} - \Delta t, t_{\text{peak},30} + \Delta t]$, then merge $t_{\text{peak},30}$ and $t_{\text{peak},a \neq 30}$. Otherwise, accept $t_{\text{peak},a \neq 30}$ as a new peak.
- $a = 100$: if $t_{\text{peak},a \neq 100} \in [t_{\text{peak},100} - \Delta t, t_{\text{peak},100} + \Delta t]$, then merge $t_{\text{peak},100}$ and $t_{\text{peak},a \neq 100}$. Otherwise, accept $t_{\text{peak},a \neq 100}$ as a new peak.

It should be mentioned that positive and negative peaks are considered separately. Therefore, only peaks with the same sign are compared with each other. The circles in Figure 5.17 show the accepted peaks at each specific scale a , while their counterparts at other scales were discarded due to merging. For example, the first blink was detected in $X_\psi(10, b)$ by a maximum and

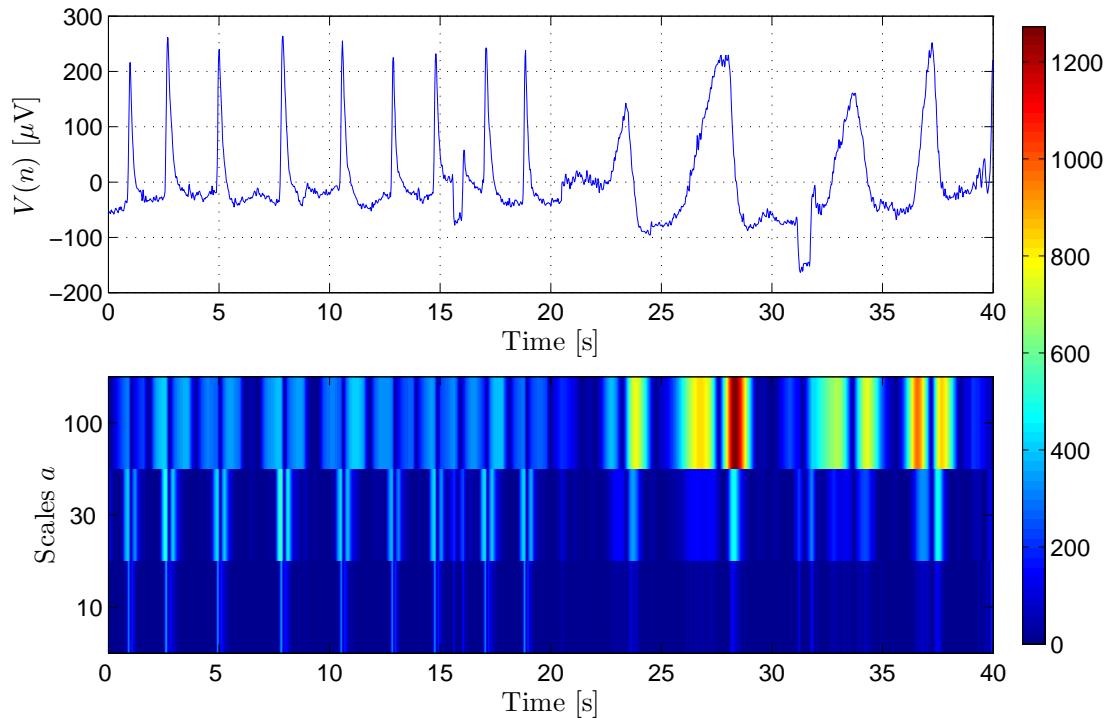


Figure 5.16.: Comparison of CWT at $a = 10, 30$ and 100 for the detection of fast (the first 20 s) and slow (the last 20 s) blinks

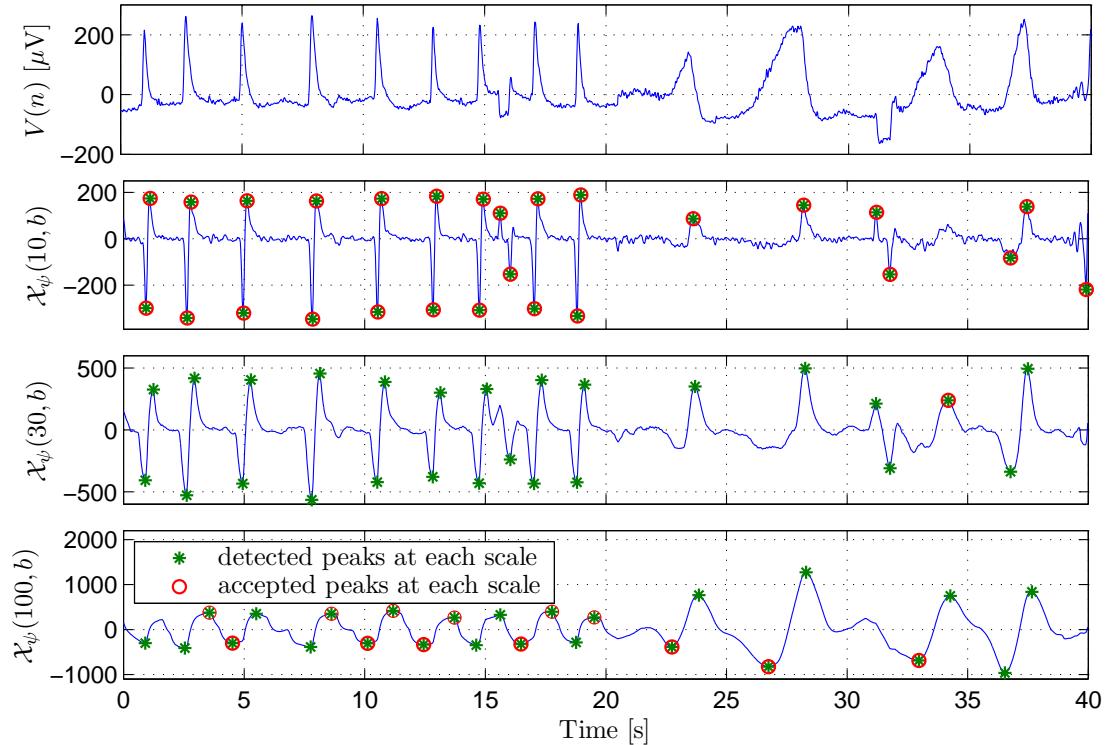


Figure 5.17.: Detected and accepted peaks at different scales of $\chi_\psi(a, b)$ signals

a minimum peak. Thus, the corresponding peaks detected in $\chi_\psi(30, b)$ and $\chi_\psi(100, b)$ were ignored. At $t \approx 27$ s, however, only in $\chi_\psi(100, b)$ a negative peak was detected. This figure also

shows that during the first 20 s, a lot of peaks in $\mathcal{X}_\psi(100, b)$ have fulfilled our condition, although no blink in EOG signal can be related to them. These peaks refer to the time interval between blinks which are all non-relevant. The events related to these peaks can be then discarded by considering some constraints on the amplitude and duration of a detected event.

Figure 5.18 summarizes the proposed blink detection algorithm.

For all accepted peaks at $a = 10$, the start, middle and end points of a blink can be either extracted from $\mathcal{X}_\psi(10, b)$ or from $V'(n)$ based on sign changes as explained before. For all other peaks, we only use $\mathcal{X}_\psi(30, b)$ for improving time localization information. In fact, $\mathcal{X}_\psi(100, b)$ is only used for roughly finding the occurrence of a potential slow blink.

In general, the efficiency of the CWT method in event detection applications highly depends on the selection of a suitable scale value. We saw that depending on the value of a , an event might be missed thoroughly. Kadambe et al. (1999) found the suitable value of a for detection of events of the ECG signals by comparing the number and the location of detected peaks in different scales. If similar number of peaks at similar locations were detected at two successive scales, then increasing the scale values is suggested to be of no benefit. Otherwise, the scale should be increased. This method, however, is not suitable for our application, since different events are not necessarily recognizable at a specific scale. Therefore, similar to our algorithm in Figure 5.18, we suggest the usage of three scales for the full coverage of all fast and slow blinks. The first scale, should be low enough ($5 \leq a_{\text{fast}} \leq 10$) to pronounce fast blinks. On the contrary, the second scale needs to be large to highlight slow movements ($a_{\text{slow}} \approx 10 a_{\text{fast}}$). In order to improve the time localization information of detected blinks by a_{slow} , a middle-valued a ($2 a_{\text{fast}} \leq a_{\text{time-loc}} \leq a_{\text{slow}}/2$) is needed as well.

The next step will be the extraction of start and end of detected events from the corresponding wavelet signal. We showed that Haar wavelet processing of the EOG signal is similar to the high pass filtering and consequently its results are very similar to the time derivative of the EOG signal. Thus, the detection algorithm explained in Section 5.2 can now be applied to the result of the wavelet-transformed EOG signal as well for extracting start and end points of a detected event.

5.3.3. Discrete wavelet transform

Discrete wavelet transform (DWT) can be introduced based on two independent approaches. The first approach defines it as the discretization of a and b in the CWT which is performed based on the *dyadic grid* as follows

$$a = 2^{-j}, \quad b = k \tau_0 2^{-j}, \quad (5.9)$$

where $j, k \in \mathbb{Z}$ and $k \tau_0$ is the dilation step. Consequently, $\psi_{a,b}(t)$ changes to

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^{-j}}} \psi\left(\frac{t - k \tau_0 2^{-j}}{2^{-j}}\right) = 2^{j/2} \psi(2^j t - k \tau_0). \quad (5.10)$$

For simplicity, we consider $\tau_0 = 1$ here. Clearly, for $j = k = 0$, we have $\psi_{0,0}(t) = \psi(t)$.

In addition to the mentioned approach, which introduces DWT as the counterpart of the CWT, the second approach, however, defines the DWT independently based on the Haar scaling functions and wavelet functions and the idea of nested spaces. In the following, this approach is studied. It will be shown how DWT can be applied to EOG signals during preprocessing step to perform noise and drift removal.

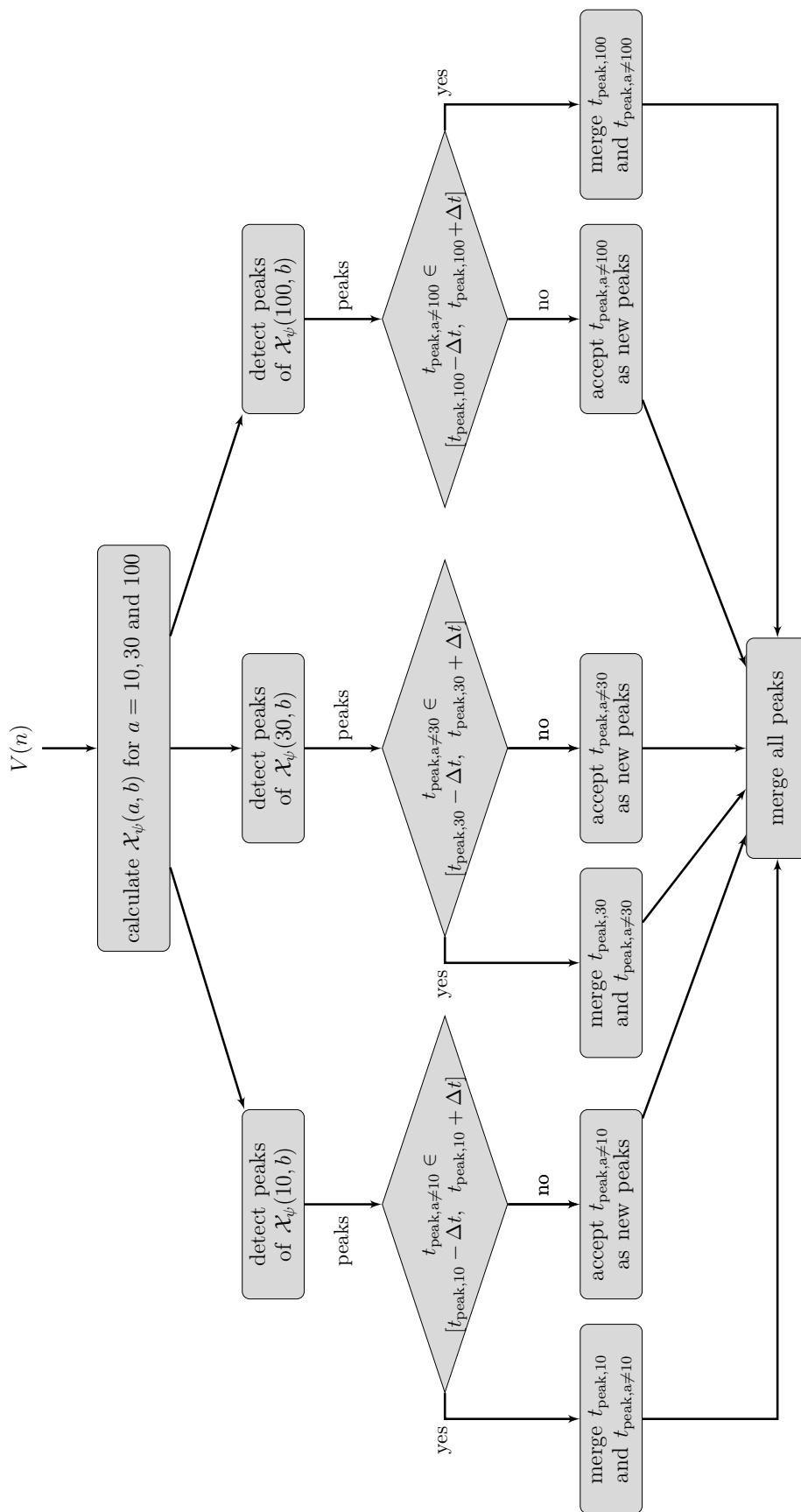


Figure 5.18.: Flow chart of the cwt-based method for blink detection

Haar scaling function

The *Haar scaling function* $\phi(t)$ is defined as

$$\phi(t) = \begin{cases} 1 & 0 \leq t \leq 1 \\ 0 & \text{otherwise.} \end{cases} \quad (5.11)$$

It is clear that the product of $\phi(t - m)$ and $\phi(t - n)$ for $m \neq n$ and $m, n \in \mathbb{Z}$ is equal to zero for all values of t . This expresses the *orthonormality* for the set of time-shifts of $\phi(t)$, namely

$$\int_{-\infty}^{+\infty} \phi(t - m) \phi(t - n) dt = 0 \quad \text{for } m \neq n \text{ and } m, n \in \mathbb{Z} \quad (5.12)$$

$$\int_{-\infty}^{+\infty} \phi(t - m) \phi(t - m) dt = 1. \quad (5.13)$$

For $j = 0$, the mentioned set spans the space \mathbf{V}_0 as

$$\mathbf{V}_0 \equiv \overline{\text{Span}_k \{\phi(t - k)\}} = \overline{\text{Span}_k \{\phi_{0,k}\}}, \quad (5.14)$$

where $k \in \mathbb{Z}$ and $\overline{\text{Span}\{\cdot\}}$ denotes the closed space¹. In addition to the previous space, for $j = 1$, the space \mathbf{V}_1 can be defined which contains $\phi(t - k)$ scaled by 2, namely

$$\mathbf{V}_1 \equiv \overline{\text{Span}_k \{\sqrt{2} \phi(2t - k)\}} = \overline{\text{Span}_k \{\phi_{1,k}\}}. \quad (5.15)$$

The factor $\sqrt{2}$ is a normalization constant which keeps the energy of signals in both spaces similar.

$\phi(t - k)$ in \mathbf{V}_0 can easily be represented by the shrunk $\phi(2t - k)$ in \mathbf{V}_1 leading to $\mathbf{V}_0 \subset \mathbf{V}_1$, e.g.

$$\phi(t) = \phi(2t) + \phi(2t - 1) \quad (5.16)$$

$$\phi(t - 1) = \phi(2t - 2) + \phi(2t - 3). \quad (5.17)$$

Similarly, for the j -th scale, the space is defined as

$$\mathbf{V}_j \equiv \overline{\text{Span}_k \{2^{j/2} \phi(2^j t - k)\}} = \overline{\text{Span}_k \{\phi_{j,k}\}} \quad (5.18)$$

which leads to

$$\cdots \mathbf{V}_{-1} \subset \mathbf{V}_0 \subset \mathbf{V}_1 \subset \cdots \subset \mathbf{V}_\infty. \quad (5.19)$$

This is depicted in Figure 5.19(a). $\mathbf{V}_{j-1} \subset \mathbf{V}_j$ implies that in comparison to \mathbf{V}_j , a part is missing in \mathbf{V}_{j-1} . This missing part can be explained in terms of the *Haar wavelet function*.

In general, $\mathbf{V}_0 \subset \mathbf{V}_1$ leads to the following expression of $\phi(t)$

$$\phi(t) = \sum_n h(n) \sqrt{2} \phi(2t - n), \quad (5.20)$$

¹The space containing the set of functions expressed by the linear combination of $\phi(t - k)$ is called the *span* of the basis set $\phi(t - k)$. If the space comprises the limits of the expansions as well, then it is a *closed* space.

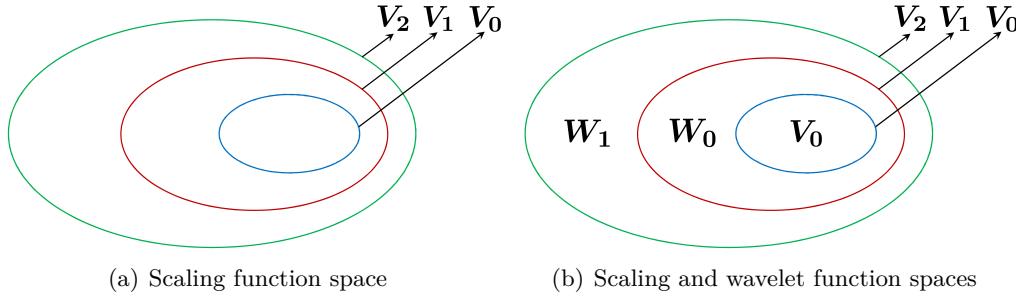


Figure 5.19.: Schematic of spaces spanned by scaling and wavelet functions

where $n \in \mathbb{Z}$ and $h(n)$ denotes the normalized coefficients. Accordingly, in example (5.16), we have $h(0) = h(1) = \frac{1}{\sqrt{2}}$ such that $\phi(t) = \frac{1}{\sqrt{2}}\sqrt{2}\phi(2t) + \frac{1}{\sqrt{2}}\sqrt{2}\phi(2t - 1)$.

Haar wavelet function

The *Haar wavelet function* $\psi(t)$ is defined as

$$\psi(t) = \begin{cases} 1 & 0 \leq t \leq \frac{1}{2} \\ -1 & \frac{1}{2} \leq t \leq 1 \\ 0 & \text{otherwise.} \end{cases} \quad (5.21)$$

Obviously, the set of translates of $\psi(t)$, namely $\psi(t - k)$, $k \in \mathbb{Z}$ also contains orthonormal pairs.

Similar to \mathbf{V}_0 and \mathbf{V}_1 , \mathbf{W}_0 and \mathbf{W}_1 can be defined which are spanned by $\psi(t - k)$ and $\psi(2t - k)$, i.e.

$$\mathbf{W}_0 \equiv \overline{\text{Span}_k \{\psi(t - k)\}} \quad (5.22)$$

$$\mathbf{W}_1 \equiv \overline{\text{Span}_k \{\sqrt{2}\psi(2t - k)\}}. \quad (5.23)$$

However, due to the characteristics of $\psi(t)$, we have $\mathbf{W}_0 \not\subset \mathbf{W}_1$ which means that the translates of $\psi(t)$ cannot be represented by translates of $\psi(2t)$ despite the fact that $\mathbf{W}_0 \perp \mathbf{W}_1$. Generalization of this point to other spaces formed by the j -th scale of $\psi(t)$ leads to

$$\cdots \mathbf{W}_{-1} \perp \mathbf{W}_0 \perp \mathbf{W}_1 \perp \mathbf{W}_2 \cdots. \quad (5.24)$$

Previously, it was mentioned that since $\mathbf{V}_0 \subset \mathbf{V}_1$, only the members of \mathbf{V}_0 can be represented by the members of \mathbf{V}_1 . Now, however, by adding members of \mathbf{W}_0 to \mathbf{V}_0 , it is also possible to represent functions $\phi(2t - k)$ of \mathbf{V}_1 by members of \mathbf{W}_0 and \mathbf{V}_0 , e.g. $\phi(2t) = \frac{\phi(t) + \psi(t)}{2}$. In fact, \mathbf{W}_0 is the missing part in \mathbf{V}_0 which makes \mathbf{V}_0 to be a subset of \mathbf{V}_1 . Mathematically, it is expressed as $\mathbf{V}_1 = \mathbf{V}_0 \oplus \mathbf{W}_0$ with $\mathbf{V}_0 \perp \mathbf{W}_0$. \oplus denotes the direct sum of the vector spaces. Similarly, for \mathbf{V}_2 we have $\mathbf{V}_2 = \mathbf{V}_1 \oplus \mathbf{W}_1$. This is shown schematically in Figure 5.19(b). Finally, generalization

of this truth for the j -th scale expresses \mathbf{V}_j as

$$\begin{aligned}\mathbf{V}_j &= \mathbf{V}_{j-1} \oplus \mathbf{W}_{j-1} \\ &= \mathbf{V}_{j-2} \oplus \mathbf{W}_{j-2} \oplus \mathbf{W}_{j-1} \\ &\quad \vdots \\ \mathbf{V}_j &= \mathbf{V}_0 \oplus \mathbf{W}_0 \oplus \cdots \oplus \mathbf{W}_{j-2} \oplus \mathbf{W}_{j-1}\end{aligned}\tag{5.25}$$

(5.19) and (5.25) make the *multi-resolution analysis* by the DWT possible.

Since \mathbf{W}_0 is also a subset of \mathbf{V}_1 , namely $\mathbf{W}_0 \subset \mathbf{V}_1$, $\psi(t)$ can also be represented by functions of \mathbf{V}_1 such as $\psi(t) = \frac{1}{\sqrt{2}}\sqrt{2}\phi(2t) - \frac{1}{\sqrt{2}}\sqrt{2}\phi(2t-1)$. Consequently, similar to (5.20), $\psi(t)$ can also be expressed as

$$\psi(t) = \sum_n g(n)\sqrt{2}\phi(2t-n),\tag{5.26}$$

where $n \in \mathbb{Z}$ and $g(n)$ refers to the normalized coefficients. By scaling and translating $\phi(t)$ and $\psi(t)$ in (5.20) and (5.26) by $t \rightarrow 2^j t - k$, we have

$$\phi(2^j t - k) = \sum_n h(n)\sqrt{2}\phi(2^{j+1}t - 2k - n)\tag{5.27}$$

$$\psi(2^j t - k) = \sum_n g(n)\sqrt{2}\phi(2^{j+1}t - 2k - n).\tag{5.28}$$

By replacing $m = 2k + n$, we have

$$\phi(2^j t - k) = \sum_m h(m - 2k)\sqrt{2}\phi(2^{j+1}t - m)\tag{5.29}$$

$$\psi(2^j t - k) = \sum_m g(m - 2k)\sqrt{2}\phi(2^{j+1}t - m).\tag{5.30}$$

According to (5.18) and (5.20), a signal $x(t)$ of space \mathbf{V}_{j+1} , namely $x(t) \in \mathbf{V}_{j+1}$, is defined as

$$x(t) = \sum_k c_{j+1}(k) 2^{\frac{j+1}{2}} \phi(2^{j+1}t - k),\tag{5.31}$$

where $c_{j+1}(k)$ denotes the *approximation coefficient*. In fact, in (5.31), $x(t)$ is being approximated by a set of scaling functions. Based on (5.25), which implies $\mathbf{V}_{j+1} = \mathbf{V}_j \oplus \mathbf{W}_j$, $x(t)$ can also be expressed by the functions $\phi(t)$ and $\psi(t)$ of the lower scale j as follows

$$x(t) = \sum_k c_j(k) 2^{\frac{j}{2}} \phi(2^j t - k) + \sum_k d_j(k) 2^{\frac{j}{2}} \psi(2^j t - k),\tag{5.32}$$

where $d_j(k)$ refers to the *detail coefficient*.

In order to calculate $c_j(k)$, the projection of $x(t)$ on $2^{\frac{j}{2}} \phi(2^j t - k)$ is calculated, namely

$$c_j(k) = \int x(t) 2^{j/2} \phi(2^j t - k) dt.\tag{5.33}$$

By replacing (5.29) in the previous equation we have

$$c_j(k) = \sum_m h(m - 2k) \int x(t) 2^{(j+1)/2} \phi(2^{j+1}t - m) dt. \quad (5.34)$$

By comparing (5.34) with (5.33), $c_j(k)$ and similarly $d_j(k)$ are defined as

$$c_j(k) = \sum_m h(m - 2k) c_{j+1}(m) \quad (5.35)$$

$$d_j(k) = \sum_m g(m - 2k) d_{j+1}(m). \quad (5.36)$$

(5.35) and (5.36) are very similar to the definition of discrete convolution and the concept of digital filtering which is defined as

$$y(n) = \sum_k u(n - k) x(k). \quad (5.37)$$

Consequently, according to Addison (2010), h and g in the previous equations perform the *low pass* and *high pass* filtering, respectively. Accordingly, $c_j(k)$ and $d_j(k)$ are referred to as coefficients of low pass and high pass filters applied to $x(t)$ in the framework of a digital filter bank. The factor 2 in $h(m - 2k)$ and $g(m - 2k)$ leads to *down sampling* of the signal $x(t)$.

$x(t)$ can, in addition to the representations in (5.31) and (5.32), also be represented in lower scales with regard to (5.25). Therefore, for an upper scale $j = J$ to the lowest scale $j = 0$ we have

$$\begin{aligned} x(t) &= \sum_k c_J(k) \phi_{J,k}(t) \\ &= \sum_k c_{J-1}(k) \phi_{J-1,k}(t) + \sum_k d_{J-1}(k) \psi_{J-1,k}(t) \\ &= \sum_k c_{J-2}(k) \phi_{J-2,k}(t) + \sum_k \sum_{j=J-2}^{J-1} d_j(k) \psi_{j,k}(t) \\ &\quad \vdots \\ &= \sum_k c_0(k) \phi_{0,k}(t) + \sum_k \sum_{j=0}^{J-1} d_j(k) \psi_{j,k}(t). \end{aligned} \quad (5.38)$$

Figure 5.20 pictorially shows the J -stage decomposition of (5.38). The box containing the downward arrow and 2 represents the down sampling of h and g in (5.35) and (5.36).

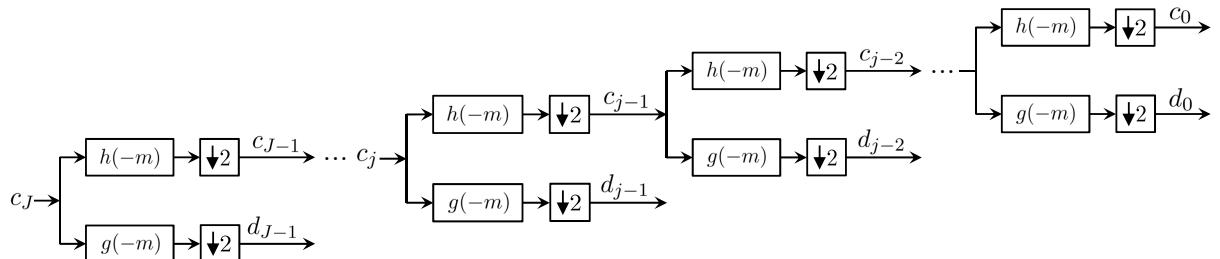


Figure 5.20.: J -stage decomposition tree

Figures 5.21 and 5.22 show the three-stage decomposition of the EOG signal with Daubechies (db4¹) wavelet (see Appendix E) for the awake and drowsy phase, respectively. Both figures show that by increasing the number of decomposition stages j (e.g. $j = 3$ instead of $j = 2$), information loss in the approximation coefficient c_0 increases which is definitely helpful for noise reduction purpose. It can be seen that in the case of $j = 3$, the information loss in c_0 is larger than that of c_1 ². Therefore, depending on the application, one of the c_2 , c_1 and c_0 coefficients in Figure 5.22 are a desired denoised version of the original EOG signal. Noise removal by DWT will be explained in this section.

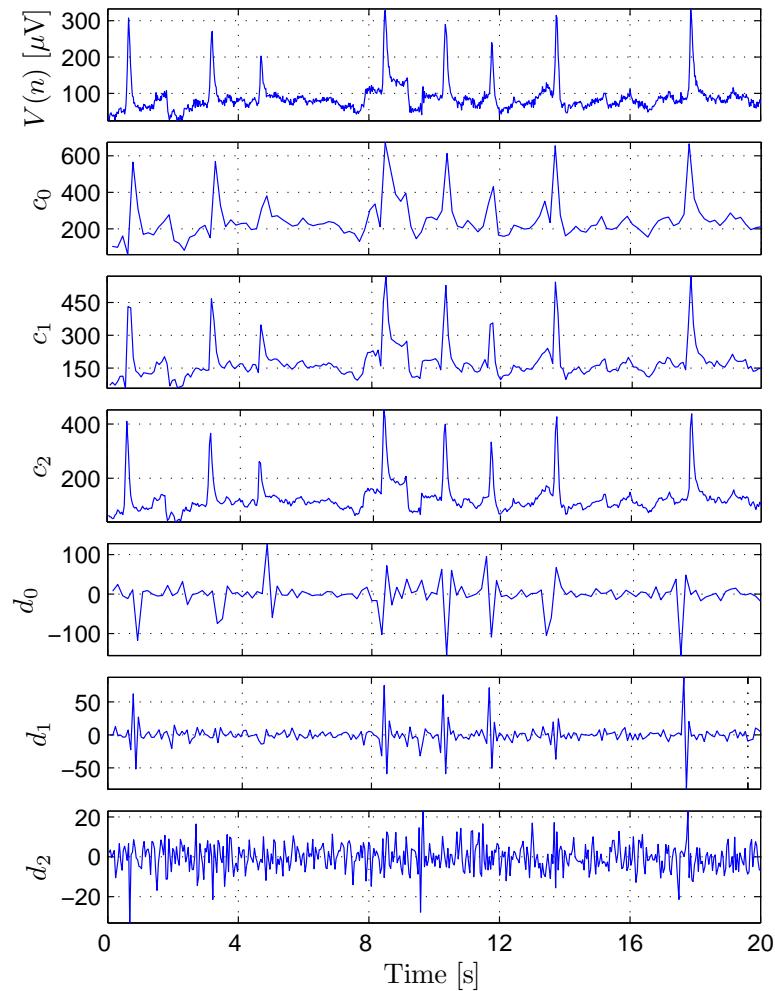


Figure 5.21.: Three-stage decomposition of the EOG signal during the awake phase by db4 wavelet

Reconstruction

As explained in the previous parts, the DWT decomposes the signal into approximation and detail coefficients at a coarse resolution. Similar to other transforms, this transform can be performed inversely to *reconstruct* the original signal. This is explained by considering (5.31) and (5.32) at

¹4th vanishing moment

² c_1 in the case of $j = 3$ corresponds to c_0 in the case of $j = 2$.

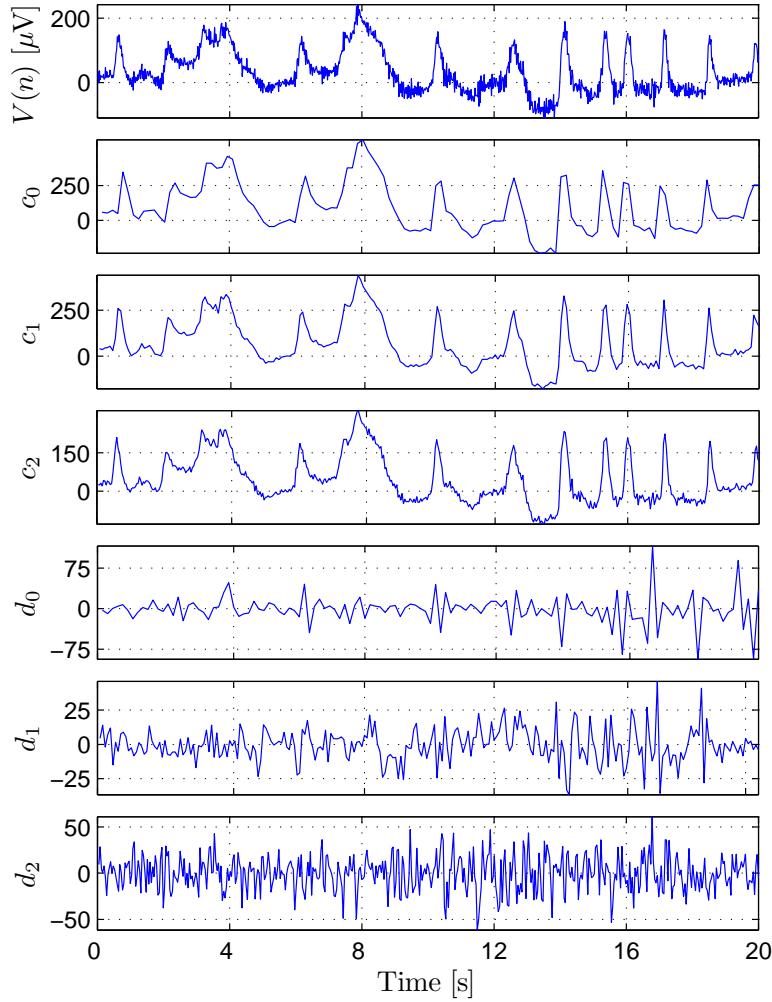


Figure 5.22.: Three-stage decomposition of the EOG signal during the drowsy phase by db4 wavelet

scales $j + 1$ and j . By replacing (5.27) and (5.28) in (5.31) and (5.32) we have

$$\begin{aligned}
 x(t) &= \sum_k c_{j+1}(k) 2^{\frac{j+1}{2}} \phi(2^{j+1}t - k) \\
 &= \sum_k c_j(k) 2^{\frac{j}{2}} \phi(2^j t - k) + \sum_k d_j(k) 2^{\frac{j}{2}} \psi(2^j t - k) \\
 &= \sum_k c_j(k) \sum_n h(n) 2^{\frac{j+1}{2}} \phi(2^{j+1}t - 2k - n) + \sum_k d_j(k) \sum_n g(n) 2^{\frac{j+1}{2}} \phi(2^{j+1}t - 2k - n).
 \end{aligned} \tag{5.39}$$

We multiply both sides of the above equation by $2^{\frac{j+1}{2}} \phi(2^{j+1}t - m)$ and integrate them which leads to (5.33) for $j + 1$, namely $c_{j+1}(k)$ on the left side. The right side equation vanishes due to orthogonality of $\phi(\cdot)$ at all values except for $m = 2k + n$ (or $n = m - 2k$) leading to

$$c_{j+1}(m) = \sum_k c_j(k) h(m - 2k) + \sum_k d_j(k) g(m - 2k). \tag{5.40}$$

The above equation differs from the convolution equation of (5.37) in the $2k$ factor. In fact, for each value of k , only the odd or the even indexed $h(m)$ and $g(m)$ are used. This corresponds to up-sampling of c and d by e.g. adding zero values between existing values and then applying $h(m)$ and $g(m)$ filters. This step can be repeated again for calculating the coefficients of the next stage as visualized in Figure 5.23. Therefore, if all coefficients are used without any changes, the signal under investigation is perfectly reconstructed. However, by removing some or parts of the coefficients c and d , which correspond to e.g. drift or noise in the signal, the reconstructed signal differs from the original one. Thus, this is a very advantageous property for noise or drift removal.

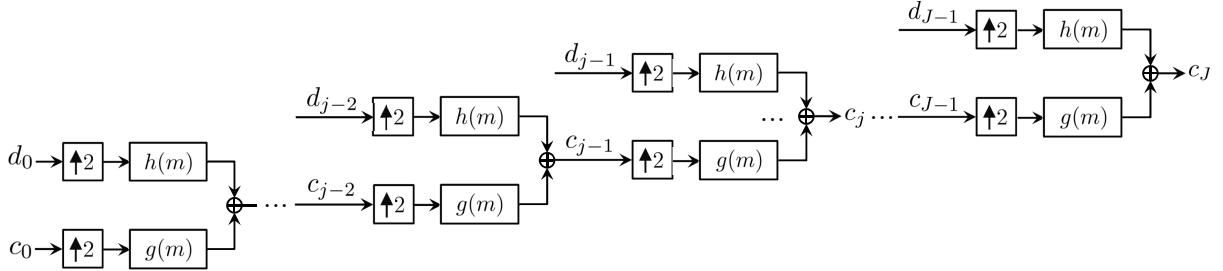


Figure 5.23.: J -stage reconstruction tree

Our approach for adaptive noise removal

As mentioned in previous parts, the multi-resolution analysis by wavelets makes the decomposition of signals possible. Therefore, by decomposing a noisy signal, the noise component can be extracted to a large extent. An example of denoising was shown in Figure 5.22. In this figure, the detail coefficients (especially d_2), unlike the approximation coefficients, seem not to contain relevant information and consequently can be considered as noise. By defining thresholds the contribution of all decomposed coefficients in the reconstruction chain can be controlled. Clearly, the goal is to remove noise and to keep the useful parts of the signals, e.g. eye blink peaks, as good as possible.

As Figures 5.21 and 5.22 show, depending on the noise level, the d_1 coefficient might also be involved in the noise reduction process. Figures 5.24 and 5.25 show examples of the EOG signals denoised by removing only d_2 (Figures 5.24(a) and 5.25(a)) and by removing both d_1 and d_2 (Figures 5.24(b) and 5.25(b)) in the reconstruction step, respectively. In both figures, $\tilde{V}_1(n)$ and $\tilde{V}_2(n)$ represent the denoised versions of $V(n)$. The denoising threshold of $\tilde{V}_1(n)$ was set such that the coefficient d_2 was completely removed during the reconstruction, while other coefficients contributed to the reconstruction step thoroughly. Accordingly, the maximum value of d_2 in Figures 5.21 and 5.22 has been considered as the corresponding threshold, i.e.

$$th_{\text{denoising},d_2} = \max(d_2(n)). \quad (5.41)$$

Hence, all values of d_2 , which were smaller than $th_{\text{denoising},d_2}$, were discarded during the reconstruction step. On the contrary to $\tilde{V}_1(n)$, $\tilde{V}_2(n)$ was calculated by removing both d_2 and d_1 , i.e.

$$th_{\text{denoising},d_2} = \max(d_2(n)) \quad (5.42)$$

$$th_{\text{denoising},d_1} = \max(d_1(n)). \quad (5.43)$$

The residuals are also plotted (third row in Figures 5.24 and 5.25) and are defined as the difference between the original and denoised signals, namely

$$\varepsilon_1(n) = V(n) - \tilde{V}_1(n) \quad (5.44)$$

$$\varepsilon_2(n) = V(n) - \tilde{V}_2(n). \quad (5.45)$$

The PSD of the residuals, namely $\mathcal{E}_1(f)$ and $\mathcal{E}_2(f)$, are also shown in the last rows of Figures 5.24 and 5.25. In the last plot of Figure 5.24(a), the highest power of $\varepsilon_1(n)$ is concentrated above 15 Hz, while in the last plot of Figure 5.24(b), frequencies between 5 to 10 Hz show dominant power. In the time domain, this is evident in larger values of $\varepsilon_2(n)$ at the blink locations which corresponds to a larger information loss of blink peaks in $\tilde{V}_2(n)$ in comparison to $\tilde{V}_1(n)$ (compare $\varepsilon_1(n)$ and $\varepsilon_2(n)$ in Figure 5.24). As a result, removing coefficient d_1 seems not to be necessary for this part of the EOG signal. However, the comparison of $\mathcal{E}_1(f)$ and $\mathcal{E}_2(f)$ in Figures 5.25(a) and 5.25(b) indicates that the information loss between 5 to 10 Hz is still smaller than the noise removed at higher frequencies. Correspondingly, the amplitude of $\varepsilon_2(n)$ is in the same range as that of $\varepsilon_1(n)$. Therefore, removing coefficient d_1 seems not to be critical in this case.

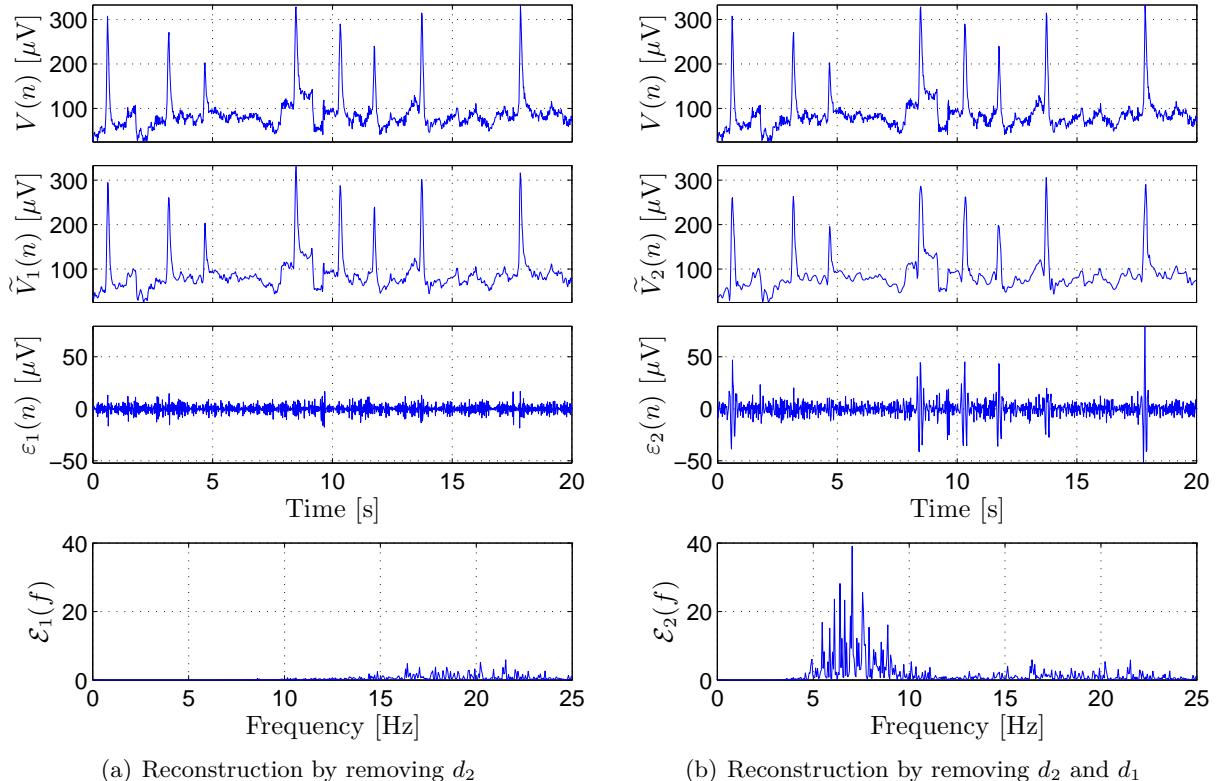


Figure 5.24.: Example 1: denoising of the EOG signal by removing different coefficients during the reconstruction

In fact, these two examples show that an *adaptive* noise removal procedure is needed due to the different levels of noise in different parts of the EOG signal. Otherwise, an inflexible reconstruction strategy causes the following problems:

- by removing both d_1 and d_2 , the information loss of less noisy parts of the EOG signal is inevitable (Figure 5.24(b)).
- by removing only d_2 , noisy parts of the EOG signals are sacrificed to save blink peak

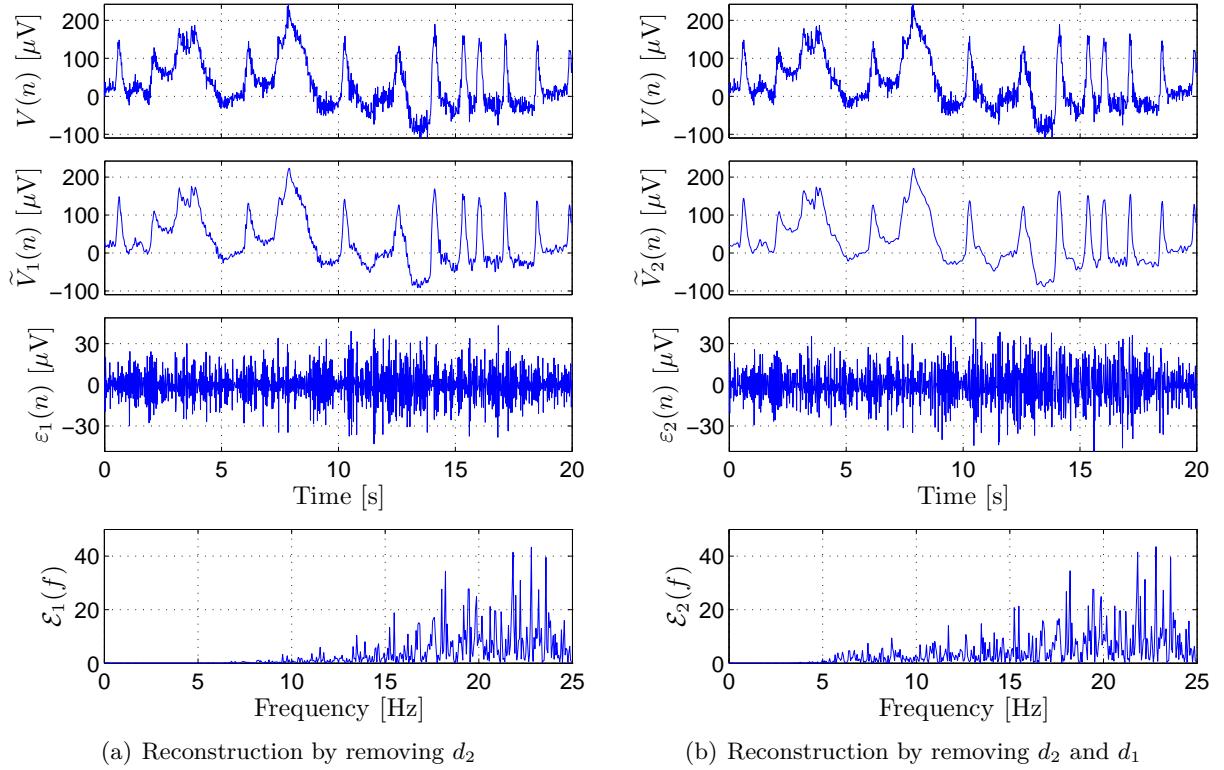


Figure 5.25.: Example 2: denoising of the EOG signal by removing different coefficients during the reconstruction

information of less noisy parts (Figure 5.25(a)).

Thus, in this work, the following compromise was made. The average value of $\mathcal{E}_2(f)$ in two frequency bands, namely $5 \text{ Hz} \leq f \leq 10 \text{ Hz}$ as the low frequency components and $f \geq 15 \text{ Hz}$ as the high frequency components have been compared with each other and the following rules have been applied

$$\text{if } \overline{\mathcal{E}_2(5 \text{ Hz} \leq f \leq 10 \text{ Hz})} \geq \overline{\mathcal{E}_2(f \geq 15 \text{ Hz})} \times 1.1, \quad \text{remove } d_2 \quad (5.46)$$

$$\text{if } \overline{\mathcal{E}_2(5 \text{ Hz} \leq f \leq 10 \text{ Hz})} < \overline{\mathcal{E}_2(f \geq 15 \text{ Hz})} \times 1.1, \quad \text{remove } d_2 \text{ and } d_1 \quad (5.47)$$

where $\overline{\mathcal{E}_2(f)}$ denotes the average of $\mathcal{E}_2(f)$. According to these rules, in Figure 5.24, only the coefficient d_2 is removed in the reconstruction step, while in Figure 5.25 both d_1 and d_2 coefficients are removed.

Figure 5.26 shows the scatter plot of $\varepsilon_1(n)$ and $\varepsilon_2(n)$ plotted in Figure 5.25. Although different coefficients were removed during the reconstruction step, the values of the residuals are very similar to each other. This underscore the strength of the proposed denoising procedure, i.e. the blink peaks have not been influenced very differently by removing different coefficients.

The introduced denoising procedure has been applied to all collected EOG signals in this work in the framework of preprocessing before applying the proposed eye movement detection methods.

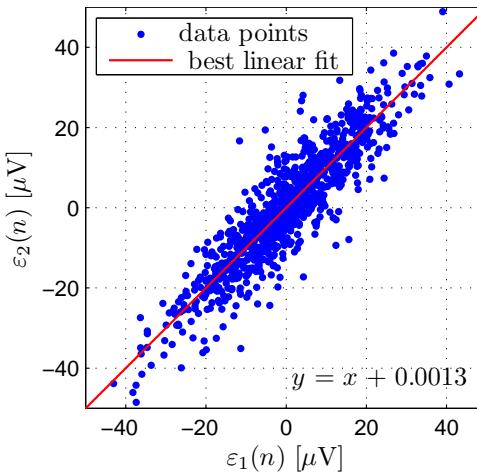


Figure 5.26.: Scatter plot: $\varepsilon_1(n)$ versus $\varepsilon_2(n)$ shown in Figure 5.25

Our approach for drift removal

Similar to the noise removal, multi-resolution analysis makes drift removal possible. Here, the question is which components should be taken into consideration in the reconstruction step in order to only remove the drift. Tinati and Mozaffary (2006) suggested a drift removal method for the ECG signals by calculating the energy of the wavelet decomposition coefficients at different levels and comparing it with a threshold. If the calculated energy is higher than a pre-defined threshold, then the current decomposition level is suitable and the signal can be reconstructed. Otherwise, the number of decomposition stages should be increased.

Based on our observations of the EOG signals, drift is a low frequency component, which is usually represented below 0.3 Hz. Therefore, 6 or 7 decomposition stages should be enough for reconstructing the drift signal. Figure 5.27 shows two examples of $V(n)$ signals in awake (top plot) and drowsy (bottom plot) phases. The drift removal in this figure is based on the same approach used for denoising. The only difference is the coefficients which are used for the reconstruction. For 6- and 7-stage decompositions, we only used the approximation coefficients c_5 and c_6 for reconstruction, respectively, and ignored all other coefficients. The resulting signals of reconstruction are drift signals which are shown in green (6-stage) and magenta (7-stage) in Figure 5.27. By subtracting the drift signal from the $V(n)$ signal, the drift is removed and the result is called $\check{V}(n)$.

In the top plot of Figure 5.27, which refers to the awake phase, it seems that the number of decomposition stages has not deteriorated $V(n)$. The form of blinks in both $\check{V}(n)$ signals is similar to $V(n)$. However, in the bottom plot, which refers to the drowsy phase, the closed phase of the microsleep event located between $t \approx 5$ s and $t \approx 8$ s (shown with an arrow) suffers from a new unwanted deformation after the 6-stage decomposition. This phenomenon also occurred in other similar events and is a representative example. Therefore, we have used the 7-stage decomposition for the rest of this work and considered only c_6 in the reconstruction for extracting the drift. Similar to the noise removal, drift removal was also applied to all collected EOG signals in this work in the framework of preprocessing before applying the proposed eye movement detection methods.

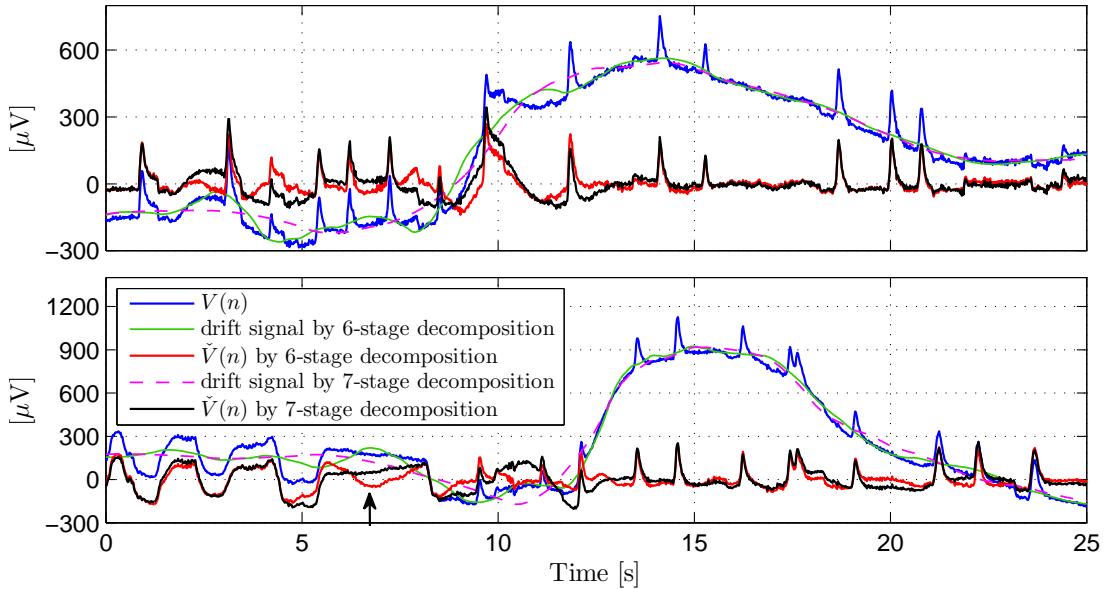


Figure 5.27.: Two examples of drift removal with the wavelet decomposition and reconstruction for awake (top) and drowsy (bottom) phases of $V(n)$

5.4. Comparison of event detection methods

In this section, the introduced event detection methods are evaluated by comparing them with a reference. To this end, we have labeled eye movement events in our offline EOG signals based on the synchronously recorded video data from the subjects' face. By comparing the detected events with the labeled ones, events are then assigned to these three categories: true positive (TP), false positive (FP) and false negative (FN) according to Table 5.1. Clearly, true negative (TN) cannot be assessed in this application.

Table 5.1.: confusion matrix: events of video labeling versus those of the proposed detection methods

		detection	
		event	detected
video labeling	labeled	True Positive (TP)	False Negative (FN)
	not labeled	False Positive (FP)	True Negative (TN)

After counting all detected and missed events, the corresponding metrics for the evaluation of detection methods are calculated. The metrics used in this work are *recall* (RC) and *precision* (PC) which are defined as follows

$$\text{RC} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \quad (5.48)$$

$$\text{PC} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100. \quad (5.49)$$

RC describes the proportion of correctly detected events among true ones ($\text{TP} + \text{FN}$), while PC represents the proportion of correctly identified events among all detected ones ($\text{TP} + \text{FP}$).

5.4.1. Median filter-based versus derivative-based method

First, we compare the performance of the event detection by the median filter-based method introduced in Section 5.1 with that of the proposed derivative-based method in Section 5.2. As mentioned before, both methods are only suitable for detection of fast eye movements. Therefore, only the detection of these eye movements, i.e. fast blinks and vertical saccades, are compared with each other. Figure 5.28 shows the RC and PC values for seven subjects.

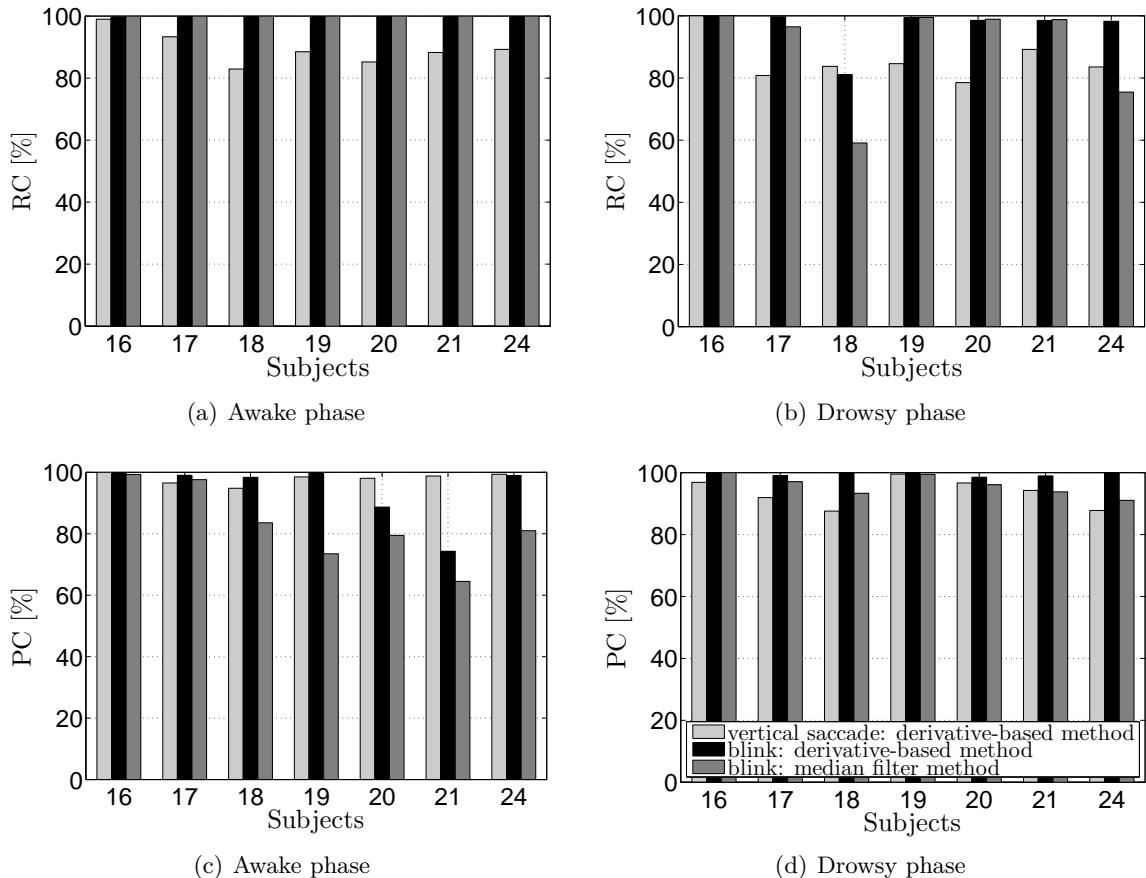


Figure 5.28.: RC and PC of vertical saccade and blink detections for the derivative-based algorithm and the median filter-based method during the awake and drowsy phases

As the goal is the detection of saccades and blinks not only during the awake phase, but also during the drowsy phase of the drive, the RC and PC during these phases were calculated separately with regard to the collected KSS values. This helps to highlight the efficiency of the detection methods, when different forms of events are present in the data. For each phase, 10 min of the collected EOG is evaluated¹. To this end, we defined the awake phase by $KSS \leq 5$ (Figures 5.28(a) and 5.28(c)) and the drowsy phase by $KSS \geq 8$ (Figures 5.28(b) and 5.28(d)). These definitions with the gap of 2 KSS steps, namely $KSS = 6$ and 7 , emphasize the difference between events of the two phases under investigation.

The calculated rates show that during the awake phase (Figures 5.28(a) and 5.28(c)), both the proposed algorithm and the median filter-based method detected all true blinks correctly (all RCS = 100% in Figure 5.28(a)). However, for all subjects, the median filter-based method always

¹Only 10 min of the collected EOG was labeled.

had a smaller PC (see Figure 5.28(c)). The reason is that most of the saccades combined with head rotation were wrongly considered as blinks, especially for subject S21. During the drowsy phase (Figures 5.28(b) and 5.28(d)), for subjects S18 and S24, the blink detection using the proposed algorithm outperformed the median filter-based method by about 20%. For saccade detection, the proposed algorithm achieved $PC > 95\%$ and $RC > 80\%$ for all subjects during the awake phase. For subject S16 all existing saccades were detected correctly. Moreover, lower PC values for saccade detection in Figure 5.28(d) in comparison to Figure 5.28(c) imply that the saccade detection during the drowsy phase is more difficult, since small amplitude blinks due to drowsiness might be mistaken for saccades.

It should be mentioned that not only the detection of events but also the quality of the extracted features out of the detected events plays a key role in assessing drowsiness. The blink amplitude as defined in (5.3) and the duration as the time difference between points C and A in Figure 5.3 were extracted from all detected blinks using the median filter-based method and the proposed algorithm. The moving average of these features over 15-min windows is shown in Figure 5.29 for subjects S15, S16 and S18 as representative examples. In addition, the numbers of blinks per minute are shown in the last row. The background colors refer to the self-rated drowsiness level by the subjects as awake ($KSS \leq 5$), medium ($6 \leq KSS \leq 7$) and drowsy ($KSS \geq 8$). In most of the plots, the ranges of the calculated features are different with respect to the applied detection methods.

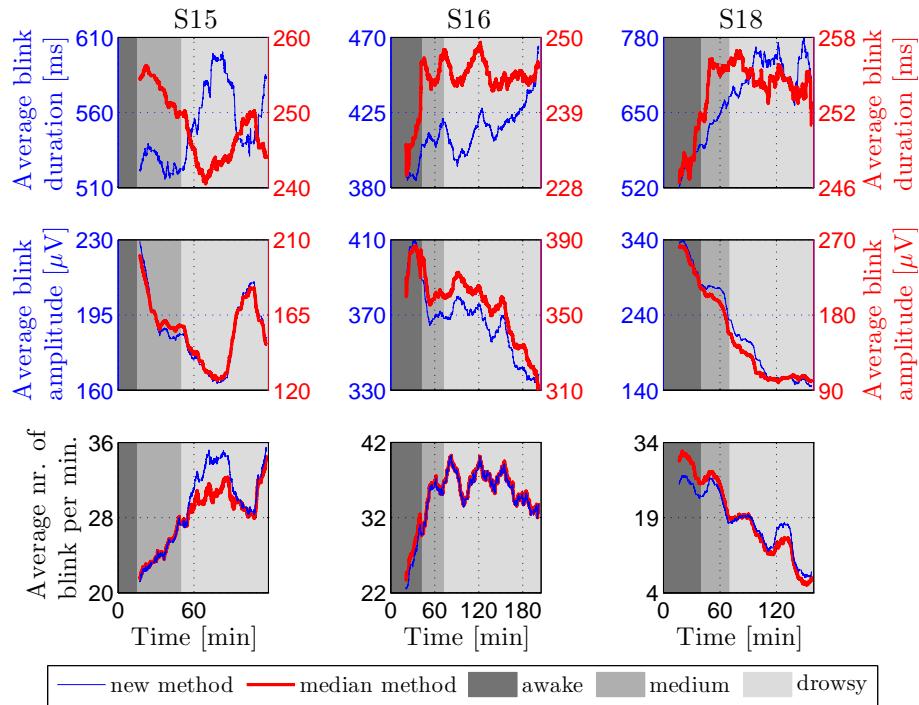


Figure 5.29.: Average duration (first row), amplitude (second row) and number of blinks (third row) versus self-estimated drowsiness level for subjects S15, S16 and S18 based on the derivative-based algorithm and the median filter-based method

For subject S15, the amplitudes of detected blinks by both methods are very similar. Nevertheless, the evolution of blink duration runs counter to the evolution of drowsiness using the median filter-based method for about one hour. This means that during this time, based on the blink detection results using the median filter-based method, blink duration is negatively correlated with drowsiness, while the proposed algorithm shows positive correlation. The reason is that

although the median filter-based method has detected some of the blinks during the drowsy phase, the start and end points of them were not extracted correctly as also shown in Figure 5.2. Moreover, for about 30 min, a large number of blinks was not detected (see left plot of last row in Figure 5.29). The differences in the evolution of blink amplitude, duration and number plotted for S15 are similar to the subjects S15, S19, S20, S21 and S24. For S16, however, it is the number of detected blinks which is similar for both methods. This is also shown in Figure 5.28 where RC and PC were very close to each other. Nevertheless, the increasing behavior of blink duration is very weak using the median filter-based method in comparison to that of the proposed algorithm. The features of subjects S22 and S23 also have the same differences as subject S16. For subject S18, despite an equal number of detected blinks during some parts of the drive using both methods, the extracted amplitudes and durations of these parts also deviate from one another. Such deviation is also the case for subjects S19 and S25.

According to Figure 5.29, the average number of blinks has increased for subjects S15 and S16, while for subject S18, it has decreased. One explanation is that, the average duration of the blinks for subject S18 increased to a larger extent in the course of time in comparison to other subjects. Therefore, in comparison to other subjects, the eyes of subject S18 were closed for a longer time during the experiment which leads to a smaller number of blinks.

All in all, according to Figure 5.29, the correlation between self-estimated drowsiness level and extracted features of eye movements based on the proposed algorithm seems to be strong enough to assess drowsiness, especially since similar results were also achieved in previous studies (Dong et al., 2011) (decreased blink amplitude and increased blink duration in the course of time). Extracting other features and applying complex classification methods are the next steps toward drowsiness detection based on eye movements which will be studied in Chapters 7 and 8.

Due to unsatisfactory results of the median filter-based method for both blinks and saccade detection and their corresponding features, it will not be analyzed further in this work. In addition, as mentioned before, the median filter-based method is unable to detect slow blinks.

5.4.2. Derivative-based method versus wavelet transform-based method

This section compares the RC and PC of the blinks detected by the derivative-based method (Section 5.2) and continuous wavelet transform method (Section 5.3.2) for the 10 subjects under study. We expect that for the awake phase, both algorithms perform similarly. On the contrary, depending on the number of slow blinks during the drowsy phase, the wavelet transform method is expected to outperform the derivative-based method.

Figure 5.30 shows the calculated RC and PC values for 10 subjects regarding both methods. Subject S3 had no awake phase according to his KSS values. These results are based on 20 min of labeled EOG data per subject, i.e. 10 min for each phase. First, for each subject, based on the KSS values, the awake and drowsy segments of the drive were defined. Afterwards, for each phase, 10 1-min segments were randomly chosen for labeling and further evaluation. On average, for the awake phase 263 and for the drowsy phase 453 blinks were labeled per subject.

As expected, during the awake phase, both methods are very similar in detection of fast blinks (all RCS > 95%) and the differences are negligible. Interestingly, the good results of the wavelet transform method are obtained at a cost of lower PC values which are the result of confusing saccades combined with head rotation with blinks. For the drowsy phase, however, for most of the subjects the wavelet method has detected more blinks, especially for subject S2. This

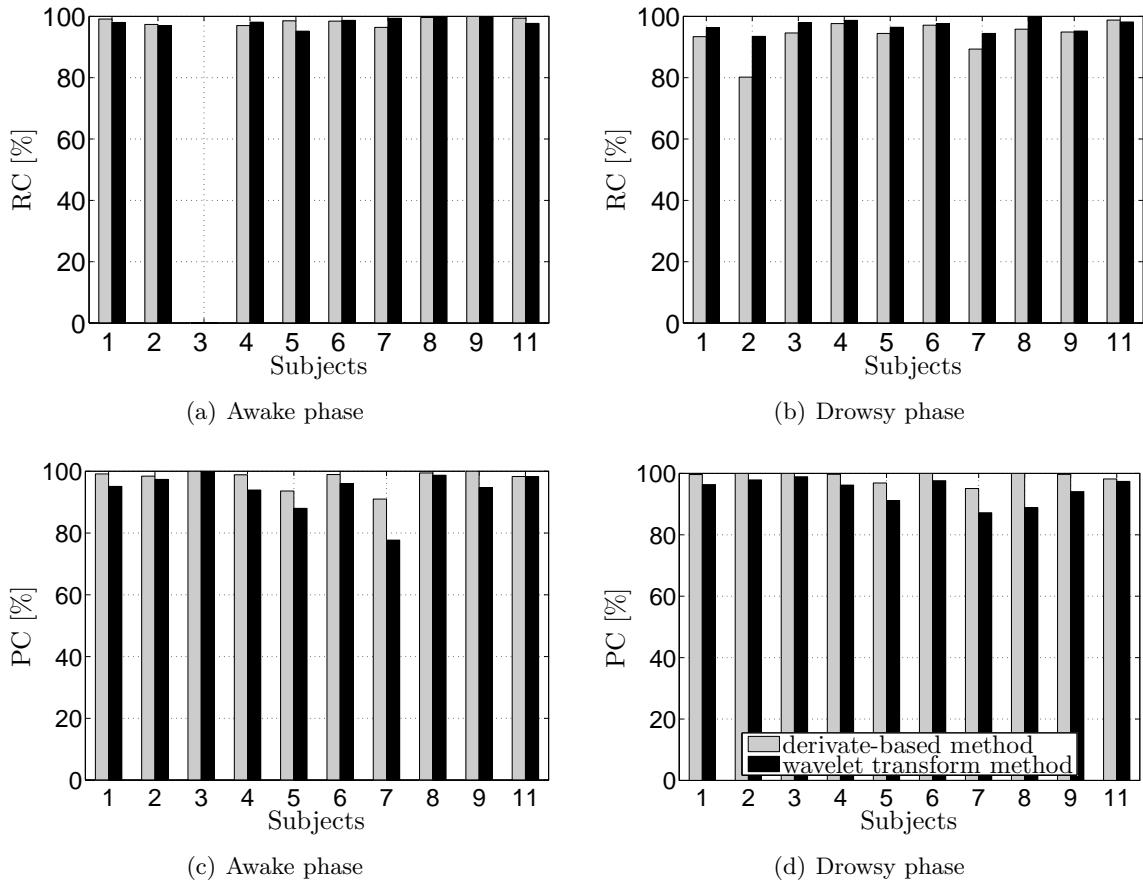


Figure 5.30.: RC and PC of blink detection for the derivate based algorithm and the wavelet transform method during the awake and drowsy phases

subject showed lots of slow blinks during the drowsy phase. The PC values of this phase are also smaller after applying the wavelet transform method.

It should be mentioned that according to our observations during the experiments, slow blinks might not occur for all subjects to the same extent during drowsiness. For some of our subjects, almost no slow blinks occurred during the experiment, although the subjects experienced lots of microsleeps. For these subjects, only the duration of the closed phase increased as shown in Figure 3.4(b). The velocity of the opening and closing phases varied to a smaller extent. For some other subjects, however, as shown in Figure 3.4(c), it was the velocity of the opening and closing phases which varied severely due to drowsiness in comparison to the duration of the closed phase. Therefore, depending on the characteristics of the blinks during the drowsy phase, different detection approaches should be applied.

Conclusions

All in all, we conclude that if the application is limited to the detection of sharp blinks, where saccades are considered *noisy* eye movements to be ignored, the median filter-based method is an appropriate detection method to be applied to the EOG or similar signals. As soon as the vertical saccades should be recognized and distinguished from blinks, we suggest the usage of the derivative-based approach. In addition, the derivative signal method is more robust against the

duration of the event to be detected in comparison to the median filter-based method. Therefore, for the detection of blinks, whose duration varies in the course of time, the median filter-based method with a fixed window length is not suitable.

Further, it was shown to be possible to detect vertical saccades and blinks simultaneously in vertical EOG signals based on the derivative signal. In addition, a 3-means clustering algorithm is recommended to distinguish between saccades and blinks in those applications where the data of both awake and drowsy phases are available. This helps to prevent confusing a driver's decreased amplitude blinks with saccades or other eye movements. Moreover, blinks with long eye closure and microsleep events, whose patterns deviate from those during the awake phase, were detected and distinguished from saccades based on the statistical distribution of the amplitude. This method is a reliable approach as long as the velocity of eye lid movements is larger than the velocity of non-relevant phases of the EOG signal.

We saw that slow blinks are also another issue, if the derivative signal is used for their detection. In order to detect all fast and slow blinks simultaneously, we suggest the continuous wavelet transform method. This method covers the detection of all types of blinks by selecting appropriate scaling values a . However, by applying the wavelet transform approach, smaller values of the PC in both awake and drowsy phases are expected in comparison to the derivative-based detection method as shown in Figures 5.30(c) and 5.30(d).

Online versus offline eye movement detection methods

Another aspect for comparing each of the studied methods with each other is whether they can be implemented online to be evaluated directly during the experiment. This is the case for the median filter-based method, which was implemented by SIMULINK to be applied online. The derivative-based method, however, has lots of parameters, which should be adjusted individually based on the entire available data. The question is, how many events are needed for finding a suitable threshold, e.g. for distinguishing between blinks and saccades (see Figure 5.4). To answer this question, we implemented the derivative-based detection algorithm online and updated the mentioned threshold for each subject after detecting 20 new events. In other words, after detecting 20 new events, the threshold was recalculated with respect to all available detected events up to that instant of the time. Figure 5.31 shows two representative results of the found online thresholds $th_{2\text{-means}}$ for subjects S36 and S39.

These two subjects were awake during the whole drive and did not suffer from lack of vigilance. This information is based on the collected KSS data and the offline video analysis of the drives. The area of $\pm 15 \mu\text{V}$ tolerance around the true threshold, namely offline calculated $th_{2\text{-means}}$, is also shown in dashed lines. For subject S36, the first values of the found online thresholds are not reliable, because they fluctuate to a large extent due to lack of available events in the clusters, i.e. vertical saccades versus blinks. After detecting 160 events, the online calculated $th_{2\text{-means}}$ is very close to the true offline $th_{2\text{-means}}$. On the contrary, for subject S39, the online $th_{2\text{-means}}$ does not fluctuate at the beginning, but starts with a large value and converges to the offline $th_{2\text{-means}}$ after roughly 2000 events.

It will be explained in Section 7.2 that the number of blinks for humans is between 15 and 20 blinks per minute. If we consider 5 vertical saccades per minute as well on average, in total 25 events should be available for detection per minute. As a result, the found values of 160 and 2000 events correspond to about 6 and 80 min of EOG data or drive time, respectively.

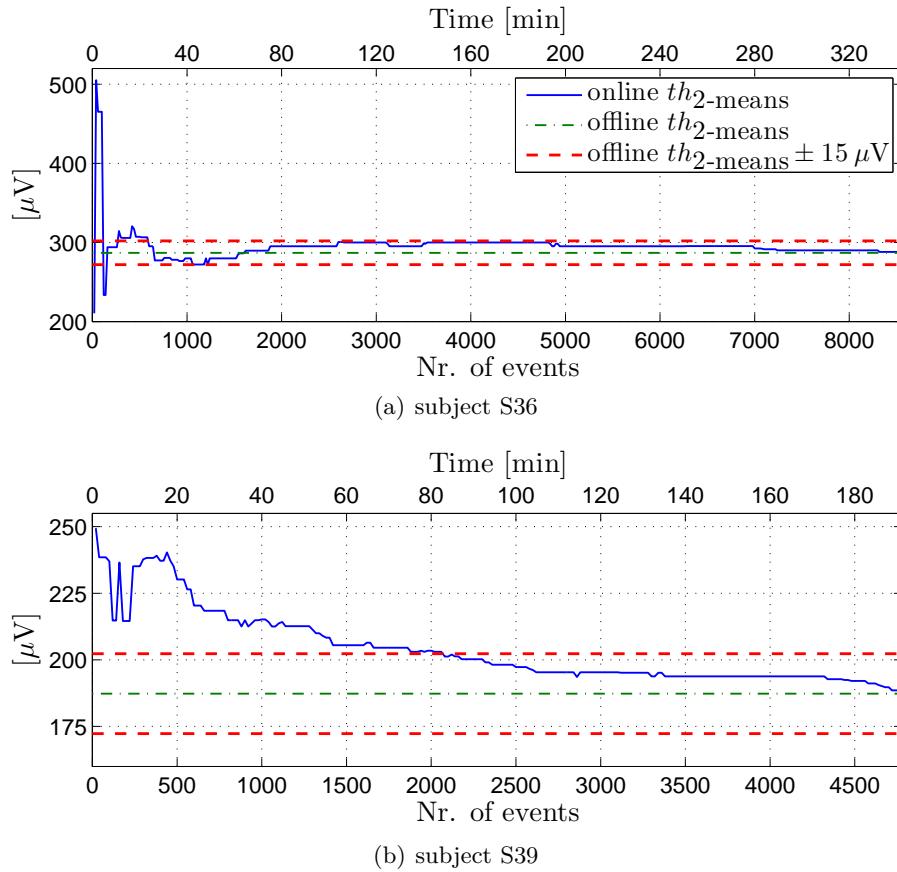


Figure 5.31.: Setting the threshold for distinguishing between blinks and vertical saccades in an online implementation of the detection method

The detection algorithm based on the continuous wavelet transform method is indeed more time-consuming for the online application, because several coefficients have to be calculated in addition to the peak detection procedure. Therefore, among these methods, the derivative-based eye movement detection method seems to detect events more precisely with regard to the time needed by the algorithm in its online implementation.

Final applied algorithms for preprocessing of EOG signals and eye movement detection

In previous sections, we studied three eye movement detection methods and discussed the strengths and weaknesses of them in detail. In addition, due to the decomposition and reconstruction properties of DWT, two preprocessing steps for EOG data were investigated. As mentioned before, prior to applying any of the detection methods, first, we preprocessed all collected EOG data by applying noise and drift removal algorithms to them. These steps improve the event detection procedure.

For the detection of fast eye movements, we applied the derivative-based algorithm explained in Section 5.2. In order to overcome the shortcoming of the derivative-based algorithm in detection of slow blinks, we combined the wavelet transform method with the detection algorithm based on the derivative signal to improve the DR values. The reason is the smaller values of the PC

in both awake and drowsy phases for the wavelet transform approach in comparison to the derivative-based detection method. The combination has been performed such that only slow blink events, which were detected in $\mathcal{X}_\psi(30, b)$ and $\mathcal{X}_\psi(100, b)$ were added to the detected events based on the derivative signal. All events detected by $\mathcal{X}_\psi(10, b)$ were considered to be fast eye movements which were already detected using the derivative signal. In other words, a fast blink, which was only detected using the wavelet transform, was discarded.

6. Blink behavior in distracted and undistracted driving

Among different physiological symptoms for assessing driver drowsiness or inattention, drivers eye movements, in particular blinks, are more associated with drowsiness (Hargutt, 2003; Schleicher et al., 2008; Dong et al., 2011). As an example, previous studies introduced increased blink frequency as a measure related to drowsiness over time (Stern et al., 1994; Summala et al., 1999; Sirevaag and Stern, 2000; Hargutt, 2003).

Can all blinks during driving be interpreted as drowsiness-related patterns? Or is the occurrence of some blinks situation-dependent? Do all blinks during driving reflect the alertness of the driver to the same extent? This chapter aims to answer these questions which describe the situation dependency of the blink occurrence. If the occurrence of some blinks is situation-dependent, then such blinks might (locally) affect the amount of correlation between the extracted blink features and driver's state of vigilance in warning systems. Therefore, they should be taken out of consideration or handled differently for assessing driver drowsiness. Phases of visual or cognitive distractions during data entering into the navigation system or during a mobile phone conversation are examples of situation dependency of the blink occurrence. In this context, Liang and Lee (2010) reported increased blink frequency in a driving simulator during cognitive and combined cognitive and visual distraction.

In order to investigate the aforementioned questions, first, it is necessary to know which type of blinks is relevant for drowsiness detection. As mentioned in Section 3.3, blinking occurs either voluntarily or involuntarily, i.e. spontaneously and reflexively. Among all blinks, we have mentioned that spontaneous blinks are the relevant blinks to be explored for drowsiness detection.

In this chapter, we study the blink behavior of subjects who participated in the experiment described in Section 4.2.2. These subjects performed the primary driving task and the visuomotor or auditory secondary tasks. This means that they were instructed to experience fixed-position gaze shifts, namely gaze shifts between pre-defined fixed positions. These gaze shifts were the result of the visuomotor secondary task which represents visual distraction.

Evinger et al. (1994, page 342) stated based on Beideman and Stern (1977):

Gaze-evoked blinks rarely occur when making a saccade to a rear view mirror before shifting lanes, but often accompany the saccade back to straight ahead position.

Inspired by this fact, it is also studied whether the occurrence of gaze shift-induced blinks is direction-dependent.

Apart from the mentioned points, during undistracted driving, blinks mainly occur spontaneously or due to gaze shifts, e.g. during a look in the rear-view mirror. Moreover, we know that the driver needs to scan the scene and the road ahead frequently to perform the driving task properly. This raises a question: which gaze shifts (with respect to their amplitude) are more probably accompanied by a blink during driving (without performing a secondary task)?

This question will be investigated by following the experiment described in Section 4.3, in which eye movements of drivers were collected by the EOG in a driving simulator. In this experiment, we specifically study the relationship between the amount of gaze shift and the blink occurrence. The results of this chapter are taken from Ebrahim et al. (2013c). All blinks and saccades were detected based on the algorithm described in Section 5.2 out of the EOG signals.

6.1. Time-on-task analysis of the saccade rate during the visuomotor task

In this section, it is studied whether the saccade rate (number of saccades per minute) during the visuomotor secondary task changed over time, i.e. the four blocks of the experiment. For investigation of the variable time-on-task, first, the number of all detected saccades out of the $H(n)$ signal during the visuomotor secondary task was calculated. As mentioned in Section 4.2.2, this task was repeated four times by each subject and each time for 3 min. The saccade rate was calculated during each block for each subject separately. Figure 6.1(a) shows the mean and the standard deviation of the calculated values for each block over all subjects.

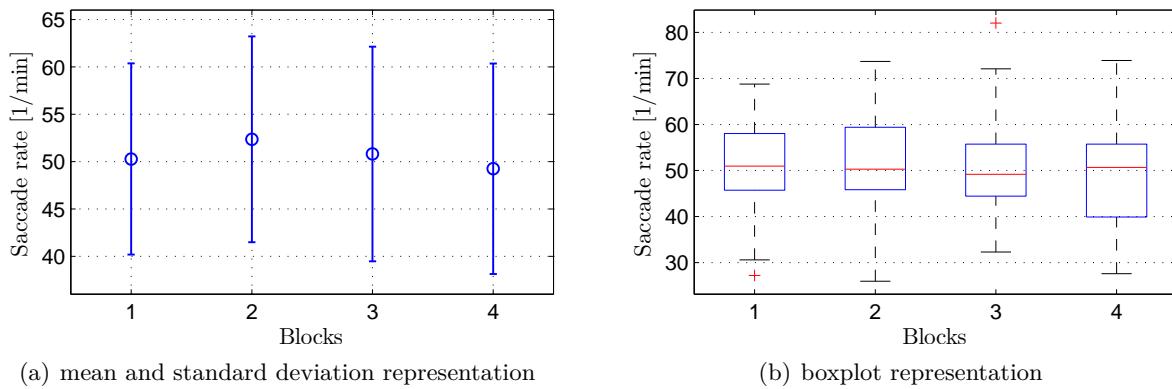


Figure 6.1.: Saccade rate for the variable time-on-task (four blocks) for all subjects

By applying the analysis of variance (one-way repeated measures ANOVA), explained in Appendix D.5, it is examined whether the saccade rate has changed significantly over time. It should be mentioned that all assumptions of the ANOVA are fulfilled. They are as follows: independency¹ and normality of observations (see Appendix D.2) and the homogeneity of their variances, i.e. sphericity (see Appendix D.6). The corresponding H_0 hypothesis states that the means of different groups of measurements (blocks) are equal. Since the calculated p -value is 0.30 at the confidence level of 95% ($\alpha = 0.05$) with $F_{3,75} = 1.24$, the H_0 hypothesis is not rejected ($0.30 > 0.05$). Consequently, we cannot show the dependency of the numbers of gaze shifts during the visuomotor task upon the variable time-on-task considering all subjects and all blocks.

6.2. Time-on-task analysis of the blink rate

Similar to the previous step, here we explore whether the blink rate (number of blinks per minute) changed over time (four blocks) during the primary and secondary tasks separately.

¹We suppose that the occurrence of an eye movement, e.g. a saccade, at a time instant is independent of the occurrence of another saccade at a different instant.

The calculated values by applying the ANOVA for repeated measures are $F_{3,75} = 1.05$ with p -value = 0.37 for the visuomotor task, $F_{3,57} = 0.27$ with p -value = 0.85 for the driving task and $F_{3,75} = 1.08$ with p -value = 0.36 for the auditory task. According to the p -values, which are all larger than $\alpha = 0.05$, we cannot show that the difference between the means of blink rates during four blocks are significant.

6.3. Saccades time-locked to blinks during the visuomotor task

This section studies the number of saccades time-locked to blinks, i.e. saccades occurring simultaneously with blinks (see Figure 3.5(c)), during the visuomotor task. A saccade was considered as time-locked to a blink, if it overlapped a blink within at least 80% of its duration. Therefore, all horizontal saccades and all blinks during the visuomotor task were first detected in $H(n)$ and $V(n)$ signals, respectively. Then, the number of saccades accompanied by blinks (in percent) was calculated as shown in Figure 6.2 for each subject and each block separately. On average over all blocks, for 14 subjects (2, 4, 6, 7, 8, 10, 12, 13, 14, 16, 21, 22, 23, and 25) over 80% of all saccades due to the visuomotor task were time-locked to blinks. For 9 subjects (1, 3, 5, 9, 15, 19, 20, 24 and 26) gaze shift-induced blinks accompanied from 50% to 75% of saccades. Interestingly, the calculated percentage was less than 50% for only 3 subjects (11, 17 and 18). All in all, for most of the subjects gaze shift induced the blink occurrence which is in agreement with the statement of Records (1979).

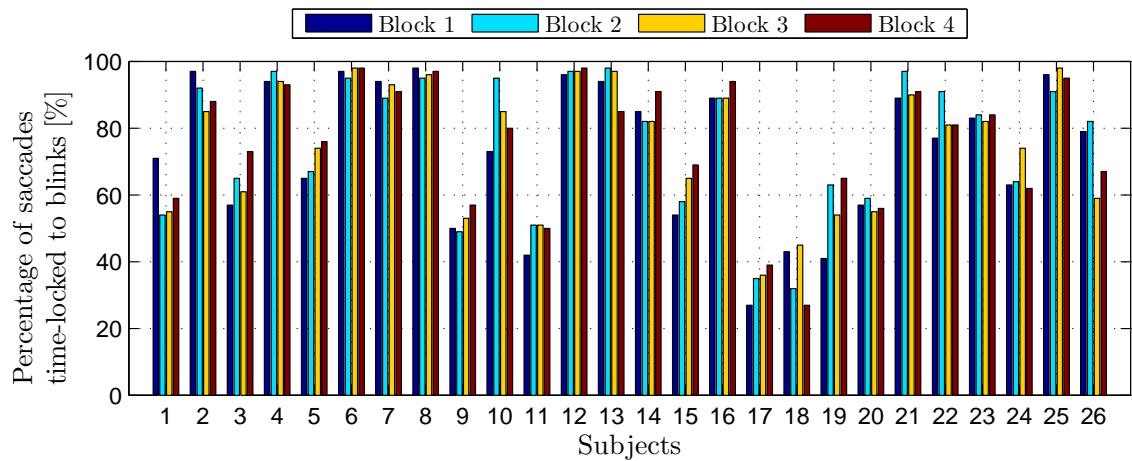


Figure 6.2.: Percentage of saccades time-locked to blinks for all subjects and all blocks during the visuomotor task

6.4. Direction-dependency of blinks time-locked to saccades

As mentioned at the beginning of this chapter, inspired by Evinger et al. (1994), Figure 6.3 shows whether the occurrence of gaze shift-induced blinks depended on the direction of the saccade. On this account, we considered only saccades either towards the road or towards the screen displaying the Landolt rings during the visuomotor task, i.e. saccades occurring between fixed positions. The bars of this figure represent the number of saccades time-locked to blinks (in percent) with respect to their direction. The results of all blocks were combined, because the saccade rate was shown to be independent of the variable time-on-task. With the exception

of subject S19, for all subjects the number of saccades time-locked to blinks was larger while moving focus towards the road (average = $88\% \pm 16.7$). However, for the other direction, namely towards the screen, we found different behaviors (average = $61\% \pm 31.1$).

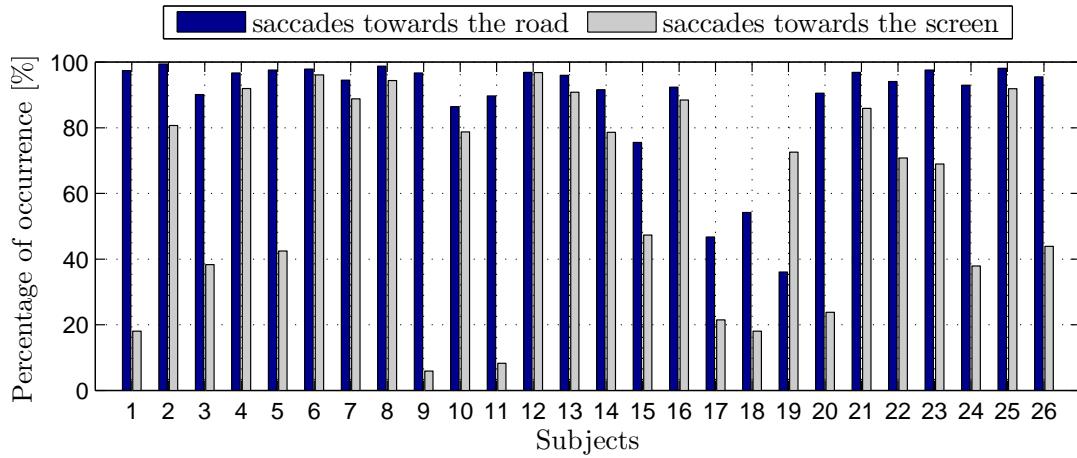


Figure 6.3.: Percentage of saccades time-locked to blinks with respect to saccade direction averaged over all blocks during the visuomotor task

In order to categorize the behaviors, the values of bars of Figure 6.3 are plotted versus each other in Figure 6.4 (dark bars as the *x*-axis, light bars as the *y*-axis).

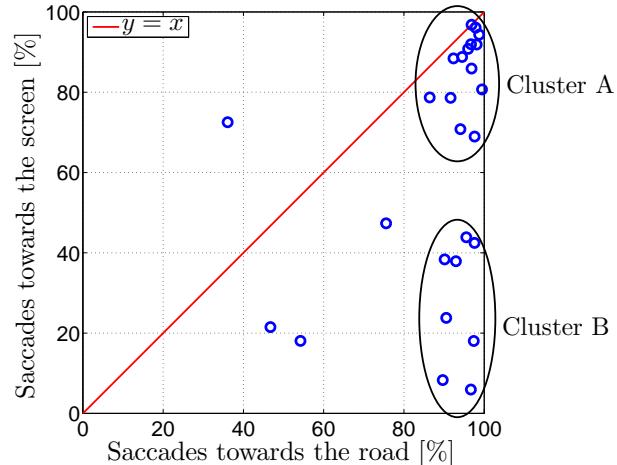


Figure 6.4.: Scatter plot: number of saccades accompanied by blinks with respect to their direction during the visuomotor task for all subjects. Ellipses show two clusters.

The red line in this plot refers to the $y = x$ line and is plotted for a better understanding. In fact, this line shows to what extent the number of saccades time-locked to blinks towards the road deviated from that of the equivalent saccades towards the screen. The ellipses show distinguishable clusters A and B. Cluster A contains 14 subjects, for whom, at least 65% of saccades induced the occurrence of blinks in both directions. On average, $95\% \pm 3.5$ of saccades towards the road and $85\% \pm 9.0$ of saccades towards the screen were accompanied by blinks for these subjects. Consequently, for this cluster, the occurrence of gaze shift-induced blinks was less direction dependent. On the other hand, for the subjects of cluster B, the direction of saccades seems to affect the occurrence of blinks to a greater extent. For this cluster, on average, $93\% \pm 3.4$ of saccades towards the road were accompanied by a blink, while only $27\% \pm 15.4$

of saccades towards the screen induced the occurrence of blinks. Therefore, in one direction (towards the road), the blink occurrence was more dominant. For subjects S15, S17 and S18, similar to cluster B, the saccades towards the road have induced more blinks than those towards the screen. However, the number of induced blinks for these subjects was less in comparison to that of the cluster B.

6.5. Blink rate analysis during the secondary and primary tasks

Here, it is studied whether performing secondary tasks affected the blink rate in comparison to the driving task. In order to analyze this, for all subjects, all detected blinks during each of these tasks were considered, independent of the fact of whether they were gaze shift-induced or not. Left plots of Figures 6.5(a) and 6.5(b) show the scatter plots of blink rates for driving versus visuomotor and auditory tasks with the correlation values of $\rho_p = 0.32$ ($p\text{-value} < 0.001$) and $\rho_p = 0.78$ ($p\text{-value} < 0.001$), respectively. ρ_p denotes the Pearson correlation coefficient which will be explained in Section 7.5. The closer the value of ρ_p to 1/0, the smaller/larger is the impact of the corresponding secondary task on the blink rate. According to the calculated ρ_p values and the scatter plots, it can be concluded that performing the visuomotor task affected the blink behavior to a larger extent in comparison to the auditory task ($0.32 < 0.78$). Moreover, based on the statistical test explained in Appendix D.4, the correlation between blink rate during the auditory task and that of the driving task is significantly different from the correlation between blink rate during the visuomotor task and that of the driving task.

Similar scatter plots for the subjects of clusters A and B are also shown in Figure 6.5, comparing the blink rates during both the visuomotor (middle and right plots in Figure 6.5(a)) and auditory tasks (middle and right plots in Figure 6.5(b)) versus the driving task. According to the correlation values, it seems that for cluster B with direction-dependent gaze shift-induced blinks, the blink rate was less affected during the visuomotor task in comparison to cluster A, because the ρ_p values are larger than those of the cluster A.

Table 6.1 shows the results of applying ANOVA for repeated measures to investigate the significant differences of the blink rate between tasks. In fact the difference between blink rate during driving task and each secondary task is studied. Overall, the mean of blink rate for visuomotor task is significantly different from that of the driving task ($p\text{-value} = 0.012 < \alpha = 0.05$). This is also the case for the auditory versus driving task ($p\text{-value} = 0.011 < \alpha$). Thus, the results state that performing a secondary task (either visuomotor or auditory) affected the blink rate. According to the values related to cluster B with direction-dependent gaze shift-induced blinks, we cannot show that the blink rate changed significantly during the secondary tasks ($p\text{-value} = 0.831 > \alpha$ and $p\text{-value} = 0.800 > \alpha$). However, for cluster A, it is the opposite which means performing both secondary tasks led to a significant variation of the blink rate in comparison to that of the primary drive task.

Table 6.1.: Values of ANOVA to assess the significant difference between means of blink rates for all tasks

	visuomotor vs. drive		auditory vs. drive	
	test statistic	p-value	test statistic	p-value
all subjects, all blocks	$F_{1,25} = 7.26$	0.012	$F_{1,25} = 7.48$	0.011
subjects of cluster A	$F_{1,13} = 14.35$	0.002	$F_{1,13} = 5.66$	0.033
subjects of cluster B	$F_{1,7} = 0.05$	0.831	$F_{1,7} = 0.07$	0.800

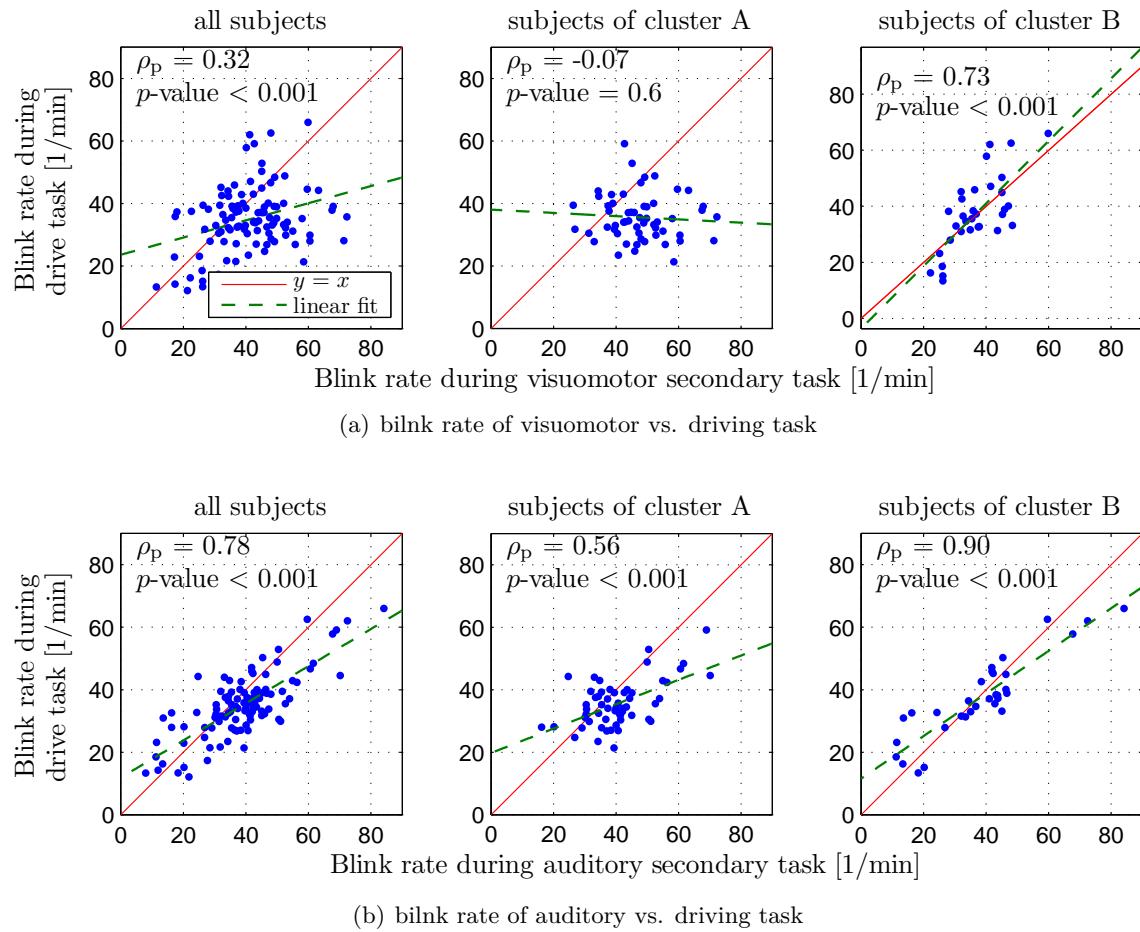


Figure 6.5.: Scatter plots of blink rate for visuomotor vs. driving and auditory vs. driving task. Pearson correlation coefficient (ρ_p) and the corresponding p -values are provided as well.

6.6. Impact of the visuomotor task on the blink behavior

In this section, it is studied how the blink behavior was affected during performing the visuomotor task. Figure 6.6 shows what percentage of blinks during the visuomotor task was gaze shift-induced on average over all blocks. In other words, after detecting all blinks in $V(n)$ signal, only those blinks time-locked to saccades of the $H(n)$ signal were considered. On average, $91\% \pm 8.7$ of blinks occurred simultaneously with a gaze shift. Consequently, during the visuomotor task the occurrence of spontaneous blinks was highly modulated by the gaze shift frequency. That means the subjects either did not blink or blinked simultaneously during the gaze shifts. To put it another way, the frequency of the gaze shifts during the visual distraction completely modulated the occurrence of blinks. This fact is shown in the left scatter plot of Figure 6.7 where the blink rate is plotted versus the saccade rate during the visuomotor task for all subjects. Middle and right plots of this figure show the same values for clusters A and B, respectively. It can be seen that the subjects of cluster A blinked as often as having gaze shifts. On the contrary, subjects of cluster B experienced a larger number of gaze shifts in comparison to the blinks.

Figures 6.8 and 6.9 show two representative examples of the EOG signals during the visuomotor and driving task. In Figure 6.8(a) (subject S8), during the visuomotor task, not only the blink frequency was completely modulated by the saccade frequency, but also the visual task led to the

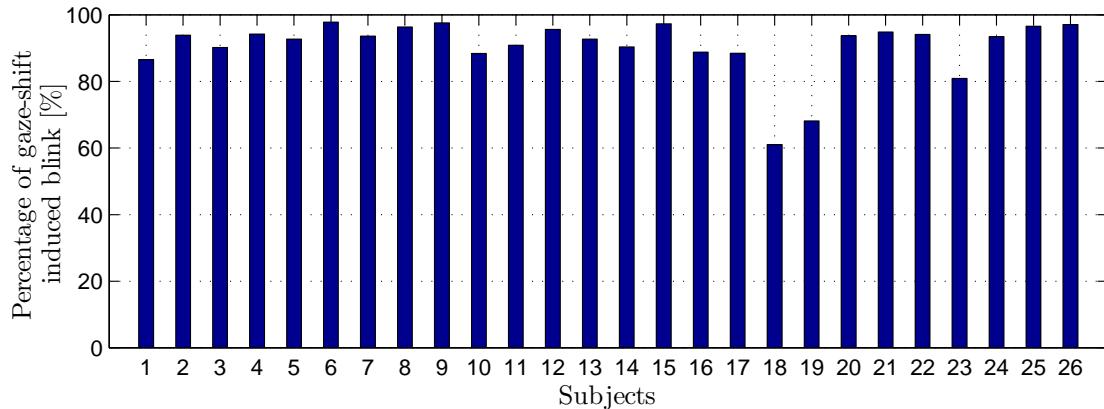


Figure 6.6.: Percentage of blinks time-locked to saccades for all subjects averaged over all blocks during the visuomotor task

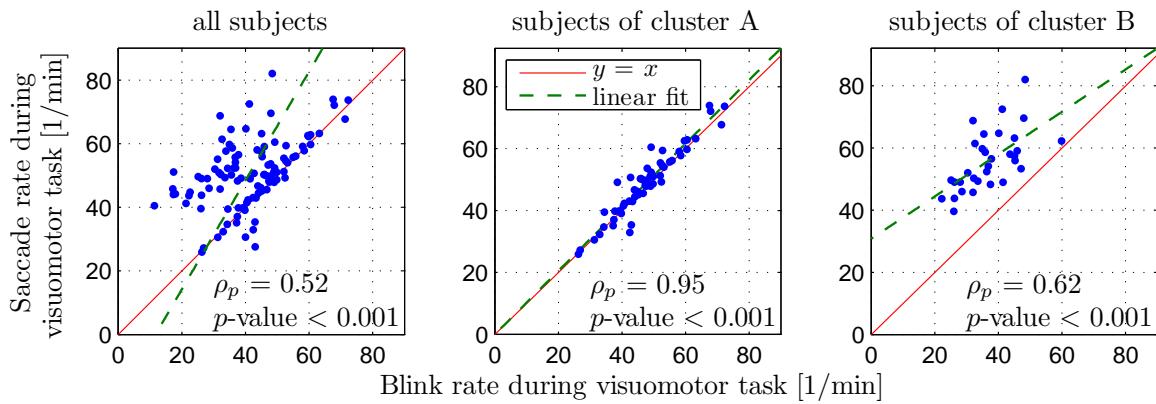


Figure 6.7.: Scatter plot: blink rate versus saccade rate during the visuomotor task

increase in the number of blinks in comparison to the driving task (Figure 6.8(b)). This subject belongs to the cluster A with direction-independent gaze shift-induced blinks. On the contrary, for subject S1 from cluster B with direction-dependent gaze shift-induced blinks, the number of blinks during the visuomotor task was decreased in comparison to that of the driving task (Figures 6.9(a) and 6.9(b)). Overall, it is clear that during the visuomotor task, blink frequency depended thoroughly on the saccade frequency.

6.7. Amount of gaze shift vs. the occurrence of gaze shift-induced blinks

This section explores whether the occurrence of gaze shift-induced blinks was correlated with the amount of the gaze shift. During the experiment in the driving simulator described in Section 4.3, the subjects experienced gaze shifts between various positions without any instruction. In order to show whether the occurrence of gaze shift-induced blinks was independent of the fact that the subjects were instructed to have gaze shift (as in the previous experiment) and whether this is positively/negatively correlated with the amount of gaze shift, all horizontal saccades during the drive were studied. This analysis is performed for the first 12 subjects of the corresponding experiment.

A single saccade, e.g. gaze shift of some degrees to the right, measured by EOG occurs with

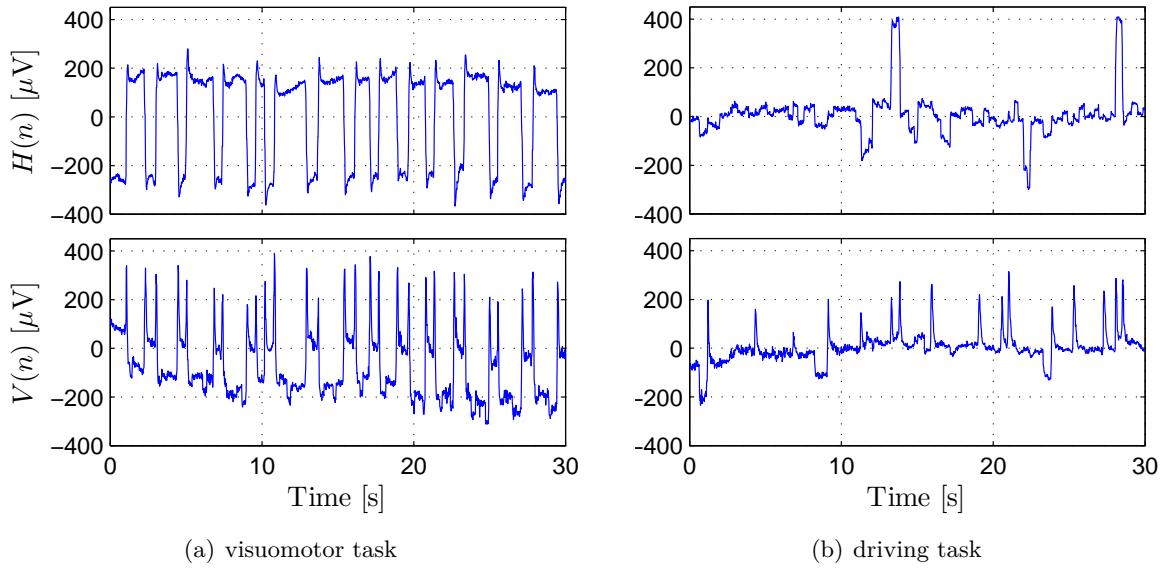


Figure 6.8.: EOG signals during the visuomotor and driving task for subject S8

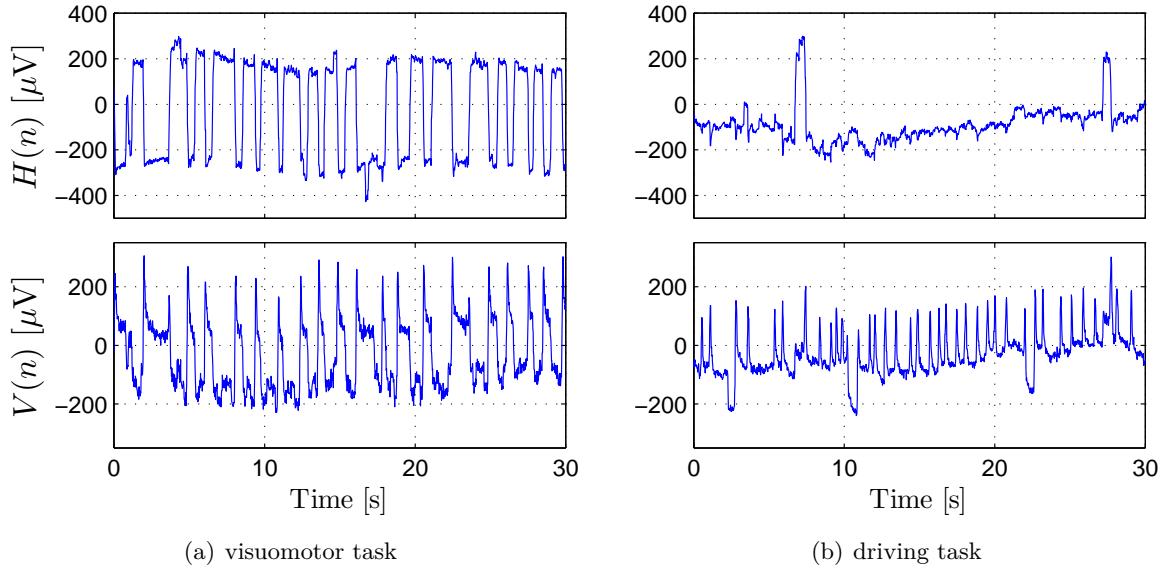


Figure 6.9.: EOG signals during the visuomotor and driving task for subject S1

different amp_{em} amplitudes in (5.4) from person to person. This measured value depends on the skin type, its cleanliness, etc. Therefore, it is possible that gaze shifts and saccadic eye movements of equal size are not detected for all subjects, because detected saccades with similar amplitudes do not necessarily refer to the equal amount of gaze shift. To overcome this, only horizontal saccades were considered whose amplitudes were equal or larger than that of a look at the speedometer¹. In fact, all glances at the speedometer out of the $V(n)$ signal were extracted as the minimum detectable gaze shift for each subject. Then one standard deviation of the mean

¹The amount of gaze shifts during moving focus to the speedometer depends also on the body size. Such gaze shifts are larger for a tall person with a large upper body part in comparison to a shorter person. Nevertheless, we suppose that the difference between body sizes is negligible among our subjects.

of them was used as the threshold for detecting saccades of the $H(n)$ signal. Thus, the threshold for the horizontal saccade detection was chosen individually based on the vertical saccades of $V(n)$, i.e. the amplitude of the glance at the speedometer. Figure 6.10 summarizes the explained algorithm.

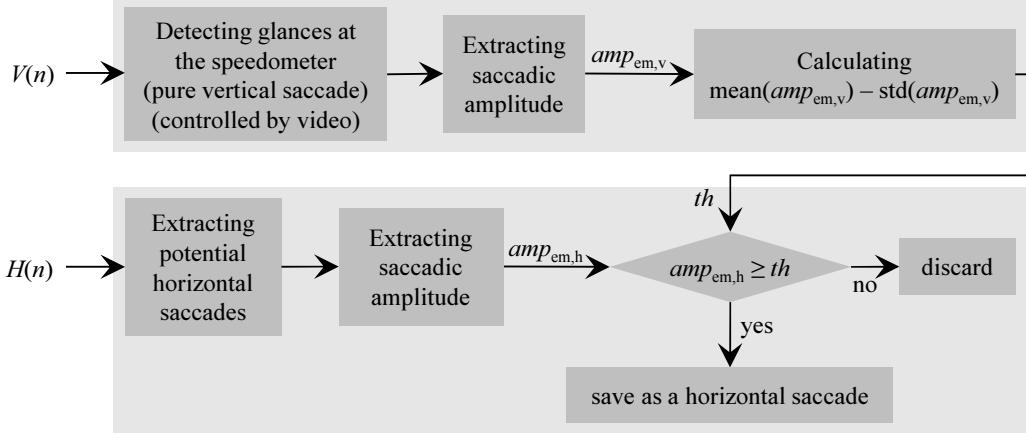


Figure 6.10.: Algorithm for determining the threshold of horizontal saccade detection

Figure 6.11 shows the histogram of the absolute amplitude of all horizontal saccades fulfilling the criterion shown in Figure 6.10 for one of the subjects. It can be seen that small-amplitude saccades (e.g. $amp_{em} < 200 \mu V$) occurred more often than the large-amplitude ones which refer to the gaze shifts during the glances at the side or rear-view mirror (e.g. $amp_{em} > 300 \mu V$). First, we need a threshold to distinguish between small and large-amplitude saccades. To this end, the k -means clustering algorithm ($k = 2$) (see Appendix C) was applied to divide the saccadic amplitude into two categories: small versus large-amplitude clusters. The threshold dividing the clusters is shown in Figure 6.11 in solid line. For most of the subjects, the number of saccades belonging to the small-amplitude cluster (N_s) was larger than that of the other cluster (N_l) which makes the further analysis and the comparison of clusters difficult. Therefore, as the second step, the numbers of occurrences of small and large-amplitude saccades were needed to be balanced. Consequently, out of N_s numbers of saccades of small-amplitude cluster, N_l of them were selected randomly. For subjects with $N_s < N_l$, the selection procedure was performed the other way around (see Figure 6.12). Afterwards, the occurrence of gaze shift-induced blinks with respect to the saccadic amplitude was studied. The selection of N_l out of N_s events (or vice versa) was repeated at least 100 times for each subject separately to ensure the independency of the result on the chosen saccades out of the small/large-amplitude cluster. Dark histograms of Figure 6.13 show the amplitude of the horizontal saccades of both clusters for the 100th iteration. The solid line also indicates the border between the clusters. After balancing the number of events in each cluster, it has been calculated how many of the saccades were accompanied by blinks. The histograms of the amplitude of these saccades are shown in light color in Figure 6.13.

The scatter plot in Figure 6.14 quantifies the result of the histograms of Figure 6.13. The numbers of small-amplitude saccades accompanied by blinks (x -axis) are plotted versus the same values for large-amplitude saccades (y -axis) in percent averaged over 100 repetitions of selecting N_s out of N_l (or vice versa). For all 12 subjects, the values are on the left side of the $y = x$ line. This implies that independent of the fact of whether the subjects were instructed to carry out a visuomotor secondary task in addition to the driving task, they automatically blinked more often during large-amplitude saccades in comparison to the small-amplitude ones. In other words, for gaze shifts with larger amounts, the probability of a simultaneous blink occurrence was higher.

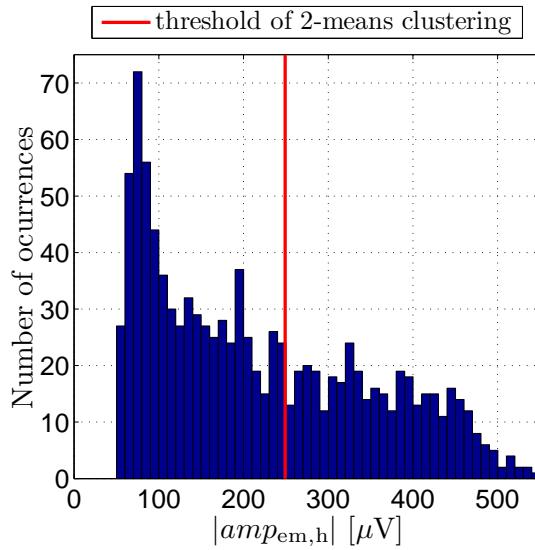


Figure 6.11.: Histogram: absolute amplitude of saccades out of $H(n)$ signal for subject S1

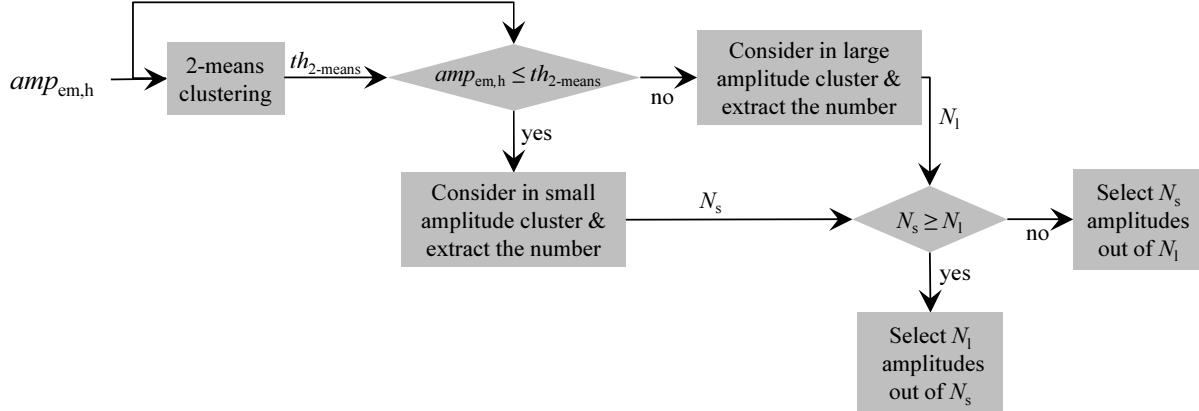


Figure 6.12.: The algorithm for balancing the number of small (N_s) and large-amplitude (N_l) saccades

To quantify whether the categorical data: “saccade amplitude: small/large” and “occurrence of gaze shift-induced blink: yes/no” were independent or not, the contingency table (cross tabulation) is studied. Table 6.2 shows the result for subject S1. By applying the Pearson’s chi-square test (see Appendix D.8), it can be shown whether the observed categorical data were related significantly to each other. Therefore, the H_0 hypothesis is formulated as “there was no relationship between two mentioned categories”. By considering the confidence level of 95% ($\alpha = 0.05$), for all subjects (except for subject S11), the p -values were always smaller than 0.001 which leads to the rejection of the H_0 hypothesis. Therefore, the amplitude of the gaze shift was responsible for inducing a blink, so that the larger the amount of the gaze shift, the more probable was the blink occurrence.

Summary

In this chapter, we discussed the occurrence of gaze shift-induced versus spontaneous blinks in real road and simulated driving. It was shown that during a visuomotor secondary task performed in a real road scenario comparable with the navigation system’s demand, gaze shifts

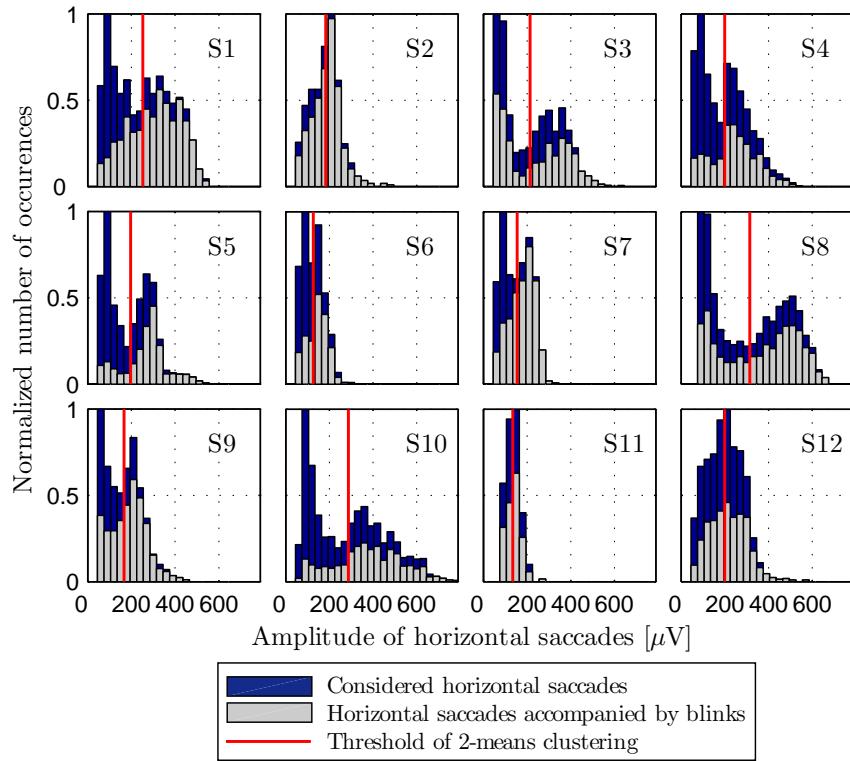


Figure 6.13.: Normalized histogram: amplitude of all horizontal saccades (dark bars) and those accompanied by blinks (light bars) for 12 subjects

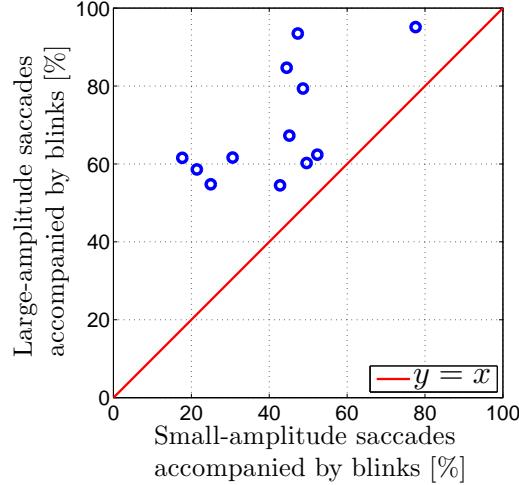


Figure 6.14.: Scatter plot: number of saccades in percent time-locked to the blinks with respect to their amplitude, i.e. small and large

induced the occurrence of blinks. This was also unrelated to the variable time-on-task. For 14 subjects out of 26, the fact was independent of the gaze shift direction. On the other hand, for 8 subjects, the occurrence of blinks was more probable during the gaze shifts towards the road. All in all, it seems that gaze shifts towards the road generally induce the occurrence of blinks to a larger extent.

Comparing the blink rate during the secondary tasks and the driving task, it was shown that the differences between the means were statistically significant stating that performing a secondary task

Table 6.2.: Contingency table: saccade amplitude versus occurrence of gaze shift-induced blinks, for subject S1, first selection procedure

		amplitude of saccades		
		events	small	large
occurrence of gaze shift-induced blink	yes	151	321	472
	no	228	58	286
	total	379	379	

(either visuomotor or auditory) affects the blink rate. Moreover, it was shown that the frequency of the gaze shifts during the visual distraction modulated the occurrence of blinks. On the other hand, results of the experiment in a driving simulator led to a positive correlation between the amount of gaze shifts and the occurrence of blinks in the case of no secondary task. This means that the larger the amount of the gaze shift, the higher is the probability of blink occurrence. Consequently, this study suggests those who solely consider blink rate as a drowsiness indicator to handle gaze shift-induced blinks differently to spontaneous ones, particularly if the driver is visually or cognitively distracted.

As mentioned in Sections 4.2 and 4.3, except for the experiment, which was studied in this chapter, all other experiences were designed such that no secondary tasks were allowed to be performed during the driving task. Therefore, the number of gaze-shift induced blinks was much smaller in comparison to the dominant number of the non-gaze shift induced blinks. However, in real life, similar scenarios to our experiment with visuomotor secondary task occur very often, e.g. while entering data into the navigation system. Therefore, the wrong interpretation of the changed blink frequency due to gaze shifts should be avoided by the drowsiness warning system. A possible solution for tackling this issue is to conditionally activate and deactivate the warning system. If the driver starts operating the center console elements, as an example, the warning system deactivates. Therefore, all induced blinks occurring due to gaze shifts between the center console and the road ahead are disregarded. As soon as, the center console is not being operated, the warning system reactivates and considers the detected blinks.

7. Extraction and evaluation of the eye movement features

In chapter 5, we explained how to detect relevant eye movements to driver drowsiness, i.e. eye blinks and saccades, out of the EOG signals. Based on the detected events, in Chapter 6, we investigated the relationship between the occurrence of the mentioned events. In this chapter, first, two approaches for aggregating attributes and features of the detected events are introduced and discussed. Our first aggregation approach is carried out with respect to collected KSS values. However, the second approach benefits from quick changes of drowsiness level in the course of time. Since features based on physiological measures are highly individual and vary from one subject to the next, we propose two baselining methods to deal with this issue. Afterwards, it is explained which features might be of interest for describing the driver's state of vigilance. Some of the features have not been defined consistently in previous studies. In this work, by considering all definitions, 19 well-defined features are introduced and extracted for each detected event. To this account, this work is one of the most comprehensive studies on eye blink features for in-vehicle applications and under real driving conditions. For well-known features, a detailed literature review is provided and our findings regarding feature's evolution due to drowsiness are compared with those of other studies. In addition, it is shown whether the extracted features change significantly shortly before the occurrence of the first safety-critical event in comparison to the beginning of the drive. To this end, the lane-keeping based and eye movement-based drowsiness detection methods are challenged. Afterwards, based on the correlation analysis, the linear and non-linear relationship between each feature and driver drowsiness is studied. Since the quality of driver observation cameras in detecting eye blinks will not be as high as that of the EOG, in the last section of this chapter, we investigate possible peak amplitude loss of features for the case that camera replaces EOG.

7.1. Preprocessing of eye movement features

Before introducing relevant features, here, we discuss feature aggregation approaches in order to reduce the number of samples of the feature space which is the space containing all extracted features. The reason is that drowsiness is a phenomenon which evolves over time. Therefore, it is unlikely to observe distinguishable characteristic differences between, e.g., two successive eye blinks. Consequently, for each feature a statistical measure, such as mean over a specific time interval, is calculated. This approach was also used in several studies as listed in Table 7.1. According to the table, the time interval for aggregating features varies from 1 s (Picot et al., 2010) to 6 min (Knippling and Wierwille, 1994) and the most frequently used statistical measure is the mean. In this work, we only used the mean as it seems to be more informative regarding the KSS values. Analysis of other statistical measures is provided in Appendix F.1.

Table 7.1.: Literature review of feature aggregation and the calculated statistic measure

Author	aggregation window	statistical measure
Knijppling and Wierwille (1994)	6 min	mean
Morris and Miller (1996)	task duration	mean
Dinges and Grace (1998)	1 min	—
Caffier et al. (2003)	—	mean
Johns (2003)	1 to 6 min	—
Svensson (2004)	20 s for EEG	—
Åkerstedt et al. (2005)	5 min	mean, standard deviation
Bergasa et al. (2006)	30 s	mean
Ingre et al. (2006)	5 min	mean, standard deviation
Johns et al. (2007)	1 min	mean, standard deviation
Papadelis et al. (2007)	1 min	mean
Damousis and Tzovaras (2008)	20 s	mean, sum, maximum
Schleicher et al. (2008)	5 min	mean, standard deviation, median
Hu and Zheng (2009)	20 s	mean
Friedrichs and Yang (2010a)	0.5 Hz	mean, median, EWMA, EWVAR
Rosario et al. (2010)	20 s	—
Picot et al. (2010)	1 s	mean
Sommer and Golz (2010)	3 min	mean, standard deviation
Wei and Lu (2012)	8 s	mean, maximum, minimum

7.1.1. KSS input-based feature aggregation

One of the methods applied for aggregating features is based on the KSS inputs. As mentioned in Section 2.2, the time interval around each KSS input is expected to be to the largest extent correlated with the true driver's state of vigilance. Therefore, as the first aggregation method, we calculated the mean value of each feature over the last 5 min before a KSS input, i.e. the time interval of $[t_{\text{KSS}} - 5 \text{ min}, t_{\text{KSS}}]$. t_{KSS} denotes the time instant at which the KSS value was collected. This is shown pictorially in Figure 7.1 for the 15-min time interval KSS data collection.

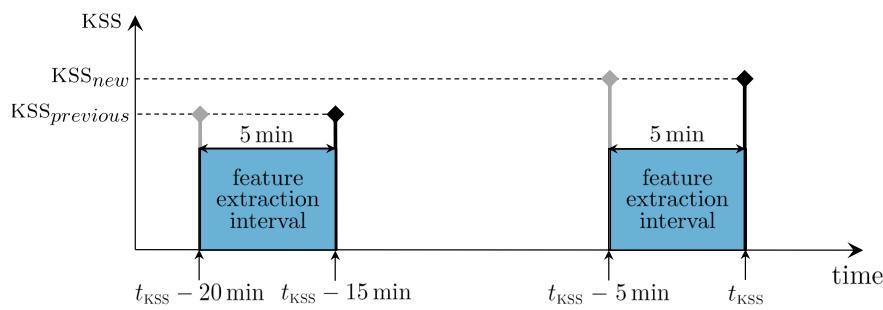


Figure 7.1.: KSS input-based feature aggregation method

Clearly, an advantage of this method is that only parts of the drive, for which the self-rating information is available, are analyzed further. However, a problem arises simultaneously by applying this approach which is ignoring valuable and expensive data outside the mentioned time interval, i.e. the intervals between KSS inputs. In fact, the resulting number of available

feature samples here depends on the number of times the KSS data is collected. We mentioned that KSS data cannot be collected very frequently for the sake of monotonicity of the driving condition. As a result, for each hour of driving, only a small number of KSS values will be available, namely only 4 values by collecting KSS in 15-min intervals. Hence, this method is not suitable, if small numbers of KSS values are available. Overall, this method might lead to a set of features which are not very informative due to lack of observation samples. Another drawback of this method is its reliance on the collected KSS values.

Schmidt et al. (2011) also used 5 min before a KSS input for evaluating the short-term effect of verbal assessment of driver vigilance which is in agreement with our approach.

In this work, in total 391 KSS values were recorded and correspondingly 391 5-min windows were available for extracting features. Figure 7.2(a) shows the distribution of the relative frequency for each KSS value in percent, i.e. the number of occurrences for each KSS value with respect to the total number of available KSS values. The numbers on the top of the bars denote the number of counts.

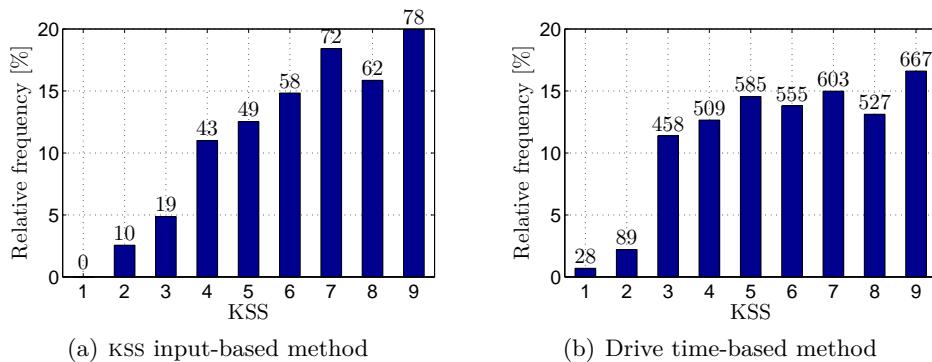


Figure 7.2.: Relative frequency of KSS values for two feature aggregation methods

7.1.2. Drive time-based feature aggregation

Another approach for aggregating features is considering a 1-min interval independent of the KSS inputs similar to Dinges and Grace (1998), Johns et al. (2007) and Papadelis et al. (2007). This method has several advantages. First of all, all parts of the drive are analyzed, without discarding any data. Moreover, on the contrary to the previous approach, for one hour of driving, 60 feature values are extracted. Consequently, by analysis of a larger amount of data, the resulting set of features is more informative.

In this work, the average 1-min intervals with no overlap was considered beginning from the first minute of driving and excluding noisy parts of the EOG data. Figure 7.3 pictorially shows this feature aggregation approach.

Non-overlapping windows make the extracted feature less statistically dependent in time¹. Two major drawbacks of this method are the underlying assumptions about the KSS values. First, similar to the previous approach, this method also strongly relies on the preciseness of the subjective measure. Secondly, in order to have a corresponding KSS value for each feature extraction interval, we have assumed that a KSS value remains unchanged between two successive

¹In Chapter 6, we supposed that the occurrence of an eye movement, e.g. a blink, is independent of that of the other blink in time. Other features of adjacent blinks, however, might be correlated with each other.

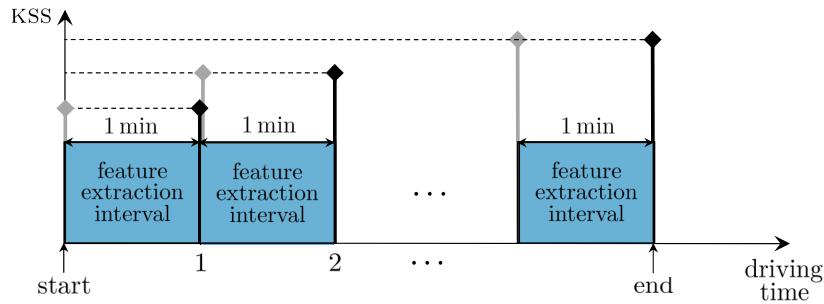


Figure 7.3.: Drive time-based feature aggregation method

KSS inputs (see Figure 7.3). On this account, a preceding KSS value was used up to the next input. In other words, we have assigned a specific self-rating value to those parts of the drive for which the subject did not rate his level of vigilance.

Comparison of Figures 7.3 and 7.1 reveals that each of the introduced approaches assigns a different KSS value to the time interval of $[t_{\text{KSS}} - 5 \text{ min}, t_{\text{KSS}}]$. The first approach uses the following KSS value, while the second approach holds the preceding value up to the time instant of the collection of a new KSS value, namely t_{KSS} .

Similar to Figure 7.2(a), Figure 7.2(b) shows the relative frequency of KSS values given the drive time-based feature aggregation method. In total, 4021 samples are available for each extracted feature which is equal to 4021 min of driving (about 70 h).

7.1.3. Feature baselining

Since biological measures like blink characteristics are highly individual and vary from one subject to the next (Dong et al., 2011), a baselining method is applied to suppress irrelevant characteristics for further analysis. Assuming that all subjects were awake at the beginning of the drive, which is not always the case in real life though, the average over the first t_{baseline} min of each feature (e.g. $t_{\text{baseline}} = 5$ or 10 min) is used as the normalization factor for the rest of that drive, namely

$$\text{normalization factor} = \text{mean}\left(\text{feature}\left([t = 0, \dots, t = t_{\text{baseline}}]\right)\right). \quad (7.1)$$

Therefore, for each sample x_i of a feature we have

$$x_{i,\text{baselined}} = \frac{\{x_i \mid x_i \in \text{feature}\}}{\text{normalization factor}}. \quad (7.2)$$

In addition to (7.2), the standard score defined as

$$x_{i,z\text{-score}} = \frac{\{x_i \mid x_i \in \text{feature}\} - \mu}{\sigma} \quad (7.3)$$

can be explored as well, where μ and σ refer to the mean and standard deviation of the corresponding feature calculated over the entire samples. Simon (2013) suggested considering μ_{awake} and σ_{awake} which are calculated with respect to the awake phase of the drive, e.g. the first 20 min of the drive.

Among these methods, we obtained the highest correlation values between each of the baselined features and KSS values for the first approach with $t_{\text{baseline}} = 10$ min. Figures 7.4(a) and 7.4(b)

show one of the extracted blink features (*MOV* will be defined in next section) versus KSS values before and after baselining, respectively. Obviously, the growing trend of this feature from KSS 1 to 3 in Figure 7.4(a) is the result of individual differences in the values of this feature and is consequently drowsiness-irrelevant. After baselining, the misleading trend is filtered out. Hence, this preprocessing step of the extracted features has a crucial contribution towards improved results for drowsiness detection, especially in the next step which is classification.

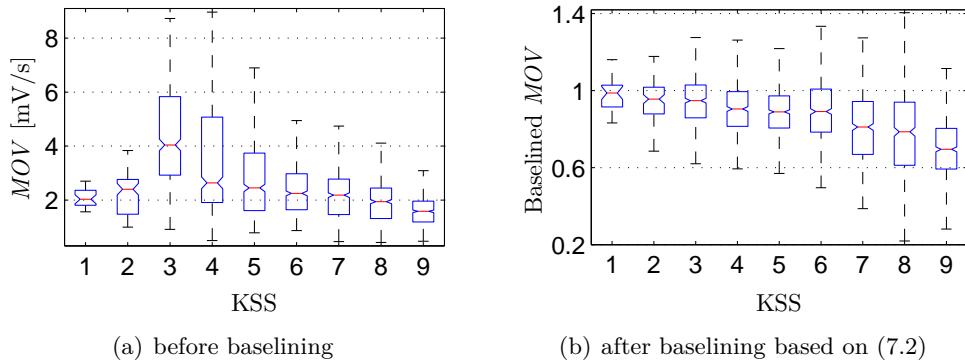


Figure 7.4.: *MOV* feature before and after baselining

7.2. Eye blink features

This section introduces the extracted features in this work and discusses their association with the driver drowsiness based on the KSS values. Additionally, for each feature a comprehensive literature review is provided to give insight into the features and to compare the findings of this study with those of the previous studies. Some of these features were called similarly in previous studies despite their different definitions. Here, we take all definitions into consideration for the sake of completeness.

Table 7.2 lists 19 features extracted in this study for detected events shown in Figure 7.5. In the following, these features are defined and it is explained how to calculate them. All plots refer to drive time-based features.

- A: Blink amplitude is defined as the minimum of the rise amplitude A_1 and the fall amplitude A_2 , i.e. $A = \min(A_1, A_2)$, where

$$A_1 = V(\text{middle}) - V(\text{start}) \quad (7.4)$$

$$A_2 = |V(\text{end}) - V(\text{middle})|. \quad (7.5)$$

Start, middle and end points of a blink are shown in Figure 7.5. This feature was also defined earlier in Chapter 5 in (5.3) for event detection. There, it was clarified why the minimum value of A_1 and A_2 was used.

The study by Morris and Miller (1996) on sleep-deprived pilots showed a significant drop in blink amplitude along with increase of pilot errors. They related this phenomenon to “*a lower starting position of the eyelid*” which leads to a reduced distance between eyelids to be traveled. This feature was also acknowledged as the best single predictor of the performance error in their study. Jammes et al. (2008), on the contrary, reported increased amplitude of very long blinks based on visual inspection. Svensson (2004), who

Table 7.2.: Extracted blink features

No.	Feature	Description
1	A	Amplitude
2	E	Energy
3	MCV	Maximum closing velocity
4	MOV	Maximum opening velocity
5	A/MCV	Ratio between A and MCV
6	A/MOV	Ratio between A and MOV
7	ACV	Average closing velocity
8	AOV	Average opening velocity
9	F	Frequency
10	T	Duration
11	T_c	Closing duration
12	T_o	Opening duration
13	$T_{cl,1}$	Closed duration (first definition)
14	$T_{cl,2}$	Closed duration (second definition)
15	T_{ro}	Delay of reopening
16	$PERCLOS$	Ratio between T and $T_{cl,2}$
17, 18 ,19	T_x	Duration from $x\%$ of the rise amplitude to $x\%$ of the fall amplitude ($x = 50, 80, 90$)

studied the relationship between A and the blink velocity, assumed that this feature evolves linearly over drowsiness. The results showed the drop of this features due to drowsiness.

Figures F.2 and F.3 show the boxplots of normalized A , namely $\frac{A}{\max(A)}$, versus KSS values for all subjects, except for the subject S2. For this subject all features are shown together in Figure F.40. Moreover, Figure 7.6 shows all baselined drive time-based features and their overall trend versus KSS values regarding all subjects. According to these figures, in our experiments, A has decreased for most of the subjects, as KSS has increased (e.g. subjects S1, S5, S13, S18 and S21). Nevertheless, for some subjects such as S4, S6 and S10, an increasing trend of A is observable. This means that drowsiness led to blinks with larger amplitude for these subjects as also reported by Jammes et al. (2008). The reason might be that the subjects tried to keep themselves awake and to fight against drowsiness by opening their eyelids to a larger degree. This results in larger blink amplitudes. Interestingly, subjects with neither an increasing nor a decreasing trend in the evolution of A (S29, S41, S40 and S43) have rated themselves mostly awake which is highly plausible.

- E : Energy of a blink is defined as

$$E = \sum_{n=\text{start}}^{\text{end}} (V(n) - V(\text{start}))^2. \quad (7.6)$$

Clearly, the energy of a blink strongly depends on the recorded $V(n)$ values. Therefore, the energy of two blinks with completely similar forms might differ depending on the drift existing in the $V(n)$ signal. In Chapter 5, approaches were introduced to remove the drift of the EOG signals. Despite applying the drift removal step to all EOG signals before event detection, subtracting the amplitude of the start value, namely $V(\text{start})$, from all other values counteract the negative effect of the drift in the calculation of E .

This feature was also used by Friedrichs and Yang (2010a). Picot et al. (2010) calculated E only for the closing phase. In addition to the mentioned definition of E , another approach is to calculate energy within different frequency bands of the EOG similar to the analysis

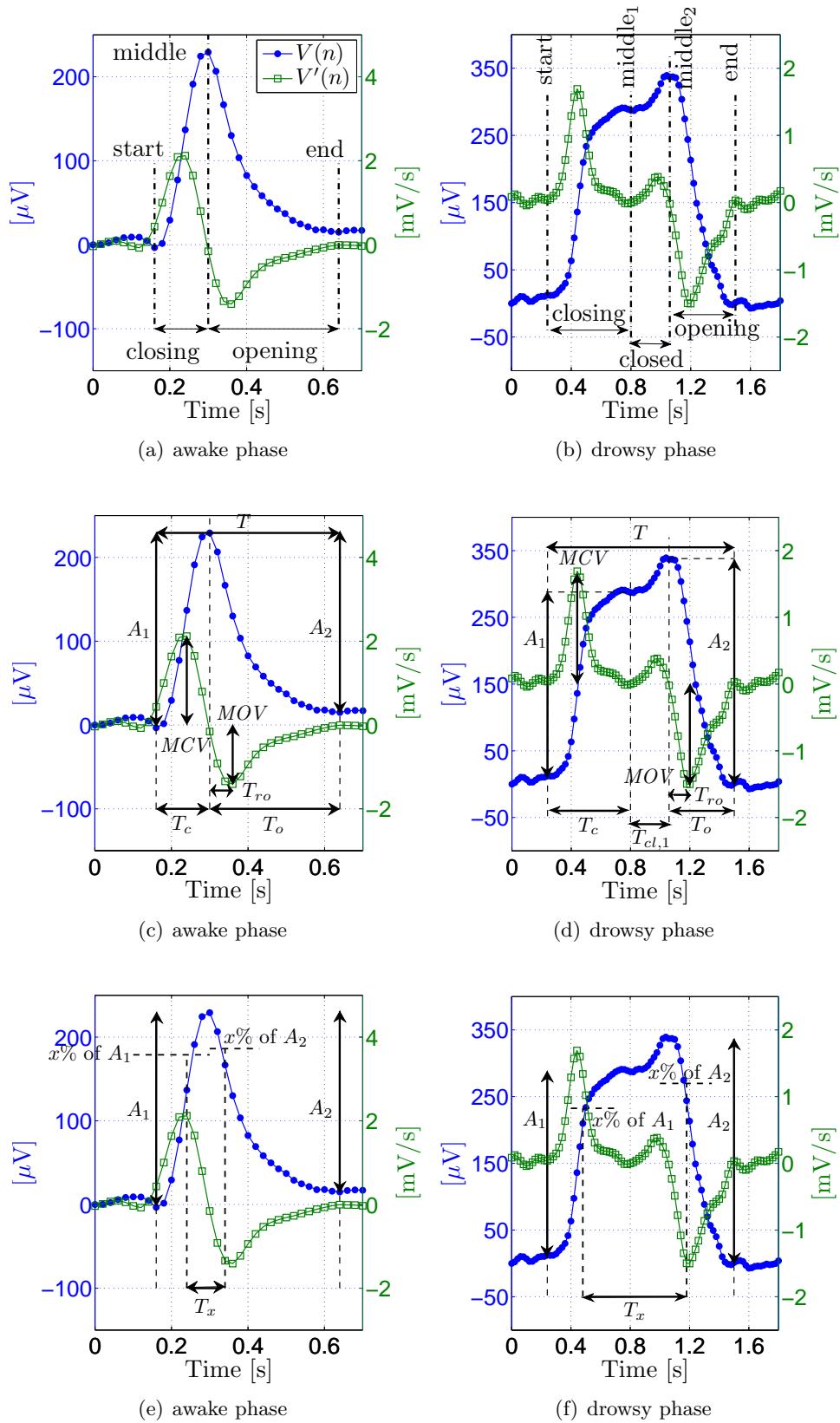


Figure 7.5.: $V(n)$ and its derivative $V'(n)$ representing eye blinks in awake and drowsy phases with the corresponding features

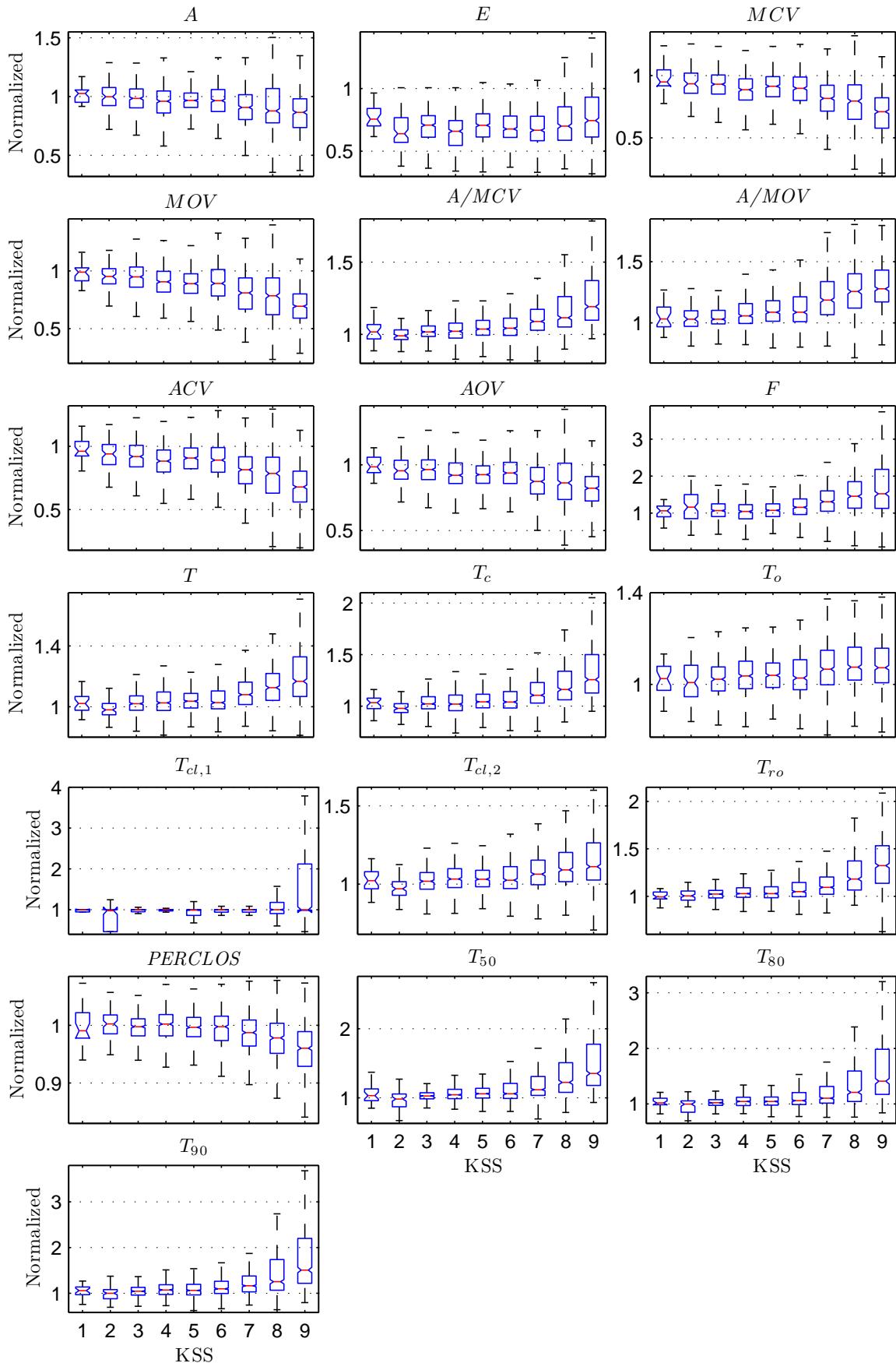


Figure 7.6.: Boxplot of normalized drive time-based features combined for all subjects versus KSS values

of EEG. This was suggested by Wei and Lu (2012) who asserted that the ratio between energy of low and high frequency bands of EOG is more important for assessing the driver vigilance than analyzing each frequency band separately. The reason is that unlike the high frequency eyelid movements, the low frequency movements occur more often during the drowsy phases.

Figures F.4 and F.5 show the relationship between this feature and KSS values. Similar to A, the overall trend of E for each individual subject is decreasing as drowsiness increases. However, the overall trend in Figure 7.6 regarding all subjects does not show any specific trend. Nevertheless, the interquartile range and the difference between whiskers (see Appendix B) increase along with the increase of KSS values.

- MCV , MOV : Maximum closing/opening velocity is the maximum value of $|V'(n)|$ during closing and opening phases as shown in Figures 7.5(c) and 7.5(d). In this work, we extracted these features out of the $V'(n)$ signal calculated by the Savitzky-Golay filter with empirically selected parameters of 5 for the polynomial order and 13 for the frame size.

According to Hargutt (2003) and Holmqvist et al. (2011), the closing phase of the eyes occurs much faster than the opening phase. In our data, this was also the case as shown in Figure 7.7 which compares the distribution of the MCV with that of the MOV . These features are shown in Figures F.6, F.7, F.8 and F.9 versus KSS values. It can be seen that the overall trends of the velocities are decreasing within increasing drowsiness, e.g. for subjects S4, S5, S12, S13 and S15.

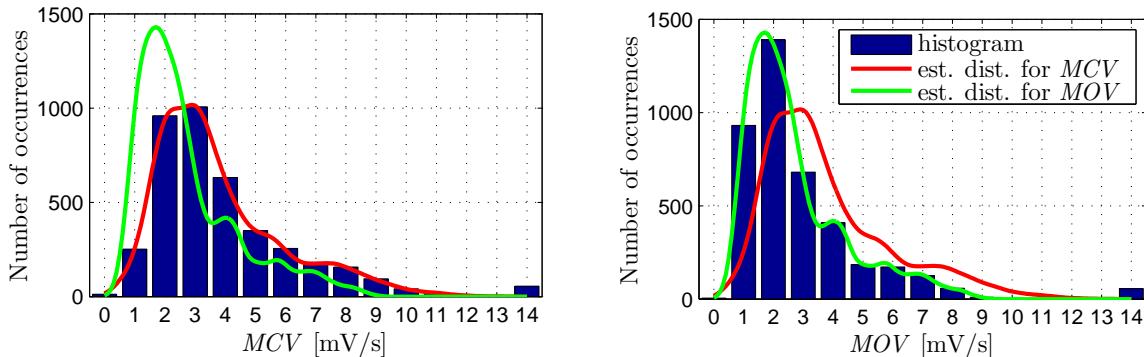


Figure 7.7.: Histogram of MCV and MOV with the estimated (est.) distribution (dist.) curves

- A/MCV , A/MOV : Johns (2003) introduced the amplitude-velocity ratio calculated by $\frac{A}{MCV}$ as a feature positively correlated with drowsiness. The ratio for MOV was defined similarly to this feature, because Johns and Tucker (2005) showed that it also increases due to drowsiness. Johns et al. (2007) asserted that these two features are almost in the same range for different subjects and, therefore, it is not necessary to “*adjust them for individuals*”. Damousis and Tzovaras (2008) used the A/MCV feature as an input to their fuzzy system for drowsiness detection.

Figures F.10, F.11, F.12 and F.13 show these features versus KSS values. For drivers, who felt drowsy at the end of the drive ($KSS \geq 7$), different trends are found. For subjects S4, S6 and S10, as an example, these features increased, while for subjects S23 and S16 the value of the features were almost stable during the drives. It can be seen that the range of both features is almost the same considering all subjects which agrees with the statement of Johns et al. (2007).

- *ACV, AOV*: The average closing and opening velocities were calculated by $\frac{A_1}{T_c}$ and $\frac{A_2}{T_o}$, respectively. T_c and T_o refer to the closing and opening duration. For blinks of the drowsy phase, the equivalent middle points were used, namely $middle_1$ and $middle_2$ as shown in Figure 7.5(b). Figure 7.8 shows that in addition to the *MCV*, the average closing velocity of the eyelids are also, in general, larger than their average opening velocity.

Figures F.14, F.17, F.16 and F.15 show these two features versus KSS values. Similar to the trends of *MCV* and *MOV*, these two features also, overall, have decreasing values versus KSS as also mentioned by Thorslund (2003). It should be mentioned that Thorslund (2003) defined $\frac{A}{T}$ as the half of the blink velocity which might not always be valid, because it takes the duration of the closed phase of the blink into consideration as well. Damousis and Tzovaras (2008) used the inverse of these two features together with other features in a fuzzy logic-based system.

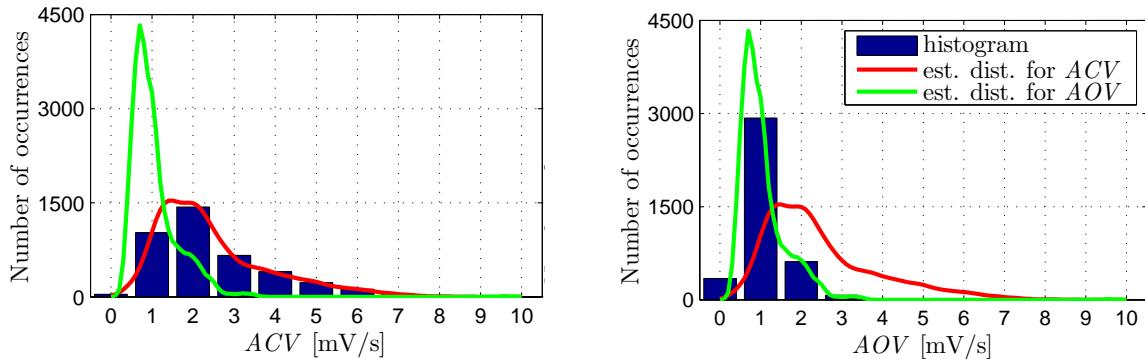


Figure 7.8.: Histogram of *ACV* and *AOV* with estimated (est.) distribution (dist.) curves. The outliers (values > 10 mV/s) are not shown.

- *F*: Blink frequency is defined as the number of blinks per minute (or another pre-defined interval) which increases in earlier phase of drowsiness and decreases as drowsiness increases, i.e. similar to the shape of an inverse U (Platho et al., 2013). Here, we calculated *F* within 1-min intervals. This feature is also called *blink rate* and since its value only depends on the detection of blinks (not the corresponding start and end points), it is referred to as the easiest measurable blink feature (Holmqvist et al., 2011).

On average, human blink rates range from 15 to 20 per min regarding spontaneous blinks (Records, 1979). This range decreases to 3 to 7 blinks per min during reading (Holmqvist et al., 2011). As mentioned in Chapter 6, this feature varies differently, while performing a secondary visuomotor task. According to Records (1979), the blink rate decreases as visual attention increases. The reason is the prevention of information loss during eye closure moments. Since in our drowsiness-related experiments, it was not allowed to perform any secondary tasks, we suppose that none of the occurred blinks were task-related.

This feature is also highly dependent on the humidity of the vehicle interior. With the air conditioning running or in a very dry condition, blink frequency might be different (Friedrichs and Yang, 2010a). Friedrichs and Yang (2010a) also reported a large between-subject variation of this feature. Moreover, Johns et al. (2007) believes that this feature is “*too task dependent and too subject-specific*” to be considered for drowsiness detection. The study by Hargutt (2003) also showed that this features is more correlated with information processing and how demanding a task is (e.g. task duration and time-on-task). Regarding these findings, some researchers have serious doubts about the usefulness of blinks and their

corresponding features as a drowsiness indicator due to their variation and external stimuli such as “*road lighting, oncoming headlights, the air temperature and state of the ventilation system*” (according to Horne and Reyner (1996) cited by Liu et al., 2009). In a study by Papadelis et al. (2007), also no significant difference of this feature between the awake (the first 15 min) and the drowsy (the last 15 min) phases was found. Moreover, driving time showed no significant interaction with F . Interestingly, right before a driving error, however, F increased significantly in comparison to the beginning of the measurement. Similarly, in a study on 60 subjects, blinks were measured at the beginning and end of a working day (Caffier et al., 2003). Although no significant difference between alert and drowsy phases was found, a slightly decreased blink rate during drowsy phase was reported. However, these results contradict the findings of Morris and Miller (1996) about the efficiency of F in a moving base-flight simulator with 10 sleep-deprived pilots.

Apart from the mentioned studies, which analyzed the interaction of F with drowsiness separately, other studies showed that the combination of this feature with other features leads to a satisfactory estimation of driver’s state of vigilance (Bergasa et al., 2006; Suzuki et al., 2006; Friedrichs and Yang, 2010a).

Figures F.18 and F.19 show the boxplot of F versus KSS values. By visual inspection of the figures, it can be inferred that blink frequency has increased due to drowsiness for 8 subjects such as S3, S13 and S15. Nevertheless, it suffers from large individual differences. For subjects S1 and S18, it has dropped, while for subjects S20, S21 and S22 it has not changed significantly. The inverse U-form shape mentioned in Platho et al. (2013) is also not found in our data set.

- T : Blink duration is the time interval between the start and the end point of a blink. Clearly, the value of this feature depends on the considered start and end points of the blink which are not defined consistently among researchers. Our definition of start and end points of a blink is in agreement with the points chosen by Hu and Zheng (2009) and Wei and Lu (2012). On the contrary, Schleicher et al. (2008), as an example, considered the point at which the maximum velocity of the eye opening phase occurs as the end point. Other works also used our 17th feature, namely T_{50} , as the blink duration (Morris and Miller, 1996; Thorslund, 2003; Johns, 2003; Åkerstedt et al., 2005; Ingre et al., 2006; Damousis and Tzovaras, 2008; Picot et al., 2010; Friedrichs and Yang, 2010a).

Caffier et al. (2003) assumed 50 ms as the shortest possible blink duration and blinks with $300 \text{ ms} < T < 500 \text{ ms}$ as the “*long-closure blink duration*”. In their study with 60 subjects, they found a significant increase of the blink duration before ($T = 192 \pm 39 \text{ ms}$) and after ($T = 316 \pm 62 \text{ ms}$) a usual working day using ANOVA. Moreover, T was significantly correlated with subjective self rating (Pearson correlation coefficient $\rho_p = -0.358$).

Figures F.20 and F.21 show the relationship between feature T and KSS values. The overall observed trend of T is an increasing one for drowsy subjects. For some subjects, however, the trend of T is neutral, such as subjects S16, S17, S19, S20 and S21, although they rated themselves as drowsy. Surprisingly, for subject S8 the trend is negative meaning that the blinks were shorter as the subject felt drowsy. For awake subjects, the feature remained almost stable.

- T_c, T_o : Duration of the closing and opening phases as shown in Figure 7.5. The distributions of these two features are shown in Figure 7.9. In our experiments, on average $T_c = 120 \pm 30 \text{ ms}$, while $T_o = 247 \pm 40 \text{ ms}$, i.e. closing time is about two times shorter than the opening time. In the study by Caffier et al. (2003), T_o was called the *reopening time*

and was shown to be highly correlated with T ($\rho_p = 0.939$), contrary to T_c ($\rho_p = 0.310$). This is clearly due to the fact that T_o covers a larger part of T in comparison to T_c . In their study, they found a 10% and 30% increase of T_c and T_o , respectively, before and after a working day. However, these findings were not significant for all of their subjects. Based on the weak correlation of T_c with the subjective self-rating, they had doubts on the performance of T_c as a single measure for drowsiness detection. Moreover, they used $T_c < 150$ ms as an additional criterion for blink detection. Damousis and Tzovaras (2008) aggregated these features over 20-s time intervals in a study based on a fuzzy system.

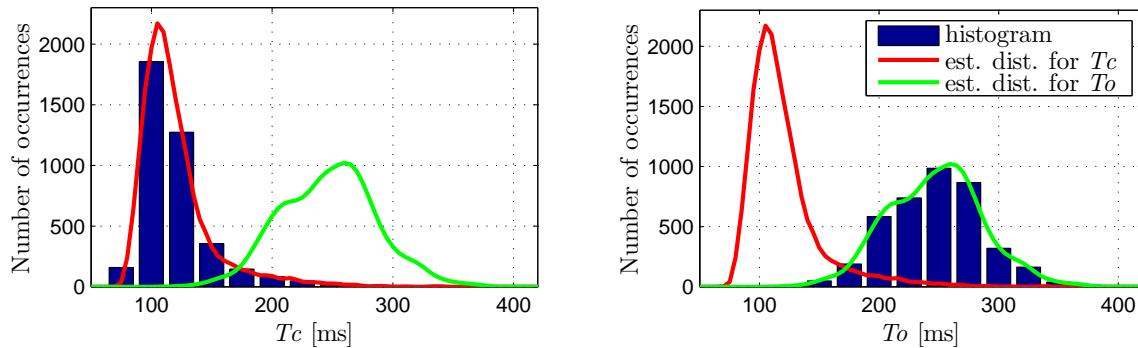


Figure 7.9.: Histogram of T_c and T_o with the estimated (est.) distribution (dist.) curves

Figures F.22, F.23, F.24 and F.25 show T_c and T_o versus KSS values. The plots indicate that overall, T_c increased while T_o varied differently during drowsiness. This agrees with the overall boxplots shown in Figure 7.6. For S11 as an example (Figures F.22 and F.24), despite the increasing trend of T_c , T_o has evolved constantly.

By assuming that drowsiness affects both features linearly, we fitted two lines to the scatter plot of these baselined features versus KSS values as shown in Figure 7.10. They show that from KSS = 1 to KSS = 9, T_c increased up to 1.5 times, while T_o scaled to 1.1 times.

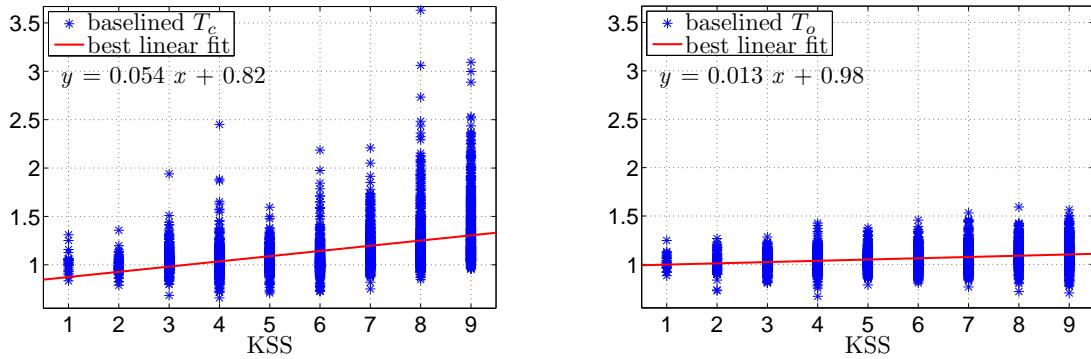


Figure 7.10.: Best linear fit to all baselined feature values of T_c and T_o

- $T_{cl,1}$, $T_{cl,2}$: Closed duration is the time interval during which the eyes are closed. Two definitions are used here. $T_{cl,1}$ is the time interval between the end of the closing phase and the start of the opening phase. This definition is very similar to the plateau duration defined by Friedrichs and Yang (2010a). Schleicher et al. (2008) called this feature “*delay of reopening*”. The other definition is taken from Wei and Lu (2012) as $T_{cl,2} = T_{cl,1} + T_o$. Caffier et al. (2003), however, introduced our 19th feature, namely T_{90} , as the closed time.

Figures F.26, F.27, F.28 and F.29 show both of the introduced features versus kss values. It can be seen that during the awake phase, $T_{cl,1}$ did not change at all and remained the same as one measured sample at 50 Hz, i.e. 20 ms. At $kss \geq 7$, larger values of $T_{cl,1}$ were measured e.g. for S10, S15 and S16.

Figure 7.11 shows how each element of the blink duration evolved due to drowsiness regarding all of our subjects. The top and bottom plots show the boxplot and the mean values of T_c , $T_{cl,1}$ and T_o , respectively.

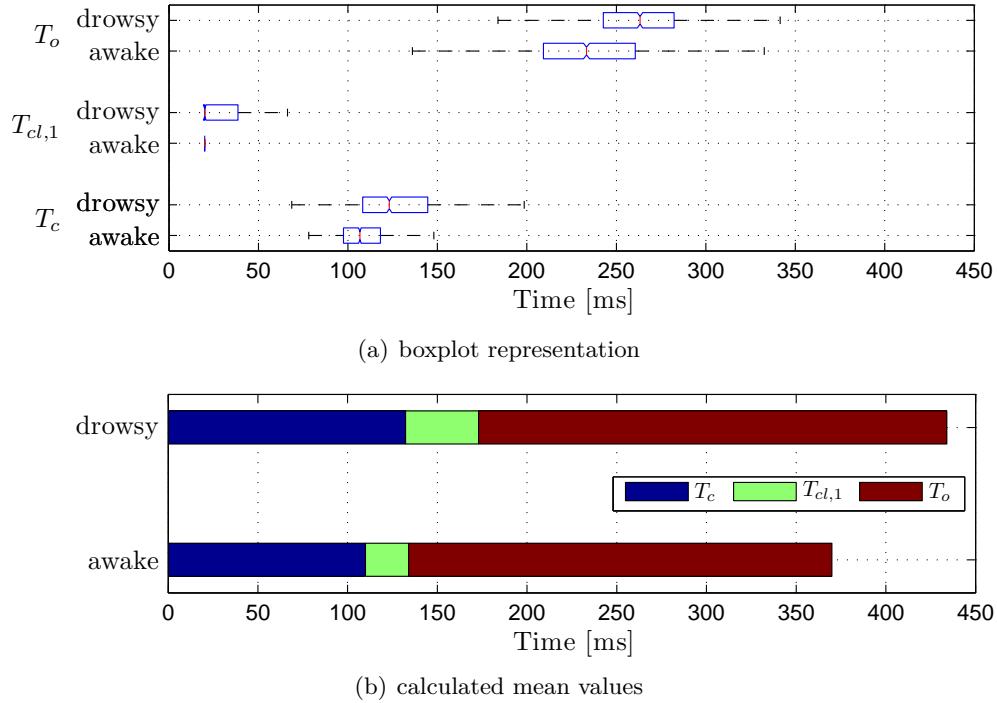


Figure 7.11.: Comparison between T_c , $T_{cl,1}$ and T_o during the awake and drowsy phases of the drive for all subjects

- T_{ro} : T_{ro} denotes the delay of reopening of the eye and is defined as the time interval between the start of the opening phase and the point of maximum velocity during this phase (Damousis and Tzovaras, 2008; Hu and Zheng, 2009) as shown in Figures 7.5(c) and 7.5(d). As also mentioned before, Schleicher et al. (2008) used the stopping point of this feature as the end point of a detected blink.

Figures F.30, F.31 and 7.6 show that this feature increased during the drowsy phases of the drive. The best linear fit to the values of this feature shown in Figure 7.12 indicates an increase of 1.6 times due to drowsiness.

- **PERCLOS**: As one of the most popular features for drowsiness detection, it was first introduced by Knipling and Wierwille (1994) and refers to the proportion of time for which the eyes are more than 80% closed (PERcentage of eye CLOSURE). This feature is originally a camera-related feature and is usually calculated accumulatively over a pre-defined interval between 1 and 5 min (Sommer and Golz, 2010).

In this work, on the contrary to eye tracking cameras, this feature is calculated by duration-

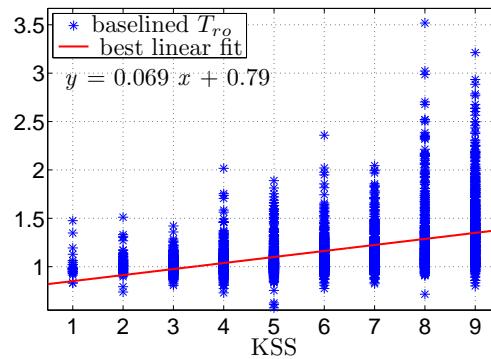


Figure 7.12.: Best linear fit to all baselined feature values of T_{ro}

based features of EOG (instead of amplitude-based ones) (Wei and Lu, 2012), i.e.

$$PERCLOS = \frac{T_{cl,2}}{T}. \quad (7.7)$$

Another definition of *PERCLOS*, which is mostly used given that the eye movement data is collected by a camera (Li et al., 2011b; Horak, 2011), is as follows

$$PERCLOS = \frac{T_{80}}{T_{20}} \times 100. \quad (7.8)$$

$T_x, x \in \{20, 80\}$ is our last feature and refers to the duration between following points: from $x\%$ of the absolute rise amplitude $|A_1|$ to $x\%$ of the absolute fall amplitude $|A_2|$.

In the following, the studies on the camera-based *PERCLOS* are reviewed. Li et al. (2011b) reported *PERCLOS* feature as the best indicator of drowsiness. Sigari (2009), who also studied a camera-based drowsiness detection method, suggested comparison of this feature with a threshold for drowsiness detection as its increase is expected due to drowsiness. In agreement with the previous study, Bergasa et al. (2006) and He et al. (2010) also reported an increase of *PERCLOS* during drowsiness in both driving simulator and real night drives. “*The PERCLOS measure indicated accumulative eye closure duration over time, excluding the time spent on normal eye blinks*”, is the definition used by Bergasa et al. (2006, page 68) and (Rosario et al., 2010, page 283). Rosario et al. (2010) studied the combination of *PERCLOS* and EEG (power of θ -waves) aggregated over 20-s intervals. Based on the high correlation found between them during drowsy driving, they suggested *PERCLOS* as a non-intrusive reliable ground truth. Bergasa et al. (2006), however, calculated the moving average of *PERCLOS* within 30-s windows. In their study, although *PERCLOS* was the best single feature among other studied features, by combining it with other features and applying a fuzzy classifier to them even better results were achieved than considering it as a single feature.

Picot et al. (2009) showed that *PERCLOS* can be measured with a 200-Hz video camera with the same accuracy as with the EOG. In a later study, Picot et al. (2010) found *PERCLOS* as the best feature for drowsiness detection based on an experiment in a driving simulator with 20 alert and sleep-deprived subjects who drove for 90 min. By applying a fuzzy logic-based fusion method, the true positive rate had a negligible increase, while the false positive rate was improved up to 8%. Dinges and Grace (1998) also acknowledged this feature for its correlation with vigilance and categorized following metrics as the *PERCLOS*:

- *PERCLOS70*: proportion of time for which the eyes are more than 70% closed. This

feature was used by Friedrichs and Yang (2010a) and Li et al. (2011b) as well.

- *PERCLOS80*: as mentioned at the beginning.
- *EYEMEAS (EM)*: mean square percentage of the eyelid closure rating.

Interestingly, Papadelis et al. (2007) considered “*per minute averaged blink duration*” as a feature called both *PERCLOSE* and *PERCLOS* in their work. It is not clear, if it denotes a new feature other than the common *PERCLOS*, or whether they presented their own definition of it, since both notations were used in their work. In their study with 20 sleep-deprived subjects during a real night drive, this feature increased significantly comparing the first and the last 15 min of the drive based on ANOVA. Moreover, they reported its increase shortly before a lane-based driving error.

In addition to the mentioned works, which found *PERCLOS* as a very reliable drowsiness measure, there are also some works which criticized it. Sommer and Golz (2010) believed that *PERCLOS* does not take the decreased average/maximum velocity of eye movements (for both closing and opening phases) into consideration which is an observable consequence of drowsiness. In addition, integrating this measure over a period makes it less dynamic to temporary changes. They compared a camera-based *PERCLOS* with a combination of EEG and EOG for drowsiness detection and found the former as less informative in terms of a drowsiness indicator. In addition, the high local correlation, which they achieved between *PERCLOS* and KSS, seemed to heavily depend on the length of the segment under investigation.

Another disadvantage of this feature is that it correlates better with drowsiness at the late phase in comparison to the earlier phases (Bergasa et al., 2006; Friedrichs and Yang, 2010a). In general, for those, whose eyes remain wide open despite severe drowsiness, this feature is not a good solution.

Barr et al. (2005) reviewed and introduced drowsiness detection systems based on *PERCLOS*.

Figures F.32 and F.33 show the relationship between KSS and this feature. On the contrary to the findings in other studies, which asserted an increase of *PERCLOS* due to drowsiness, in our experiments, we have found a drop of it. This might be due to the definition that we have used for *PERCLOS* and the fact that this feature was originally defined in the field of camera-based drowsiness detection rather than EOG.

- $T_x, x \in \{50, 80, 90\}$: T_x reflects the duration of the blink from $x\%$ of the absolute rise amplitude $|A_1|$ to the $x\%$ of the absolute fall amplitude $|A_2|$ of the blink (see Figures 7.5(e) and 7.5(f)). As mentioned before, some studies used T_{50} as the feature reflecting the duration of the blink (Morris and Miller, 1996; Thorslund, 2003; Johns, 2003; Åkerstedt et al., 2005; Ingre et al., 2006; Damousis and Tzovaras, 2008; Picot et al., 2010; Friedrichs and Yang, 2010a). Caffier et al. (2003) also introduced T_{90} as the closed time and showed that it is correlated with T and increases due to drowsiness. T_{80} was also used by Hu and Zheng (2009) together with other features as the input to the classifier system.

Morris and Miller (1996) found the combination of T_{50} with A and long closure rate, defined as the number of closures longer than 500 ms per minute, as the best 3-feature combination for predicting pilot errors due to drowsiness. Damousis and Tzovaras (2008) used an accumulated value of T_{50} over a 20-s window in a “*fuzzy expert system*” with other features, but only for those eye blinks whose duration were longer than 0.2 s. In a study by Anund (2009), the alteration of T_{50} for 17 sleep-deprived and non-sleep-deprived subjects during free driving and car following situations was studied. In fact, a complex

scenario was designed which forced the subjects to takeover. During all situations, T_{50} was always higher for sleep-deprived subjects, except for the takeover maneuvers which remained similar for both groups. Ingre et al. (2006) studied the interaction between KSS and T_{50} and found large individual differences between T_{50} values. They asserted that this is to a large extent “*independent of subjective sleepiness*”. This finding is in agreement with that of Friedrichs and Yang (2010a) who also observed large between-subject variation in T_{50} and consequently suggested applying a baselining step prior to its analysis.

Similar analysis to the one performed for T_c and T_o in Figure 7.10 was also performed for T_{50} with the best linear fit of $y = 0.083x + 0.69$. This shows that in our study, overall, T_{50} increased up to 1.95 times due to drowsiness. However, Åkerstedt et al. (2005), who studied the influence of sleep deprivation due to night shift working on 10 subjects, found up to 1.4 times increase of T_{50} after 120 min of driving.

Figures showing the features T_{50} , T_{80} and T_{90} versus KSS are F.34, F.35, F.36, F.37, F.38 and F.39. It seems that all three features have similar increasing trends for each subject. Moreover, if T_{50} has increased at higher KSS values, T_{80} and T_{90} have also followed that trend.

Figure 7.13 shows the scatter plot of T versus T_x , $x \in \{50, 80, 90\}$ with the best linear fit which underscores the relationship between T_x and T . The calculated Pearson correlation coefficients (will be defined in Section 7.5) between them are as follows¹: $\rho_{T,T_{50}} = 0.90$, $\rho_{T,T_{80}} = 0.82$, $\rho_{T,T_{90}} = 0.81$ (all p -values < 0.001). According to the ρ_p values, these features are highly and significantly correlated with each other.

In addition to the features mentioned here, *blink interval* and the occurrence of *blink flurries* are also mentioned in previous works as drowsiness indicator features. Wei and Lu (2012) defined the blink interval as the time between two successive blinks. Blink flurries are at least three blinks occurring within 1 s (Platho et al., 2013). It is clear that these features are highly correlated with each other and with the blink frequency. Shorter blink time interval within 1 min is the result of increased number of blinks or the occurrence of flurries. Similarly, longer blink intervals are the result of a lower blink frequency or the absence of blink flurries within a window.

We mentioned in Sections 7.1.1 and 7.1.2 that all features have been extracted within a pre-defined time window (1 min for drive time-based features and 5 min for KSS input based features). As a result, the blink flurries (if they occur) might be either located within one window or might totally be missed, if they are located on the window boundaries. In other words, flurries which are distributed outside the window under investigation will be lost. Moreover, the blink interval feature cannot be measured for the first and the last blinks of the feature extraction window. Therefore, these two features are not explored in this work.

The 19 introduced features can be categorized in two groups: *base* versus *non-base* features. We define base features as those whose values can be extracted directly from the measured EOG signal or its derivative. In other words, a clear property of base features is that they cannot be calculated through a combination of other features, since they can only be *measured*. Here, following features are categorized as base features: F , A (including A_1 and A_2), MOV , MCV , T_c , $T_{cl,1}$, T_o and T_{ro} . We consider T_x , $x \in \{50, 80, 90\}$ also as a base feature, although the value of A_1 and A_2 are required for measuring T_x value. All other features, namely E , A/MCV , A/MOV , ACV , AOV , T , $PERCLOS$ and $T_{cl,2}$ are non-base features, because they are *calculated* (not measured) using one or two base features. This categorization is of importance for the

¹The index p of the Pearson correlation coefficient ρ_p is not shown.

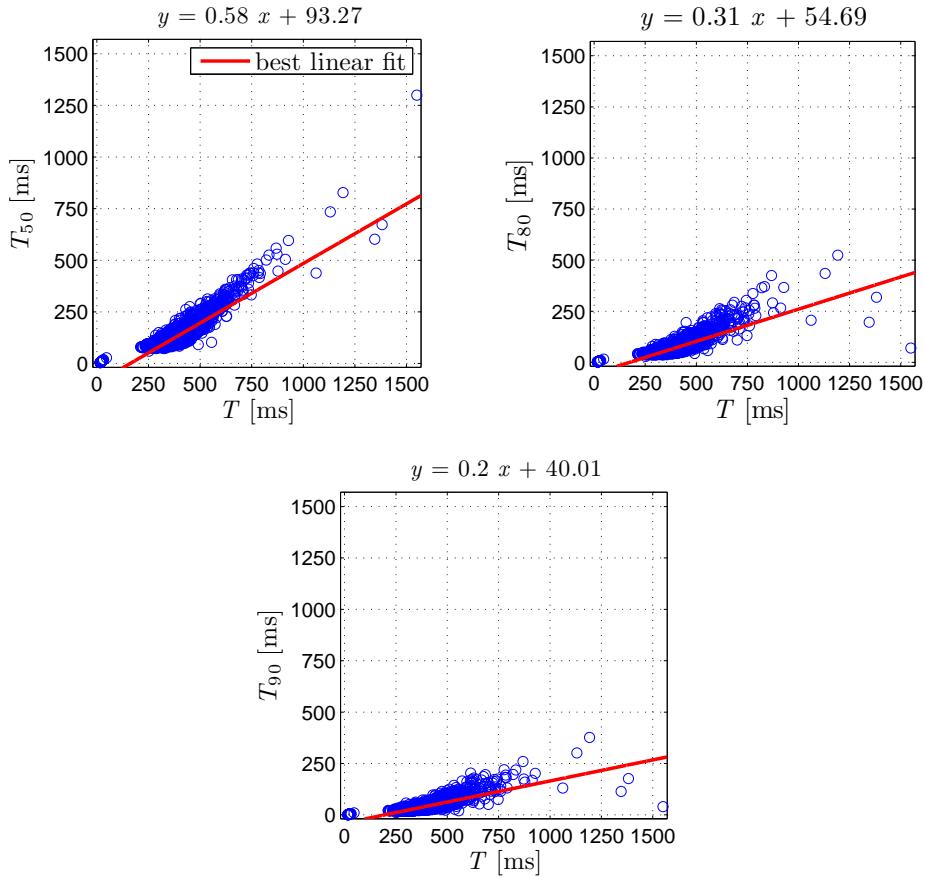


Figure 7.13.: Scatter plot of T versus T_x , $x \in \{50, 80, 90\}$. The red line shows the best linear fit with its equation on the top of each plot.

implementation of an in-vehicle driver drowsiness detection system as a product for sale. In fact, the measuring system, i.e. a driver observation camera, should be able to provide an image quality which makes the extraction of base features with an acceptable precision technically possible. If these features are of poor quality, the remaining non-base features will be even less precise.

Table 7.3 summarizes the experimental setup of most of the works reviewed in this section. The works, for which none of the information was provided, were excluded. According to this table, drowsiness detection has mostly been explored under simulated driving. Only two studies were conducted on real roads, namely Bergasa et al. (2006) and Friedrichs and Yang (2010b). Except for the studies by Caffier et al. (2003) and Schleicher et al. (2008), all previous researches had a smaller number of participants than ours in their experiments. It should be mentioned that the study by Caffier et al. (2003) was not designed as a driving task. Another issue is the participant's state of vigilance prior to the start of the experiment. Although this information was not provided in all studies, in most of them, sleep-deprived subjects were participated in the experiments. This leads to an imbalanced data set in terms of availability of information about different levels of driver drowsiness and vigilance. Section 8.1.5 will provide an in-depth discussion about this issue. As a result, a larger number of subjects, considering both real and simulated driving and including different levels of driver vigilance in the data set are the strength points of the conducted experiments in this work.

Table 7.3.: Literature review of the experiment setups. n. s.: not specified

Author	nr. of subjects	real or simulated driving	sleep-deprived subjects?
Knipling and Wierwille (1994)	12	simulated	yes
Morris and Miller (1996)	10	simulated	n. s.
Dinges and Grace (1998)	14	none	yes
Caffier et al. (2003)	60	none	no
Johns (2003)	12	simulated (only 4)	yes
Thorslund (2003)	10	simulated	no
Svensson (2004)	20	simulated	yes
Åkerstedt et al. (2005)	10	simulated	no
Johns and Tucker (2005)	5	n. s.	n. s.
Bergasa et al. (2006)	n. s.	real	n. s.
Suzuki et al. (2006)	21	simulated	n. s.
Ingre et al. (2006)	10	simulated	no
Johns et al. (2007)	8	simulated	yes (not all)
Papadelis et al. (2007)	22	simulated	yes
Damousis and Tzovaras (2008)	35	simulated	n. s.
Jammes et al. (2008)	14	simulated	n. s.
Schleicher et al. (2008)	129	simulated	n. s.
Hu and Zheng (2009)	37	simulated	yes
He et al. (2010)	n. s.	simulated	n. s.
Friedrichs and Yang (2010a)	30	real	no
Rosario et al. (2010)	20	simulated	yes
Picot et al. (2010)	20	simulated	yes
Sommer and Golz (2010)	16	simulated	n. s.
Wei and Lu (2012)	5	n. s.	n. s.

A literature review of all introduced features is listed in Table 7.4. This table shows which features have been analyzed in different studies and which trends have been found for them with respect to drowsiness.

7.3. Saccade features

Saccades were defined in Section 3.3 and were characterized as essential movements for performing the driving task properly. The detection method of the saccades and the corresponding start and end points were defined in Section 5.2 according to Figure 5.5. Many of the features extracted for blinks can also be extracted similarly for saccades, such as frequency, amplitude, duration, maximum and average velocity. In the following, we define these features based on $H(n)$:

- frequency or saccade rate: defined as the number of saccades which occur within a specified time interval. In Chapter 6, we calculated this feature within the time interval of 1 min. As mentioned in Section 2.1.2 and shown in Figure 2.5, performing an auditory secondary task, which is representative of a cognitive task, leads to a smaller number of horizontal saccades. In other words, cognitive load shrinks the scanning scene which is in agreement with the findings of Rantanen and Goldberg (1999).
- amplitude: this feature was defined in (5.4) as the difference between the amplitude $H(n)$ of the start and end points of a detected saccade. It also characterizes the amount of

Table 7.4: Literature review of the features introduced in this work. Trends versus drowsiness are either pos.: positive or neg.: negative. n. s.: the feature was studied without its trend being specified. * reduced vigilance, ** before a driving error, *** based on another end point for blinks

Author	Features	A	E	MCV	MOV	A/MCV	ACV	AOV	F	T	T _c	T _o	T _{cl,1}	T _{cl,2}	T _{ro}	PERCLOS	T ₅₀	T ₈₀	T ₉₀
Knipping and Wierwille (1994)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Morris and Miller (1996)	neg.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	n. s.	-	-
Dinges and Grace (1998)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Caffier et al. (2003)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	pos.
Johns (2003)	-	-	-	-	-	pos.	-	-	-	-	-	-	-	-	-	-	-	-	-
Hargutti (2003)	neg.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Thorslund (2003)	neg.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Svensson (2004)	neg.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Åkerstedt et al. (2005)	-	-	-	-	pos.	pos.	-	-	-	-	-	-	-	-	-	-	-	-	-
Johns and Tucker (2005)	-	-	-	-	pos.	pos.	-	-	pos.	pos.	pos.	-	-	-	-	-	-	-	-
Bergasa et al. (2006)	-	-	-	-	-	-	-	-	pos.	pos.	-	-	-	-	-	-	-	-	-
Suzuki et al. (2006)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ingre et al. (2006)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Johns et al. (2007)	-	-	-	-	n. s.	n. s.	-	-	-	-	-	-	-	-	-	-	n. s.	-	-
Papadellis et al. (2007)	-	-	-	-	-	-	-	-	pos.**	pos.	-	-	-	-	-	-	-	-	-
Damousis and Tzovaras (2008)	-	-	n. s.	-	-	-	-	-	-	-	n. s.	-	-	-	-	-	n. s.	-	-
Jammes et al. (2008)	pos.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Schleicher et al. (2008)	-	-	-	-	-	-	-	-	pos.***	-	-	-	-	-	-	-	-	-	-
Hu and Zheng (2009)	n. s.	-	n. s.	n. s.	-	n. s.	n. s.	-	n. s.	n. s.	n. s.	-	-	-	-	-	n. s.	n. s.	-
He et al. (2010)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sigari (2009)	-	-	-	-	pos.	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Friedrichs and Yang (2010a)	n. s.	pos.	-	-	-	-	-	-	n. s.	-	-	-	-	-	-	-	pos.	-	-
Rosario et al. (2010)	-	-	-	-	-	n. s.	-	-	-	n. s.	-	-	-	-	-	-	pos.	-	-
Picot et al. (2010)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	n. s.	n. s.	-
Sommer and Goltz (2010)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	pos.	-	-
Horak (2011)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	n. s.	-	-
Li et al. (2011b)	-	-	n. s.	n. s.	n. s.	-	-	n. s.	n. s.	n. s.	n. s.	-	-	-	-	-	pos.	-	-
Wei and Lu (2012)	-	n. s.	n. s.	n. s.	-	-	n. s.	n. s.	n. s.	n. s.	n. s.	-	-	-	-	-	n. s.	n. s.	-

gaze shift. However, the contribution of the head rotation remains undetermined. On the contrary to blink amplitude, which is affected by drowsiness, saccadic amplitude is a function of the angular distance to be traveled towards any destination angle. As a result, it cannot be a drowsiness indicator in the general case.

- duration: defined as the time difference between start and end points of a saccade in $H(n)$. Schleicher et al. (2008) found that the standard deviation of the saccade duration correlates best with the video-labeled drowsiness in a driving simulator study.
- maximum velocity: similar to MCV and MOV , this feature is calculated using the derivative signal. Rowland et al. (2005) reported the drop of this feature due to sleep-deprivation.
- average velocity: defined as the ratio between amplitude and duration of a saccade.

Schleicher et al. (2008) believed that for the data collection and measurement of saccades, a sampling rate between 500 and 1000 Hz is required. Moreover, they stated that drowsy drivers scan the scene ahead “*unsystematically*” in comparison to the awake ones. According to our observations in the conducted experiments, the occurrence of saccades depends most of the time directly on the surrounding events. Under highly monotonous driving conditions, saccades occur seldom and irregularly, because almost nothing outside of the vehicle attracts driver’s attention. Therefore, in this case, a smaller number of saccades is irrelevant to drowsiness. On the contrary, under real driving conditions with high traffic density or on urban streets with lower vehicle speed, the drivers scan the environment more frequently leading to a larger number of saccades which is again independent of driver drowsiness. This emphasizes the fact that assessment of frequency and amplitude of saccades make sense only if the events occurring outside of the vehicle are to some extent quantifiable in terms of traffic density with sensors such as radars. In addition, reproducible driving scenarios, which are only possible in the driving simulators, are crucial for concluding conducted experiments. Therefore, real road experiments are not suitable for assessing saccadic features due to varying traffic density for each subject. Here, the experiment conducted in the driving simulator was designed to be monotone to accelerate driver drowsiness. Hence, a low saccade rate, as an example, cannot necessarily be the consequence of driver’s low vigilance. Apart from the mentioned points, saccades are very fast eye movements and, as a result, a higher sampling frequency is needed for assessing their duration or velocity. Considering all these accounts, in this work, the saccadic features are not studied.

7.4. Event-based analysis of eye blink features

In this section, the variation of the introduced features before the occurrence of an event, e.g. a safety-critical one, is studied and compared with a reference condition. Afterwards, based on a statistical test analysis, it is studied whether the variation of the corresponding feature, i.e. its decrease or increase, is statistically significant or not. Such analyses provide information about predictability of the safety-critical events based on the feature variation in the course of time.

As mentioned in Chapter 2, Schmidt et al. (2011) considered the moments of verbal assessment of the driver’s state of vigilance as an event and compared the variation of blink duration during the event with a baseline and with the time interval after the event. In other studies, a safety-critical event was investigated instead, such as a lane departure, defined as one or two wheels outside the lane marking, hitting the rumble strip (Papadelis et al., 2007; Anund et al., 2008) and eye closures longer than 500 ms (Schleicher et al., 2008). Afterwards, the variation of physiological measures such as blink or EEG features were explored within the time interval shortly before

the event occurrence in comparison to a baseline or to the time interval shortly after the event occurrence.

In Sections 7.4.1 and 7.4.2, similarly, the moment of the first unintentional lane departure and the first unintentional eye closure longer than 1 s are considered as safety-critical events and the alteration of blink features shortly before their occurrences are studied.

7.4.1. Event 1: Lane departure

Here, a lane departure event was found by analyzing the lane lateral distance signal measured by a multi-purpose camera system (Seekircher et al., 2009). This signal represented the distance between the vehicle center and middle of the lane. Since, in real life, not all lane departures are unintentional, an offline video labeling was performed additionally to validate detected lane departure events. Figure 7.14 shows examples of intentional and unintentional lane departure events.

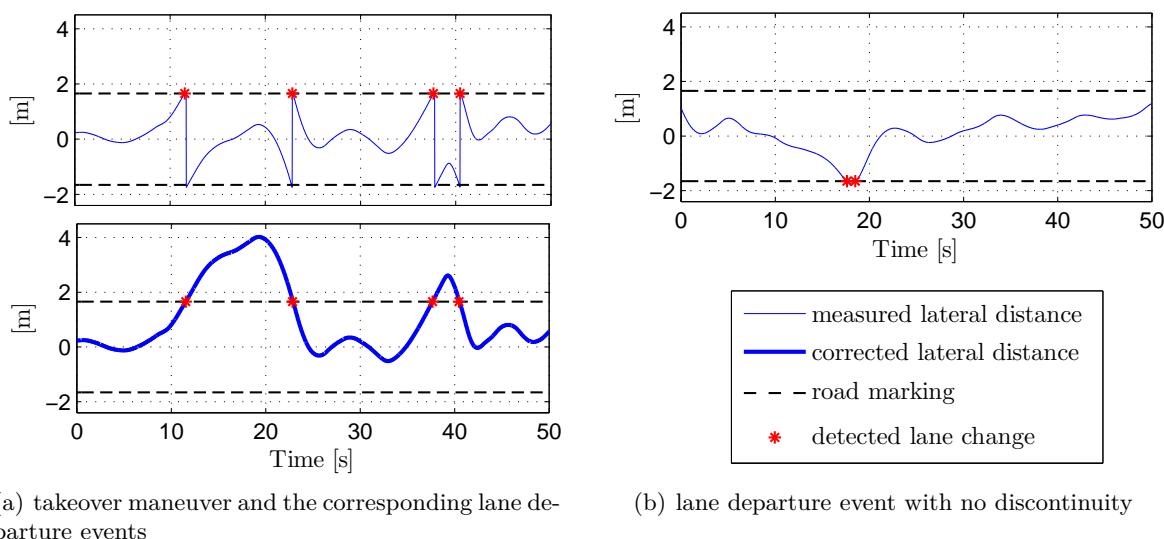


Figure 7.14.: Examples of intended (takeover maneuver) and unintended (lane departure) lane change events visible in the vehicle's lateral distance signal

Top plot of Figure 7.14(a) depicts four discontinuity moments which occurred due to the change to a new reference lane marking by the camera (Ebrahim, 2011). The red asterisks show the moment when the vehicle center passed the lane marking. These discontinuities were corrected in the bottom plot. The first two lane changes show a representative example of a takeover maneuver lasting for about 10 s. However, based on the offline video analysis, the third and the forth asterisks represent an unintended lane departure due to drowsiness with the length of 2.5 s. The durations are calculated at the time instant that 40% of the vehicle's width crossed over a lane marking. This threshold, i.e. 40% of the vehicle's width, was found based on the 99th percentile of all measured lateral distances, excluding takeover maneuver events¹. On the other hand, offline video analysis of the drives showed that most of the subjects did not interpret

¹Two succeeding lane changes, which occurred within at least 4 s and the maximum gap of 10 s, were considered as potential takeover maneuver and were consequently excluded for the threshold calculation. Moreover, the participants were instructed to keep right as much as possible during the experiment in accordance with the German official traffic regulations. Therefore, lane changes occurred always in pairs.

crossing the lane marking within a certain limit as safety-critical, e.g. one tire over the lane marking. This was evident, as their steering movements for correction were not large enough, although they partly left the lane. In fact, the subjects steered with larger movements towards the middle of the lane only after having reached a certain limit, not as soon as they slightly crossed the lane marking with one tire. Since the width of a tire of the vehicle used in our experiments (an S-Class) was 15% of the total vehicle width, a larger threshold was needed for the definition of an unintended lane departure event. In addition to the given points, we were looking for safety-critical lane departure events which did not occur frequently. For this reason, the 99th percentile of all measured lateral distances is a representative value as the threshold which is equivalent to the moment that about 40% of the vehicle crossed the lane marking. The first time that the threshold was exceeded and validated by the offline video analysis has been considered as a safety-critical lane departure event. Figure 7.14(b) also shows an example of a lane departure, although no discontinuity is evident in the measured lateral distance signal. Such lane departures were included in the detection of safety-critical lane departure events, if they exceeded the mentioned threshold and were validated by the offline video analysis.

Figure 7.15 shows the mean value of EWVAR (Friedrichs and Yang, 2010a) (see (4.5)) of the lateral distance for each KSS value regarding 25 subjects (S1 to S25) who drove in the driving simulator. The parameter N_{σ^2} was set to 250 samples. The error bars refer to the standard deviation of the EWVAR values. The moments of intended lane changes were excluded for the calculation of EWVAR. It can be seen that the variance of lateral distance increased due to drowsiness in our experiment which agrees with the studies mentioned in Section 2.1.1. Despite the fact that some subjects have rated themselves as very drowsy, this feature does not follow any trend for them (e.g. subjects S16, S21 and S23) or the trend is very weak (subjects S6, S9, S10, S13 and S17).

It is interesting to know at which subjective drowsiness level, i.e. KSS, the first unintended lane departure has occurred. These values are listed in Table 7.5. Two subjects are excluded, because the offline video analysis was not possible for them. For subjects, who never left the lane unintentionally with respect to our threshold, the maximum rated subjective drowsiness levels are listed. This clarifies two points: 1) If the maximum KSS value for these subjects is smaller than 6, then the nonexistence of the unintentional lane departure is due to high driver vigilance, 2) If the maximum KSS value for these subjects is larger than 6, the nonexistence of the unintentional lane departure shows that the calculated feature is either not meaningful to assess their drowsiness level or these subjects overestimated their drowsiness level. According to the listed values, the lane departure event occurred mostly at a time that the subject also believed that he was drowsy (17 subjects). However, there are 6 subjects who never left the lane according to our criteria, although they rated themselves as drowsy.

Table 7.5.: Left table: number of occurrences of KSS values at the time of first unintended lane departure and number of occurrences for the maximum value of KSS, if no lane departure was detected.
Right table: confusion matrix.

		KSS									
		awake			drowsy						
		1	2	3	4	5	6	7	8	9	
lane departure	first	0	0	0	0	0	0	1	6	10	
	none	0	0	0	0	0	0	1	1	4	

		KSS									
		awake			drowsy						
		yes	no		yes	no		total		total	
lane departure	yes	0	17		0	17		17		17	
	no	0	6		0	6		6		6	
	total	0	23								

Figure 7.16 shows the boxplots which compare the mean value of a feature within the first 5 min

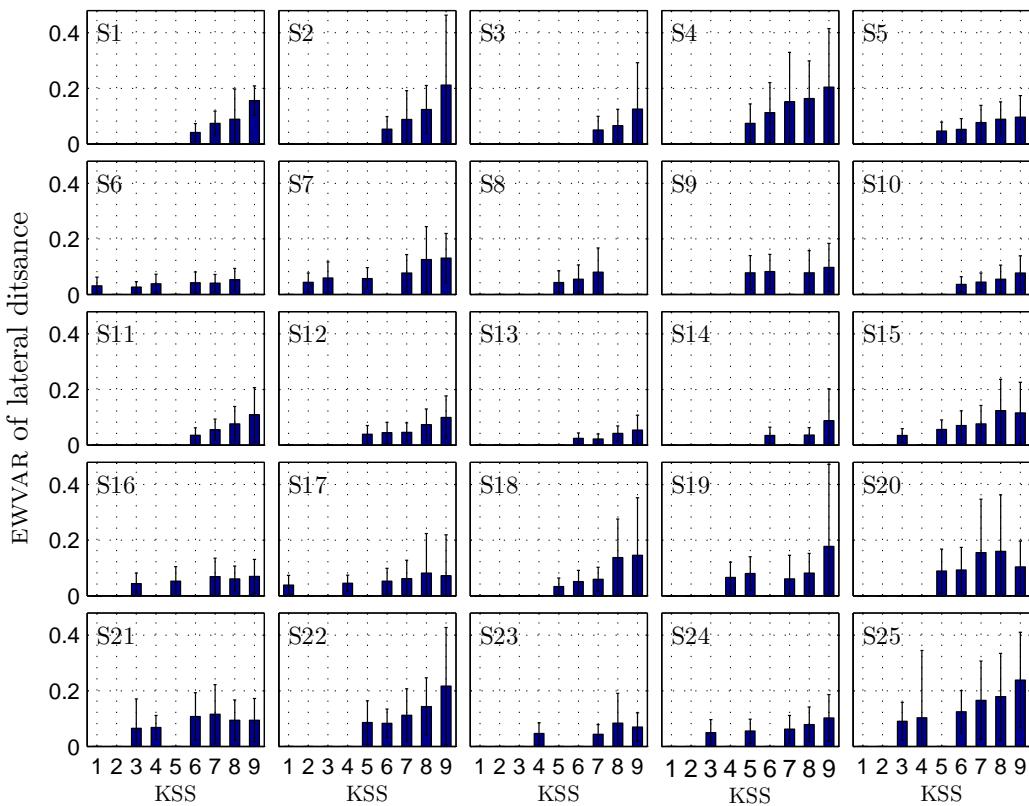


Figure 7.15.: The mean of the EWVAR of lateral distance versus KSS values for 25 subjects who drove in the driving simulator. The standard deviations are also shown.

of the drive¹ with the 5-min interval before the lane departure for 23 subjects. Participants of real road drives were not considered, because their collected data were mainly related to awake driving. Therefore, no lane departure due to drowsiness occurred for them. For most of the features, an increasing or a decreasing trend is apparent.

With a statistical test such as the *paired-sample t-test* (see Appendix D.1), it is possible to show whether the difference between the means of the groups (i.e. the first 5 min of the drive and the last 5 min before the event) are significantly different (Field, 2007). If the assumption of this test is not fulfilled, i.e. normality of the difference between observations is not given, the alternative non-parametric test, namely the *Wilcoxon signed-rank test* (see Appendix D.7) is applied instead. The normality of the distribution of the difference between groups was analyzed with the *Lilliefors test* (see Appendix D.2).

Table 7.6 summarizes the results of the mentioned tests including the test significance value which is t_0 for the paired-sample *t-test* or z_0 for the Wilcoxon signed-rank test and the corresponding p -values with respect to the significance level of 5%. The test significance value of the Wilcoxon signed-rank test z_0 indicates that the difference between the means of the corresponding feature in both groups was not normally distributed. According to the results, except for *E*, all results are statistically significant with p -value < 0.05. As a result, we conclude that *A*, *MCV*, *MOV*, *ACV*, *AOV* and *PERCLOS* decreased significantly 5 min before the lane departure in comparison to the first 5 min of the drives. On the contrary, *A/MCV*, *A/MOV*, *F*, *T*, *T_c*, *T_o*, *T_{d,1}*, *T_{d,2}*,

¹As mentioned in Section 4.3, in the experiment conducted in the driving simulator, the very first 3 min of the drives were removed from the collected EOG data sets, because it was considered as the phase that the eyes needed to accommodate. Therefore, the first 5 min, which is studied here, does not include this phase.

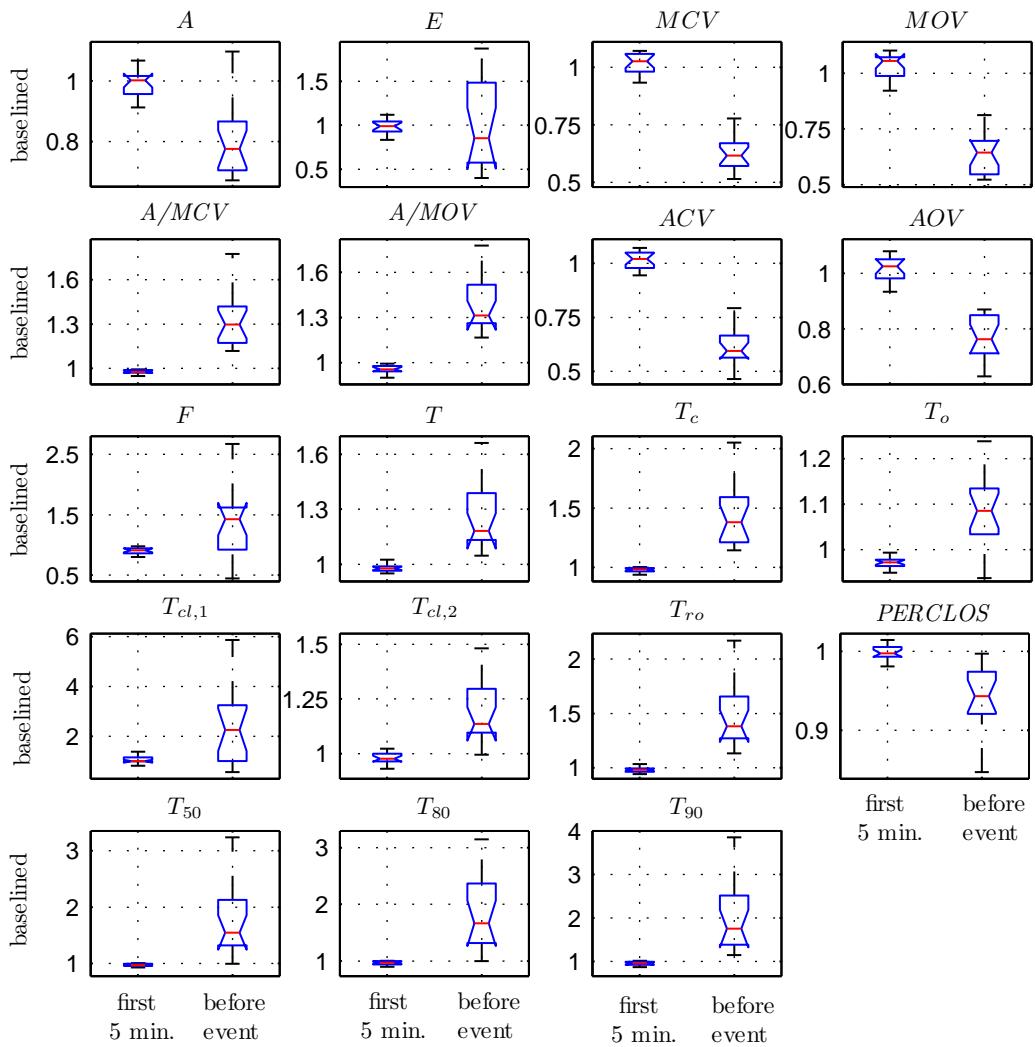


Figure 7.16.: The mean of all baselined features over the first and the last 5 min before the first unintended lane departure event for 23 subjects who drove in the driving simulator

T_{ro} , T_{50} , T_{80} and T_{90} increased significantly before the mentioned event.

Now, the question is whether the found trends of the features are the consequence of drowsiness or rather time-on-task. To answer this question, a similar analysis was performed for 18 subjects who drove under real road conditions for covering the EOG data collection of the awake phase. Since none of them experienced a lane departure event, the last 5 min of the drive has been compared to the first 5 min. Almost all subjects rated their drowsiness level with a higher value of KSS at the end of the drive in comparison to the start of the driving. Nevertheless, overall, none of them reported severe drowsiness.

Figure 7.17 shows the boxplots of all features for these subjects. Comparing this figure with Figure 7.16, it can be concluded that first of all, larger overlap between the boxplots is evident. Moreover, some features, such as T or T_c , show contradicting trends. Interestingly, the decreasing trends of A , MCV , MOV , ACV and AOV are still visually distinguishable, comparing the first and the last 5 min of the drives. However, a larger decrease of these features has occurred in the interval before the lane departure shown in Figure 7.16. Similar to Table 7.6, Table 7.7 shows the statistical comparison of the mean values of the features shown in Figure 7.17. As can be

Table 7.6.: Results of paired-sample t -test (t_0) and Wilcoxon signed-rank test (z_0) for all features shown in Figure 7.16. Red color indicates non-significant features.

Feature	test significance	p-value
<i>A</i>	$t_0 = 4.07$	< 0.05
<i>E</i>	$z_0 = -0.17$	0.87
<i>MCV</i>	$t_0 = 10.81$	< 0.05
<i>MOV</i>	$t_0 = 11.31$	< 0.05
<i>A/MCV</i>	$t_0 = -6.69$	< 0.05
<i>A/MOV</i>	$t_0 = -9.46$	< 0.05
<i>ACV</i>	$t_0 = 12.52$	< 0.05
<i>AOV</i>	$t_0 = 6.03$	< 0.05
<i>F</i>	$t_0 = -2.82$	< 0.05
<i>T</i>	$z_0 = -3.57$	< 0.05
<i>T_c</i>	$t_0 = -6.51$	< 0.05
<i>T_o</i>	$t_0 = -5.03$	< 0.05
<i>T_{cl,1}</i>	$t_0 = -3.40$	< 0.05
<i>T_{cl,2}</i>	$z_0 = -3.53$	< 0.05
<i>T_{ro}</i>	$z_0 = -3.62$	< 0.05
<i>PERCLOS</i>	$t_0 = 5.74$	< 0.05
<i>T₅₀</i>	$t_0 = -5.29$	< 0.05
<i>T₈₀</i>	$t_0 = -4.71$	< 0.05
<i>T₉₀</i>	$z_0 = -3.48$	< 0.05

seen, the difference between the means of the first and the last 5 min of the drives is for most of the features non-significant (p -value > 0.05). However, the mentioned features related to the amplitude of a blink, namely *A*, *E*, *ACV*, *AOV*, *MCV* and *MOV*, show a significant decrease. This might be due to the fact that these features are subject to time-on-task to a larger extent in comparison to the duration-based features. Considering these findings, we conclude that the significant variation of features before the unintended lane departure event is mostly due to the driver drowsiness.

We emphasize that the goal of comparing Figures 7.16 and 7.17 was to highlight the contribution of drowsiness towards the changes in feature mean values before the occurrence of the lane departure event. Studying the intrinsic difference between driving simulator and real road conditions is outside the scope of this analysis and will be explored in the next chapter.

7.4.2. Event 2: Microsleep

Similar to the previous approach, an unintended eye closure longer than 1 s, called a microsleep, was considered as a safety-critical event. We investigate this event due to the following reason. In our driving simulator experiment, we observed that at the earlier phase of drowsiness some subjects had longer eye closures, although they did not leave the lane. In other words, not all long eye closures led to a lane departure, although their occurrence was the consequence of driver drowsiness.

For the first 14 subjects, who participated in the driving simulator experiment, an additional *Dikablis glasses* (Ergoneers GmbH, 2014) was used. It is an eye tracking measurement system which provides a signal with either zero or non-zero values. Non-zero values of the signal are not of interest in this work. However, zero values occur, if the eye-tracker does not detect the pupil. The pupil will not be detected either due to an eye closure (both intentional and unintentional) or due to technical problems. Hence, zero sequences lasting longer than 1 s refer to potential

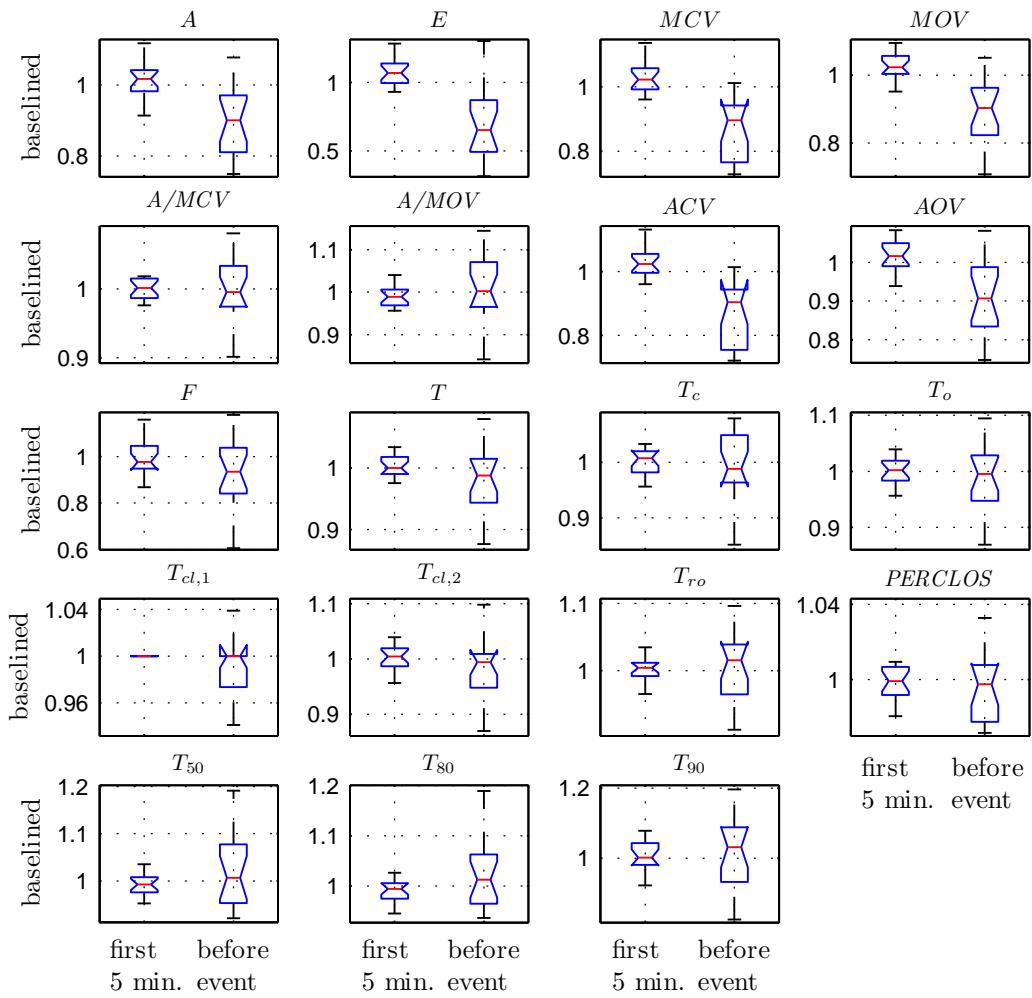


Figure 7.17.: The mean of all baselined features over the first and the last 5 min of the drive for 18 subjects who drove under real conditions

microsleep events. Since some long eye closures might have occurred intentionally or due to technical problems, by offline video labeling, non-relevant events were discarded. At the end, the first unintended eye closure lasting at least 1 s, namely a microsleep, was considered as a safety-critical event. The microsleep event detection by the Dikablis glasses instead of the EOG makes this analysis independent of the event detection approach in this work.

Similar to the lane departure events, we want to know at which subjective drowsiness level, i.e. KSS, the first microsleep occurred. It is also possible that some subjects never closed their eyes unintentionally for at least 1 s. For these subjects, it is interesting to know what their maximum subjective drowsiness level was given that no microsleep was detected. Here, it also clarifies whether nonexistence of a microsleep was the result of high driver vigilance or whether the subjects overestimated their drowsiness level. These values are shown in Table 7.8. Interestingly, it seems that two subjects have underestimated their drowsiness level by choosing KSS = 4 and 5, although they unintentionally closed their eyes longer than 1 s. This was also validated by the offline video analysis. Moreover, three subjects rated themselves as very drowsy, although no eye closure longer than 1 s was detected in their data. On the contrary, nine subjects found themselves drowsy during the occurrence of the first microsleep. There was no subject who rated himself as awake during the entire drive given that no microsleep was detected.

Table 7.7.: Results of paired-sample t -test (t_0) and Wilcoxon signed-rank test (z_0) shown in Figure 7.17. Red color indicates non-significant features.

Feature	test significance	p-value
<i>A</i>	$t_0 = 4.61$	< 0.05
<i>E</i>	$t_0 = 5.17$	< 0.05
<i>MCV</i>	$t_0 = 5.69$	< 0.05
<i>MOV</i>	$t_0 = 4.03$	< 0.05
<i>A/MCV</i>	$z_0 = -0.37$	0.71
<i>A/MOV</i>	$t_0 = -0.79$	0.44
<i>ACV</i>	$t_0 = 5.32$	< 0.05
<i>AOV</i>	$t_0 = 3.62$	< 0.05
<i>F</i>	$t_0 = 2.44$	< 0.05
<i>T</i>	$t_0 = 1.40$	0.18
<i>T_c</i>	$t_0 = 0.69$	0.50
<i>T_o</i>	$t_0 = 1.13$	0.28
<i>T_{cl,1}</i>	$z_0 = -1.86$	0.08
<i>T_{cl,2}</i>	$t_0 = 1.45$	0.17
<i>T_{ro}</i>	$t_0 = -0.45$	0.66
<i>PERCLOS</i>	$t_0 = 0.51$	0.62
<i>T₅₀</i>	$t_0 = -1.46$	0.16
<i>T₈₀</i>	$t_0 = -2.04$	0.06
<i>T₉₀</i>	$t_0 = -0.50$	0.63

Table 7.8.: Left table: number of occurrences of KSS values at the time of first microsleep and the number of occurrences for the maximum value of KSS, if no microsleep was detected. Right table: confusion matrix.

KSS						
		awake		drowsy		
		1	2	3	4	5
		7	8	9		
microsleep		1	2	3	4	6
first		0	0	0	1	1
none		0	0	0	0	0
		1	6	2		
		1	0	2		

KSS						
		awake		drowsy		total
		yes	no	yes	no	
microsleep		2	9	11		
		0	3	3		
		2	12			

Similar to Section 7.4.1, the mean value of the features within the first 5 min of the drive is compared with the 5-min interval before the occurrence of the microsleep event as shown in Figure 7.18. It can be seen that the interquartile range of the boxplots (see Appendix B) for most of the features are not overlapping, except for *A*, *E*, *AOV* and *T_{cl,1}*. Comparing plots of this figure with those of Figure 7.16, it can be deduced that a lane departure occurs during later phases of drowsiness. As an example, before a lane departure event, *ACV* decreased to 60% of its magnitude at the beginning of the drive. However, before the first microsleep, it has only dropped to 75% of its initial value. This is also the case for *MCV*, *MOV* and *T_c*. In general, the comparison of both figures should be done with care, because different numbers of subjects were considered for the event analyses. Apart from that on average, the first unintended microsleep occurred after 110 ± 26 min of driving, while the first unintentional lane departure occurred after 128 ± 40 min. These values agree with the conclusions made based on the alteration of blink features. Moreover, the time difference between the occurrence of each safety-critical event implies that driver physiological measures outperform the driver performance measures in early prediction of the driver drowsiness.

Table 7.9 shows the results of the statistical tests. Most of the results are based on the Wilcoxon signed-rank test which means that most of the differences between distributions were not normally distributed. Clearly, it is due to the small number of available samples for each feature.

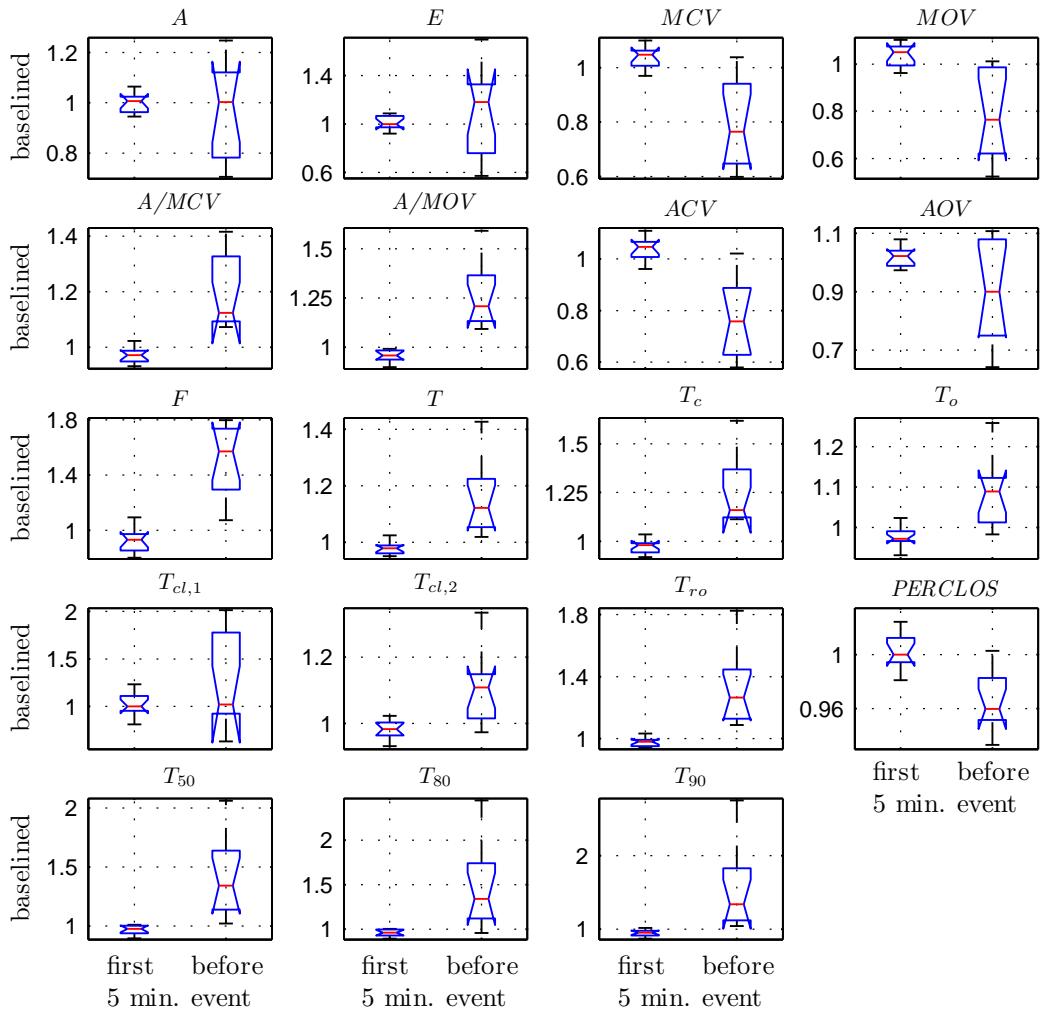


Figure 7.18.: The mean of all baselined features over the first and the last 5 min before an unintended microsleep for 11 subjects who drove in the driving simulator

Except for A and E , all features varied significantly in comparison to the beginning of the drive.

7.5. Correlation-based analysis of eye blink features

In this section, the relationship between extracted features is analyzed statistically. First, based on the correlation analysis, it will be shown to what extent each feature correlates with the KSS values. This is done for both drive time-based and KSS input-based features to show how informative they are in terms of a drowsiness indicator. Moreover, it is also studied which features are correlated with each other. Such analysis is, in general, important for knowing the amount of redundant information available in the extracted feature set.

7.5.1. Case 1: correlation between a feature and KSS values

Since our goal is the detection of driver drowsiness, first of all, the relationship of each feature with the drowsiness is explored. The *Pearson product-moment correlation coefficient* and

Table 7.9.: Results of paired-sample t -test (t_0) and Wilcoxon signed-rank test (z_0) for all features shown in Figure 7.18. Red color indicates non-significant features.

Feature	test significance	p-value
<i>A</i>	$t_0 = 1.15$	0.26
<i>E</i>	$z_0 = -1.07$	0.28
<i>MCV</i>	$z_0 = -5.74$	< 0.05
<i>MOV</i>	$z_0 = -5.32$	< 0.05
<i>A/MCV</i>	$z_0 = -6.45$	< 0.05
<i>A/MOV</i>	$z_0 = -6.45$	< 0.05
<i>ACV</i>	$z_0 = -5.99$	< 0.05
<i>AOV</i>	$z_0 = 3.91$	< 0.05
<i>F</i>	$t_0 = -7.01$	< 0.05
<i>T</i>	$t_0 = -6.01$	< 0.05
<i>T_c</i>	$z_0 = -6.45$	< 0.05
<i>T_o</i>	$z_0 = -5.06$	< 0.05
<i>T_{cl,1}</i>	$z_0 = -2.30$	< 0.05
<i>T_{cl,2}</i>	$z_0 = -4.55$	< 0.05
<i>T_{ro}</i>	$z_0 = -6.32$	< 0.05
<i>PERCLOS</i>	$z_0 = 7.09$	< 0.05
<i>T₅₀</i>	$z_0 = -5.86$	< 0.05
<i>T₈₀</i>	$z_0 = -5.81$	< 0.05
<i>T₉₀</i>	$z_0 = -5.76$	< 0.05

Spearman's rank correlation coefficient are two possibilities for quantifying the linear and non-linear association between features and KSS values, respectively. This analysis is also called *inter-correlation*.

It should be mentioned that, in general, for stepwise and ordinal values like KSS, the mentioned coefficients are not necessarily an optimal evaluation method. The reason is that the values of a feature might vary, even though their corresponding KSS values do not change. This fact affects the values of the correlation coefficients and their interpretation severely.

Pearson product-moment correlation coefficient

The Pearson product-moment correlation coefficient $\rho_p(\mathbf{x}, \mathbf{y})$, which quantifies the linear association between two vector variables \mathbf{x} and \mathbf{y} (Artusi et al., 2002; Field, 2007), is defined as follows

$$\rho_p(\mathbf{x}, \mathbf{y}) = \frac{\text{Cov}(\mathbf{x}, \mathbf{y})}{\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}} = \frac{\sum_{i=1}^N (x_i - \mu_{\mathbf{x}})(y_i - \mu_{\mathbf{y}})}{\sqrt{\sum_{i=1}^N (x_i - \mu_{\mathbf{x}})^2 \sum_{i=1}^N (y_i - \mu_{\mathbf{y}})^2}}. \quad (7.9)$$

Here, vectors \mathbf{x} and \mathbf{y} denote all N samples of a feature and the corresponding KSS values. $\text{Cov}(\mathbf{x}, \mathbf{y})$ refers to the covariance calculated for \mathbf{x} and \mathbf{y} . μ and σ correspond to the mean and standard deviation of \mathbf{x} and \mathbf{y} , respectively. Values of $0 < \rho_p(\mathbf{x}, \mathbf{y}) \leq 1$ indicate that \mathbf{x} and \mathbf{y} are positively correlated. Similarly, values of $-1 \leq \rho_p(\mathbf{x}, \mathbf{y}) < 0$ denote a negative relationship between the variables. If two variables do not relate linearly to each other, $\rho_p(\mathbf{x}, \mathbf{y})$ is close to zero. The closer the value of $\rho_p(\mathbf{x}, \mathbf{y})$ to ± 1 , the stronger is the linear association between \mathbf{x} and \mathbf{y} . Field (2007) categorizes the values of ρ_p as follows: small effect for $\rho_p = \pm 0.1$, medium effect for $\rho_p = \pm 0.3$ and large effect for $\rho_p = \pm 0.5$. However, depending on the field of research and

the issue being addressed, the value of ρ_p might be interpreted differently.

Based on a hypothesis test, namely the t -test, it is possible to analyze the significance of the calculated ρ_p and to show whether ρ_p is significantly different from zero as explained in Appendix D.3.

Spearman's rank correlation coefficient

The Spearman's rank correlation coefficient ρ_s quantifies the “*general monotonicity of the underlying relationship*” between two variables (Artusi et al., 2002). This means, if one variable increases, synchronously the other one increases or decreases. In this case, the Spearman's rank correlation coefficient has a high value independent of the existing or non-existing linear relationship between the variables. Therefore, Spearman's rank correlation coefficient also quantifies the amount of non-linear relationship between two variables. This is unlike the Pearson product-moment correlation coefficient which quantifies to what extent an existing relationship is close to a linear one.

Since the similarity between an arbitrary monotonic function and the underlying relationship between variables is looked for, first, all samples of \mathbf{x} and \mathbf{y} are sorted descendingly. $\rho_s(\mathbf{x}, \mathbf{y})$ is then calculated using the ranks of sorted values as follows

$$\rho_s(\mathbf{x}, \mathbf{y}) = 1 - \frac{6 \sum_{i=1}^N (\text{rank}(\mathbf{x})_i - \text{rank}(\mathbf{y})_i)^2}{N(N^2 - 1)}, \quad (7.10)$$

where $\text{rank}(\mathbf{x})_i$ and $\text{rank}(\mathbf{y})_i$ denote the i -th rank of \mathbf{x} and \mathbf{y} . (7.10) is valid as long as there are no identical values in variables \mathbf{x} and \mathbf{y} . For identical rank values, called *ties*, the average of the found rank is used (rank_{tie}) and the calculation of $\rho_s(\mathbf{x}, \mathbf{y})$ is more complex as follows

$$\rho_s(\mathbf{x}, \mathbf{y}) = \frac{N(N^2 - 1) - \frac{1}{2} \sum_{i=1}^{N_1} (r_{\mathbf{x}, i}^3 - r_{\mathbf{x}, i}) - \frac{1}{2} \sum_{i=1}^{N_2} (r_{\mathbf{y}, i}^3 - r_{\mathbf{y}, i}) - 6 \sum_{\text{tie}}^N \left(\text{rank}_{\text{tie}}(\mathbf{x})_i - \text{rank}_{\text{tie}}(\mathbf{y})_i \right)^2}{\sqrt{\left(N(N^2 - 1) - \sum_{i=1}^{N_1} (r_{\mathbf{x}, i}^3 - r_{\mathbf{x}, i}) \right) \left(N(N^2 - 1) - \sum_{i=1}^{N_2} (r_{\mathbf{y}, i}^3 - r_{\mathbf{y}, i}) \right)}}. \quad (7.11)$$

In the above equation, N_1 and N_2 denote the numbers of elements in \mathbf{x} and \mathbf{y} excluding their duplicate values. $r_{\mathbf{x}, i}$ and $r_{\mathbf{y}, i}$ refer to the numbers of observations with identical ranks.

Similar to ρ_p , ρ_s can also be tested for showing its significant difference from zero. More details is provided in Crawshaw and Chambers (2001).

In the following, the introduced correlation coefficients are calculated for KSS input-based and drive time-based features.

KSS input-based features: All baselined KSS input-based features are shown in Figure 7.19 versus KSS values. The features are baselined with respect to the average of the first two intervals before a KSS input. Spearman's rank correlation coefficient ρ_s and Pearson correlation coefficients ρ_p between these features and KSS values are listed in Table 7.10. The values are sorted with respect to $|\rho_s|$. As mentioned before, ρ_p shows the strength of the linear relationship between features and KSS values. However, drowsiness may also evolve non-linearly in the course of time

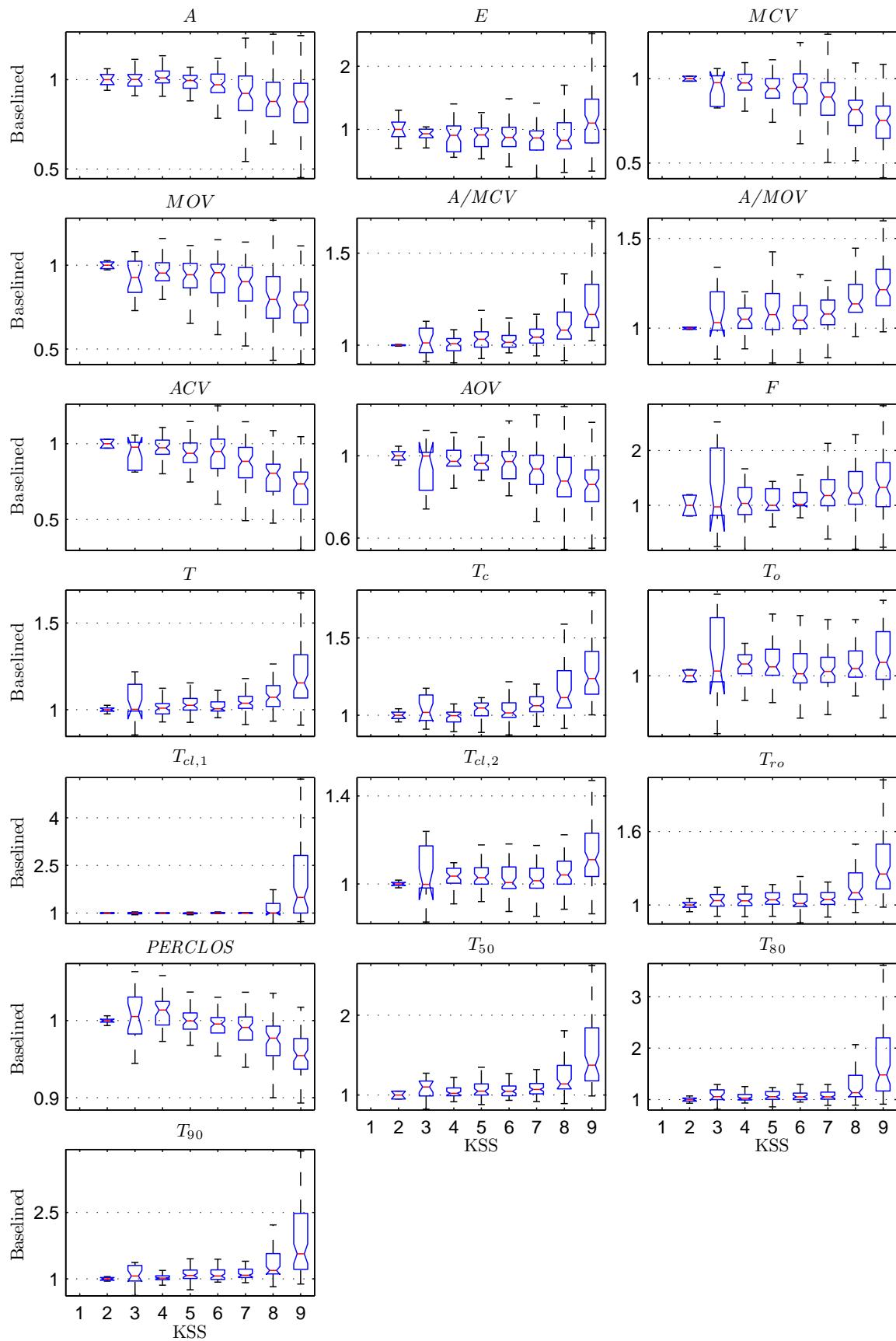


Figure 7.19.: Boxplot of baselined KSS input-based features for all subjects versus KSS values

which makes ρ_s a more suitable measure for studying the relationship between features and KSS values.

For all features, except for T_o , the calculated correlation coefficients are significantly different from zero, i.e. $p\text{-value} < 0.05$. For most of the features $|\rho_s|$ is also larger than $|\rho_p|$ which underlines the non-linear relationship between feature's evolution and KSS values. Furthermore, the highest value of $|\rho_p|$ and $|\rho_s|$ occurs for different features. The features with the highest amount of linear association with KSS is *ACV*, while the highest $|\rho_s|$ is achieved for T_c . The negative signs show the inverse relationship between features and drowsiness. For example, the drop of *ACV* occurred parallel with the increase of KSS which means that the subjects closed their eyes more slowly due to drowsiness (see Figure 7.19 for *ACV*).

Table 7.10.: Sorted Spearman's rank correlation coefficient ρ_s and Pearson correlation coefficient ρ_p between all KSS input-based features and KSS values ($N = 391$). All p -values were smaller than 0.05 except for red features.

Feature	ρ_s	ρ_p	Feature	ρ_s	ρ_p
T_c	0.60	0.50	<i>MOV</i>	-0.44	-0.41
<i>A/MCV</i>	0.57	0.48	<i>A/MOV</i>	0.44	0.40
<i>ACV</i>	-0.56	-0.52	$T_{cl,2}$	0.40	0.35
T_{90}	0.55	0.45	<i>A</i>	-0.36	-0.29
<i>PERCLOS</i>	-0.53	-0.46	<i>AOV</i>	-0.33	-0.27
T_{50}	0.53	0.45	$T_{cl,1}$	0.26	0.28
<i>MCV</i>	-0.53	-0.50	<i>F</i>	0.21	0.16
T_{ro}	0.52	0.49	<i>E</i>	0.10	0.21
T_{80}	0.50	0.43	<i>T_o</i>	0.06	0.05
T	0.45	0.40			

Drive time-based features: Table 7.11 shows the sorted values of ρ_s and ρ_p correlation coefficients between all drive time-based features and KSS values. Here, also the values are sorted with respect to $|\rho_s|$. All calculated correlation coefficients are significantly different from zero, i.e. $p\text{-value} < 0.05$. Moreover, it can be seen that for most of the features we have $|\rho_s| \geq |\rho_p|$. This confirms that the relationship between drive time-based features and KSS values is also to a larger extent non-linear. In addition, the features have different rankings with respect to the absolute values of ρ_s and ρ_p . As an example, with respect to $|\rho_p|$, *A/MOV* seems to be the best feature correlated with KSS with a linear association. However, T_{ro} is the most correlated with drowsiness based on $|\rho_s|$. Comparison of the correlation coefficient values in Tables 7.10 and 7.11 shows that the KSS input-based features are more correlated with KSS values. This is in agreement with the hypothesis that KSS values, in their best case, represent the driver's vigilance level within a short time interval prior to their collection. However, it should be mentioned that larger values of $|\rho_s|$ do not guarantee that the combination of these features with each other and using all of them simultaneously for driver drowsiness detection yields better results. This will be discussed in the next chapter.

7.5.2. Case 2: Correlation between features

In addition to the relevance of the extracted features to drowsiness and their informativity, it is also important to know whether they are redundant or not. The amount of correlation between features can be used to analyze the degree of redundancy. In general, highly correlated features are not desired, since they may carry the same information which is already provided by the other feature. The between-feature correlation analysis is also referred to as *intra-correlation*.

Table 7.11.: Sorted Spearman's rank correlation coefficient ρ_s and Pearson correlation coefficient ρ_p between all drive time-based features and KSS values ($N = 4021$). All p -values were smaller than 0.05.

Feature	ρ_s	ρ_p	Feature	ρ_s	ρ_p
T_{ro}	0.51	0.46	F	0.36	0.37
A/MOV	0.50	0.48	T_{50}	0.36	0.35
T_c	0.50	0.45	$PERCLOS$	-0.33	-0.34
A/MCV	0.47	0.40	AOV	-0.30	-0.28
MOV	-0.43	-0.42	$T_{cl,1}$	0.26	0.26
ACV	-0.43	-0.43	T_o	0.23	0.25
T_{90}	0.42	0.39	A	-0.23	-0.16
MCV	-0.41	-0.41	E	0.11	0.19
T_{80}	0.38	0.37	$T_{cl,2}$	0.06	0.21
T	0.37	0.35			

In the following, the correlation analysis for KSS input-based features and drive based-time features are studied based on ρ_p . Figures F.41 and F.42 show this analysis based on $|\rho_s|$.

KSS input-based features: Figure 7.20 shows $|\rho_p|$ calculated between KSS input-based features. The calculated $|\rho_p|$ for feature pairs with the red \times sign cannot be shown to be significantly different from zero, because for these feature pairs we have p -value > 0.05 . According to this figure, for some feature pairs such as (ACV, MCV) , (AOV, MOV) , $(T_c, A/MCV)$, $(T_{cl,1}, T)$, (T_{50}, T) , (T_{80}, T_{50}) , (T_{90}, T_{50}) and (T_{90}, T_{80}) , we have $|\rho_p| > 0.9$ with p -value < 0.05 . It is interesting that all highly linearly correlated features carry the same kind of information, i.e. they are all related either to the velocity or to the duration of a blink. For pairs, which are to a very small amount linearly associated with each other, namely $|\rho_p| < 0.1$, we found p -value > 0.05 (red \times sign in Figure 7.20). The very small amount of association, i.e. $0.1 < |\rho_p| < 0.2$, between pairs such as (F, A) , (T_{50}, A) and (F, MCV) is also comprehensible and reasonable, since each of these features has different underlying mechanism.

Drive time-based features: The absolute Pearson correlation coefficients between drive time-based features are also shown in Figure 7.21. Similar to Figure 7.20, in this figure, feature pairs with p -value > 0.05 are also shown with a red \times sign. Regardless of these pairs, other pairs such as (F, MOV) , (F, ACV) , (F, AOV) , (T, F) , (T_o, AOV) , (T_o, ACV) , $(T_{cl,2}, T_o)$, $(PERCLOS, E)$, (T_{50}, A) and (T_{80}, A) are all to a very small amount linearly correlated with each other, namely $|\rho_p| < 0.1$. As mentioned before, almost all of these features in pairs are based on different underlying mechanism. Interestingly, the drive time-based feature pairs with $|\rho_p| > 0.9$ are exactly the same as KSS input-based feature pairs with $|\rho_p| > 0.9$, except for the pair $(T_c, A/MCV)$. Thus, we conclude that the feature aggregation method has not affected the redundancy of features.

7.6. Eye blink feature's quality vs. sampling frequency

This section studies how sampling frequency of raw EOG signals affects the quality of the extracted features. This analysis is important for evaluating the features extracted from the data provided by the driver observation cameras rather than the EOG. In this context, Picot et al. (2009) studied the correlation between EOG and a high frame rate camera.

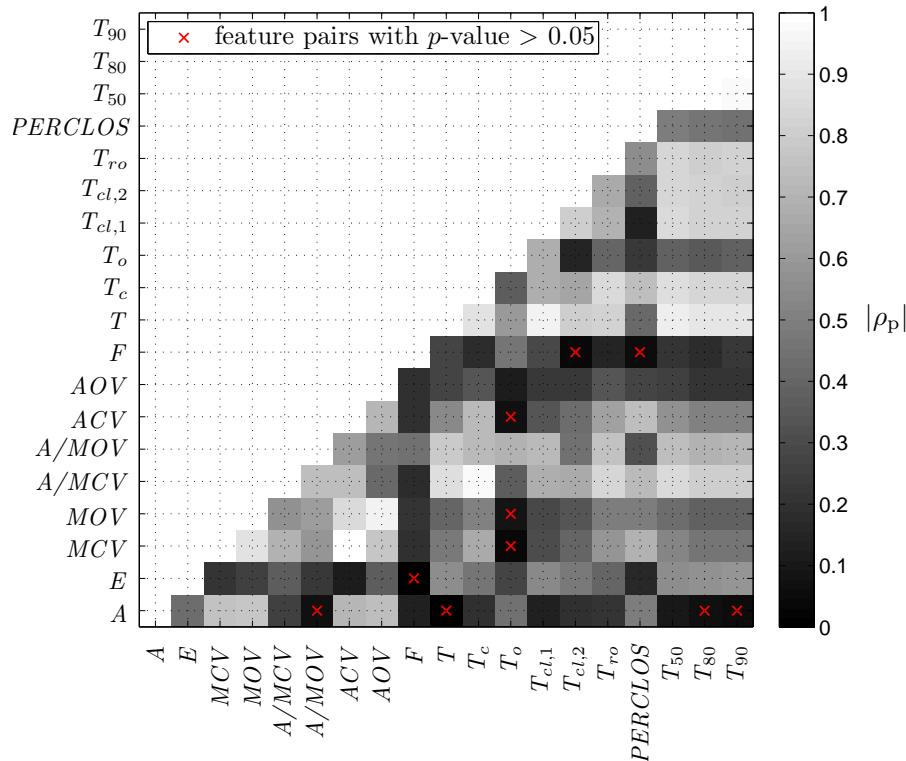


Figure 7.20.: Absolute values of Pearson correlation coefficient calculated between KSS input-based features

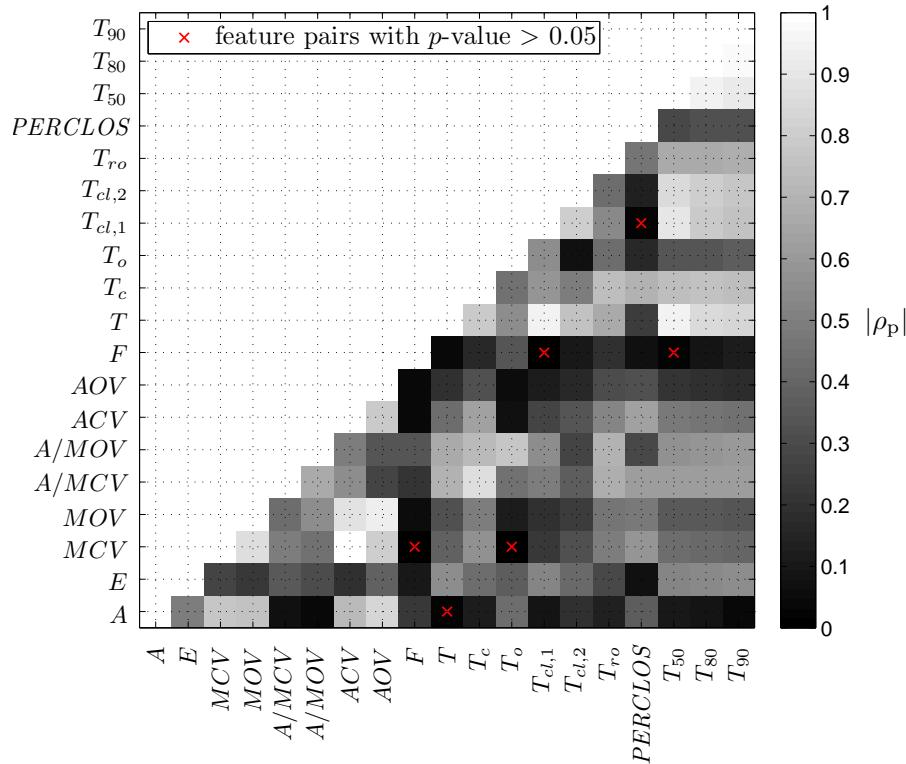


Figure 7.21.: Absolute values of the Pearson correlation coefficient calculated between drive time-based features

In the experiments conducted with the EOG measuring system, a camera cannot be used simultaneously, since the attached electrodes around the eyes disturb the image processing task of the camera. Therefore, we sampled EOG signals down to 40 Hz and 30 Hz which is comparable to the data of the driver observation cameras on the market. In fact, first, we artificially degraded the quality of the raw signals and then extracted all features as before.

Figures 7.22 and 7.23 show scatter plots of all 40 Hz and 30 Hz features versus those of the 50 Hz. Best least square linear fits are also plotted. According to the plots, for most of the features a smaller sampling frequency leads to smaller feature values. It is also clear that smaller sampling frequency results in peak amplitude loss. This fact shows itself in amplitude-based features to a larger extent, e.g. in MCV and MOV . Interestingly, T seems to be resistant to the reduction of the sampling frequency up to 30 Hz.

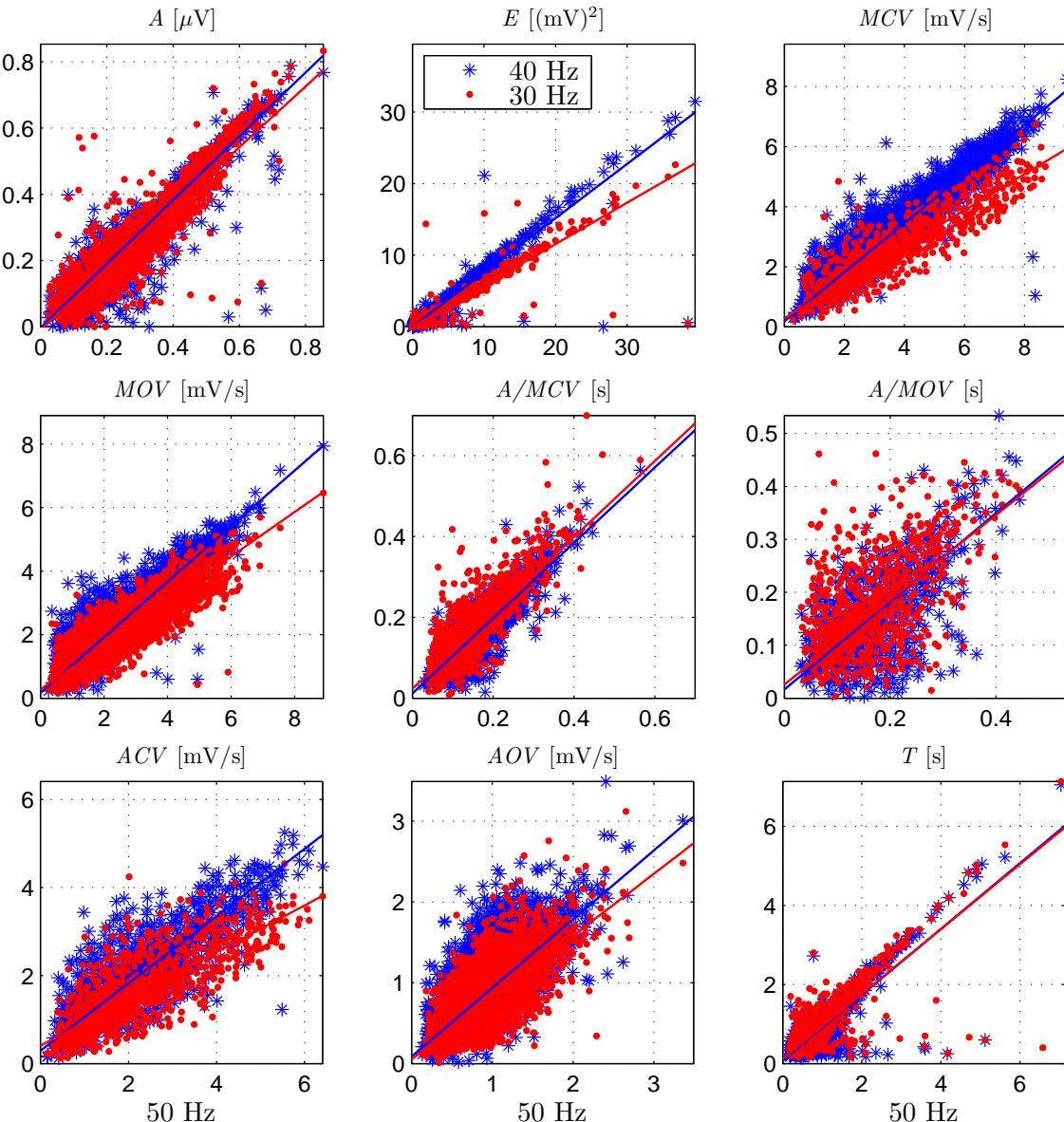


Figure 7.22.: scatter plot: comparison of 50-Hz features with 40- and 30-Hz ones for the first 12 subjects - part 1

Figure 7.24 compares AOV and MOV with respect to different sampling rates. As expected,

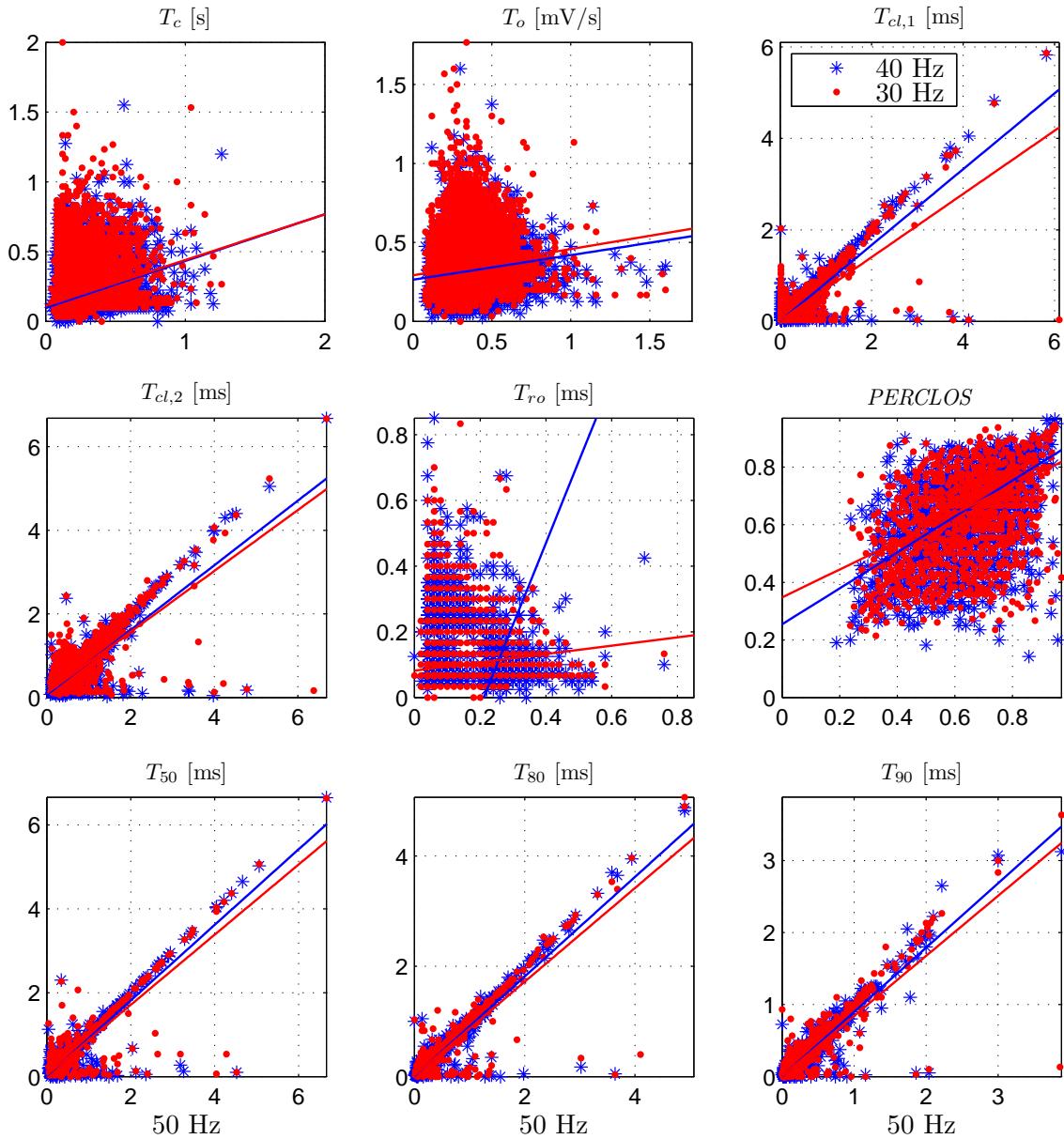


Figure 7.23.: scatter plot: comparison of 50-Hz features with 40- and 30-Hz ones for the first 12 subjects
- part 2

for smaller sampling frequencies, *MOV* values are closer to the *AOV* values which is the result of the peak amplitude loss.

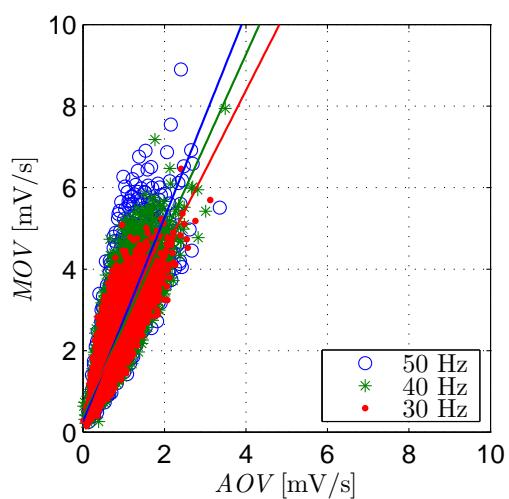


Figure 7.24.: Scatter plot: comparing AOV and MOV extracted based on 30, 40 and 50 Hz sampling rate. The lines indicate the best linear fits.

8. Driver state detection by machine learning methods

On the contrary to previous chapter, where features were explored as separate and independent source of information, in this chapter, we consider all extracted blink features of Chapter 7 together. To this end, we introduce machine learning and different state-of-the-art classifiers applied to the features for driver drowsiness detection. *Artificial neural network* (ANN), *support vector machine* (SVM) and *k*-nearest neighbors (*k*-NN) are the sophisticated classifiers used here. In order to evaluate the classification results, different metrics are also introduced. In addition, we consider different data division approaches to study the generalization aspects of the classifiers. Further, the issue of imbalanced data set for all classifiers is addressed by balancing the data sets. It is also investigated whether artificially balanced data sets can replace the expensive and demanding data collection during the awake phase. After comparing all classifiers with each other and suggesting an optimal classifier for driver drowsiness detection, the generalization of the data collected in the driving simulator to real road conditions is scrutinized and two new approaches are studied. Finally, we discuss approaches for feature dimension reduction in order to address issues of an in-vehicle warning system.

8.1. Introduction to machine learning

Machine learning methods, as their name mentions, are the algorithms and rules which *machines* like computers *learn* from some representative data with different attributes called features. By applying these learned rules to unseen data, it is possible to automatically classify the data based on its similarities with the seen data. In fact, machine learning methods are classification tools which divide the feature space into different regions by means of decision boundaries. The decision boundary is either a linear one (e.g. a line or a hyperplane) or non-linear depending on the complexity of the problem.

Here, the goal is to classify driver state based on extracted eye blink features. In fact, the idea is that it would be possible to learn drowsiness patterns in eye blink features and to use the learned patterns for predicting driver drowsiness. Eskandarian et al. (2007), Liang et al. (2007), Hu and Zheng (2009), Friedrichs and Yang (2010a,b) and Simon (2013) are examples of recent studies carried out in the field of driver state classification, i.e. classifying vigilance, drowsiness and attentiveness of the driver based on some physical or/and physiological features. A detailed review is provided by Dong et al. (2011).

In this work, the machine learning method used for driver state classification is called a *classifier*. Input of the classifier is the *feature matrix* $\mathbf{F} \in \mathbb{R}^{D \times N}$ containing N feature vectors \mathbf{x}_n , with $n \in \{1, \dots, N\}$ as follows

$$\mathbf{F} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \dots \quad \mathbf{x}_N]_{D \times N}, \quad (8.1)$$

where N is the number of samples. $\mathbf{x}_n \in \mathbb{R}^D$ represents the D -dimensional feature vector of the n -th sample, namely $\mathbf{x}_n = [x_{1,n} \ x_{2,n} \ \cdots \ x_{D,n}]^T$. Here, we have $D = 19$ features.

The output of the classifier is *class c* which corresponds to the driver state in this work. Therefore, depending on the number of the available states, we are dealing with a 2-class (binary), a 3-class or, in general, an m -class classification problem. The classes are defined in Section 8.1.1. Depending on the availability of the class membership for the samples of the feature matrix ahead to the classification step, two types of classification methods can be explored: *supervised* versus *unsupervised* classification. In this work, only supervised classification is studied.

8.1.1. Supervised classification

Supervised classification is defined as a classification problem with available information about class membership for each sample in the feature matrix. In our work, the KSS inputs collected during the experiments are the class labels. Therefore, for each sample n of the feature matrix \mathbf{F} , namely $\mathbf{x}_n \in \mathbb{R}^D$, there exists a corresponding discrete-valued class label c_n . The total data set \mathcal{S} is then expressed as follows

$$\mathcal{S} = \left\{ (\mathbf{x}_1, c_1), (\mathbf{x}_2, c_2), \dots, (\mathbf{x}_N, c_N) \right\}. \quad (8.2)$$

The task of supervised classification begins with the division of the data set into two sets called *training* and *test* sets. Based on the features and the corresponding classes belonging to the training set $\mathcal{S}_{\text{train}}$, the classifier is trained by learning rules. The complexity of these rules depends on the complexity of the classifier and the relationship between features and classes. Afterwards, the rules will be applied to the features of the test set $\mathcal{S}_{\text{test}}$, which is unknown to the classifier, in order to estimate the class of its samples. Finally, the performance of the classifier is evaluated by comparing the estimated class \hat{c} with the true class c of each sample.

A typical phenomenon, whose occurrence severely affects the performance of the classifier during the training step, is either *overfitting* or *underfitting* of the classifier to the training data set as shown in Figure 8.1. The former occurs, if the learned rules are highly adapted and fitted to the samples in the training set, such that all samples are classified correctly in the training set (Figure 8.1(a)). This leads to a zero training error rate which is defined as the ratio of wrongly classified samples during the training phase. However, this does not directly imply a low number of errors also for the test set. In fact, as soon as a new sample of the test set is applied to the classifier, it fails to classify the unseen data correctly. This is also called lack of *generalization* for the classifier. The reason is that the classifier is fitted to the noise rather than the data (Zamora, 2001). Therefore, a zero training error rate never guarantees a small test error rate. Overall, the ultimate goal is to construct a general classifier which not only classifies training data correctly, but also classifies new unseen data with the similar performance.

On the contrary to the overfitting, underfitting (Figure 8.1(b)) occurs, if the classifier is too simple and even the training error rate is high. In fact, the classifier does not fit to the underlying structure of the data. Obviously, it cannot be expected that such a classifier performs better on unseen data. Consequently, as Figure 8.1(c) shows, a compromise, as a trade-off between the mentioned phenomena, should be made. Therefore, sometimes even wrong classifications in the training phase are acceptable. Finally, in spite of the existing training error rate, the generalization aspect of the classifier improves, because the classification result of the unseen test set improves.

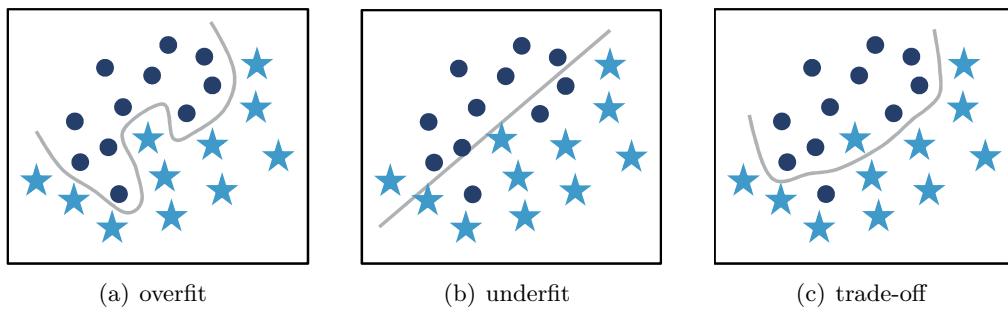


Figure 8.1.: Examples of three classification rules

KSS values as the classes

In this work, the KSS values collected during the experiments are used as available labels for the supervised classification task. As mentioned in Section 2.2, the self-estimation of drowsiness during driving is highly subjective. Moreover, for each KSS value, it is very probable that the subjects compare their current state with the previous ones for a better self-rating. Hence, depending on the preciseness of the first selected KSS value, there might be a bias shift on the other selected values until the end of the drive. Similar results concerning misjudgment of drowsiness after three hours of continuous monotonous daytime driving were reported by Schmidt et al. (2009). To this end, we suppressed the probable inaccuracy of the KSS values by grouping them together to form 2-class (binary) and 3-class (multi-class) problems as follows:

binary or 2-class:	awake					drowsy		
KSS values:	1	2	3	4	5	6	7	8
3-class:	awake					medium		drowsy

We call these classes {awake, drowsy} for the binary case and {awake, medium, drowsy} in the 3-class case.

The distribution of the classes for the binary classification is awake = 55% versus drowsy = 45% and for the 3-class case, awake = 41%, medium = 29% and drowsy = 30%, while considering drive time-based features. For the KSS input-based features, we also have awake = 46% versus drowsy = 54%. Due to small number of available samples for KSS input-based features, the 3-class case is not studied for them in this work. Figure 8.2 summarizes all class distributions and shows that in all cases, we have balanced class distributions.

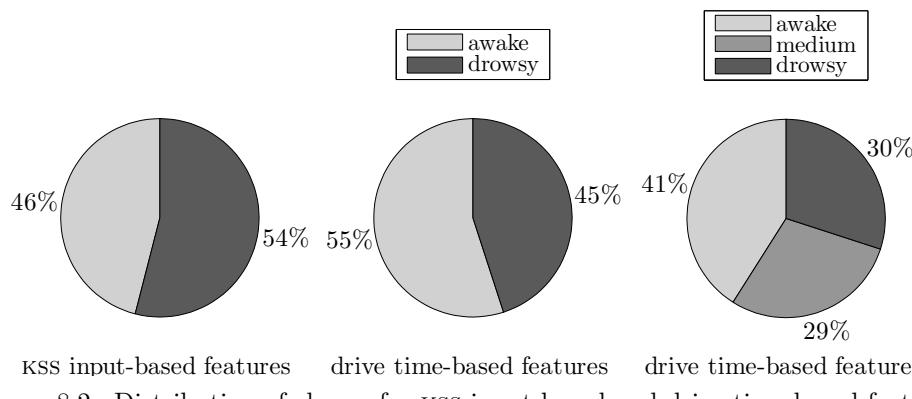


Figure 8.2.: Distribution of classes for KSS input-based and drive time-based features

8.1.2. Metrics for evaluating the performance of classifiers

As also mentioned in Section 5.4, the performance of a binary classifier can be evaluated by a confusion matrix shown in Table 8.1.

Table 8.1.: Confusion matrix of a binary classifier

		predicted state \hat{c}	
		awake	drowsy
given state c	state	True Positive (TP)	False Negative (FN)
	drowsy	False Positive (FP)	True Negative (TN)

For each class (awake or drowsy), a detection rate DR is calculated based on Table 8.1 and the values of TP, FP, TN and FN. The detection rates of classes are as follows

$$DR_{\text{awake}} = \frac{TP}{TP + FN} \quad (8.3)$$

$$DR_{\text{drowsy}} = \frac{TN}{TN + FP}. \quad (8.4)$$

They are also called *sensitivity* and *specificity* which are equivalent to probabilities $P(\hat{c} = \text{awake}|c = \text{awake})$ and $P(\hat{c} = \text{drowsy}|c = \text{drowsy})$. For the confusion matrix introduced in section 5.4, TN is not defined and as a result (8.4) cannot be calculated. Therefore, there, precision and recall were defined instead. Conventionally, the metrics in (8.3) and (8.4) are referred to as TPR and TNR, respectively. However, in this study, for a better readability of detection performance results, we use the term DR which refers to both of these metrics.

Similarly, the rates of wrongly classified samples with respect to the number of available samples in each class are

$$FNR = \frac{FN}{FN + TP} = 1 - DR_{\text{awake}} \implies P(\hat{c} = \text{drowsy}|c = \text{awake}) \quad (8.5)$$

$$FPR = \frac{FP}{FP + TN} = 1 - DR_{\text{drowsy}} \implies P(\hat{c} = \text{awake}|c = \text{drowsy}), \quad (8.6)$$

where FNR refers to the *false negative rate* or *miss rate* and FPR is the *false positive rate* and is also called the *false alarm rate*. The term 1–DR is used in this work to refers to both FNR and FPR regardless of the classes.

Moreover, we define an average DR (ADR) as the average of both DR values, namely

$$ADR = \frac{DR_{\text{awake}} + DR_{\text{drowsy}}}{2}. \quad (8.7)$$

Balanced accuracy is another notation of ADR.

In addition to the above metrics, *accuracy* (ACC) and its complementary, namely *error rate* (ER), are defined

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \implies P(\hat{c} = c) \quad (8.8)$$

$$ER = 1 - ACC \implies P(\hat{c} \neq c). \quad (8.9)$$

For a multi-class problem with m classes and confusion matrix $\mathbf{M} = [M_{i,j}]_{m \times m}$, these metrics

are calculated as follows (Kolo, 2011):

$$\text{DR}_i = \frac{M_{i,i}}{\sum_{j=1}^m M_{i,j}} \quad (8.10)$$

$$\text{ADR} = \frac{1}{m} \sum_{i=1}^m \text{DR}_i \quad (8.11)$$

$$\text{ACC} = \frac{\sum_{i=1}^m M_{i,i}}{\sum_{i=1}^m \sum_{j=1}^m M_{i,j}} \quad , \quad \text{ER} = 1 - \text{ACC} , \quad (8.12)$$

where i in DR_i refers to the i -th class.

8.1.3. Subject-dependent classification

As mentioned before, in supervised classification, data division into training and test is performed prior to the training of the classifier. Typically, training and test sets contain 80% and 20% of the total samples in the feature matrix, respectively. In addition, the feature matrix is split in a way that the class distributions of the training and test sets remain similar. To put it another way, as an example, if 65% of the total samples in the feature matrix belong to the awake class, 65% of the samples in the training and test sets also belong to this class. To keep the classification results independent of the samples selected for each set, we have randomly split the samples 100 times into 80% and 20% sets and call it *repeated random sub-sampling validation*. Due to the fact that samples of the training and test sets are randomly selected, both sets might contain samples of a specific subject. Obviously, if samples of a subject are statistically dependent or correlated, the classifier takes advantage of it and inflated classification results are expected. Therefore, due to dependency of the classifier on the subjects in this type of data division, the classification problem is called a *subject-dependent* (subj.-dep.) one. The final evaluation results are the average of DR and FDR values over all permutations.

8.1.4. Subject-independent classification

Generally, the goal of all driver assistance systems is to be able to warn all drivers, even those whose features are unknown to the warning system. Hence, another possibility for data division of training and test sets is to sort them according to the subjects unlike the subject-dependent case. In this type of data division, for the total number of s subjects, the classifier is trained by samples of $s - 1$ subjects and tested on the samples of the s -th subject. By repeating this procedure s times, all subjects appear one time in the test set. This method is similar to the *leave-one-out cross validation* (Duda et al., 2012). Since the constructed classifier model is fully independent of the subjects in the test set, we call it a *subject-independent* (subj.-indep.) classification.

In subject-independent classification, due to a varying number of samples in the test sets, misclassified samples of each subject might be differently penalized, e.g. 1 misclassification out of 5 samples corresponds to the rate of 20%, while 1 out of 10 corresponds to the rate of 10%. Thus, the DR and FDR values of different subjects cannot be directly compared with each other. On

this account, the TP, FP, TN and FN values of all test sets are summed up to the overall DR and FDR metrics.

At the first glance, it seems that subject-dependent data division is not needed to be taken into consideration in our application for the sake of generality of the classification results. However, in the following, it is explained why both subject-dependent and subject-independent data divisions are studied in this work.

We mentioned in Section 7.2 that there exist subjects, for whom the evolution of feature values due to drowsiness did not follow the overall trend found among the majority of participants. This highly individual behavior is linked to the intrinsic nature of the physiological measures. Clearly, if the data of these subjects is fed as the test set to the subject-independent classifier, the result will not be satisfactory. The reason is that the training data is not representative of the test data. Therefore, in order to cover individual differences to a very large extent, a data set with a very large number of participants is needed. This, however, is not a feasible solution in our application, because for the nighttime drowsy data collection, preparation of the participant and performing the task takes about 4 hours. Hence, subject-dependent data division can be considered as a rough estimation for the case that the training set is well enough representative of the test set.

Another motivation is related to future intelligent vehicles. If future vehicles make user adaptation possible, then each driver is a known user to the vehicle system. In other words, the driver creates a user profile for himself and the vehicle saves the settings of this user to his profile, e.g. seat setting, mirror settings, etc. Each time when the driver gets into the vehicle, he signs in and the vehicle adapts all saved settings. This idea can be extended to the warning system as well. Accordingly, the warning system is able to learn the user's behavior and benefits from his features for future use. This is similar to subject-dependent data division, because the user is not thoroughly unknown to the warning system.

8.1.5. Imbalanced class distributions

A very important issue, which severely impacts the performance of a classifier, is the distribution of the classes versus each other. In other words, it is crucial whether the class distributions are *balanced* or *imbalanced*. We showed in Section 8.1.1 that we have balanced class distributions. The reason that we study imbalanced data sets here is that it is not only a problem from the theoretical point of view for the classifiers, but also for the warning systems which aim to prevent car crashes due to driver drowsiness. He and Garcia (2009) reviewed and discussed the issue of imbalanced data which is summarized here.

A data set is referred to as imbalanced, if the proportion of one class to the other class is within the following orders: 100:1, 1000:1 or even larger. For 100:1, as an example, this means that the classes are distributed such that in one of them 100 times more samples are available than the other one. The class with the larger number of samples is called the *majority* class and the other class with the smaller number of samples is referred to as the *minority* class. According to Abdellaoui (2013), in the case of a multi-class problem, the class with the smallest number of samples is the minority class, while all other classes are considered as the majority classes.

The phenomenon of imbalanced data is mostly highlighted in medical applications such as distinguishing between ill and healthy patients as minority and majority classes, respectively. There, the wrong classification of an ill patient as a healthy one is much more crucial than the inverse

case. In our application, similarly, a drowsy driver should always be warned, otherwise a car crash is inevitable.

In general, for drowsiness detection in real life applications, most of the time, sleep-deprived or already drowsy subjects are preferred to participate in the experiment. Therefore, the collected data sets are dominant in the drowsy events. This is clearly due to two factors: cost and time. A subject, who is almost awake and fit for performing the driving task during the experiment, needs a longer time to feel drowsy which is not desirable. In spite of this fact, the state of an awake subject should be classified correctly to the same extent that a drowsy subject is classified as drowsy. In other words, not to warn an awake subject is also a part of the goals of driver state classification. Therefore, experiments solely conducted with less awake subjects suffer from imbalanced classes which also affects the classification task.

We mentioned in Chapter 4 that the driving simulator experiment (Section 4.3) and real road experiment (Section 4.2.1) covered the data collection of drowsy and awake phases, respectively. Therefore, in order to highlight the issue of imbalanced data set for less awake subjects, we have only considered the KSS input-based features collected in the driving simulator experiment and removed features of the real road drives from the feature matrix. This led to a feature matrix with 261 samples. We used 20% of this feature matrix for defining 100 test sets with balanced binary classes. Hence, each test set contained 52 samples (26 samples for each class). The remaining 209 samples of the training sets had the following imbalanced class distribution: awake = 24% and drowsy = 76%. In the end, this imbalanced training set is used to train the classifiers.

Although balanced class distributions are always desired, in most of the applications in real life, it is not possible to collect such data especially due to high cost. Therefore, one solution to tackle this problem is to *artificially* balance the data before applying a classifier to it based on some known methods. These approaches are applied to the imbalanced training set as well to balance the class distributions. In our case, it is the awake class which should be balanced artificially. If for the unseen awake data, the performance of the classifier trained by artificially balanced data is as good as that of the classifier trained by fully balanced data, then we can save time and cost by choosing sleep-deprived and drowsy subjects for our experiments and balance the class distributions afterwards artificially. Otherwise, conducting experiments for collecting the awake samples similar to the drowsy ones is inevitable despite taking a lot of time and effort. In Sections 8.2.3 and 8.3.7, this issue is studied.

In the following, two methods for artificially balancing the data set are introduced.

Random undersampling and oversampling

A very simple method for artificially balancing the class distribution is to randomly remove samples of the majority class in order to have balanced class distributions. Clearly, the random selection of samples to be removed should be repeated several times to guarantee that the classification results do not depend on the respective selections of samples of the majority class. As a very straight forward method, the random undersampling method has a disadvantage. Due to randomly removing samples of the majority class, definitely, some valuable information for the classifiers regarding this class might be removed which leads to poor classification results. Moreover, for highly imbalanced data, a large number of samples have to be removed. The undersampling method will be used as the solution for dealing with imbalanced data in Section 8.6.1.

Similar to the undersampling method, an oversampling is performed to deal with the imbalanced data issue, such that randomly selected samples of the minority class are duplicated in the data set. This method is motivated by the prevention of information loss of the undersampling approach. However, outlier samples could be selected to be added to the data and poor classification results are expected. Moreover, in general, an overfitting is inevitable, if multiple instances of a sample are added to the data. Some classifiers even require unique samples for training, like the classifier which will be explained in Section 8.3.

Synthetic minority oversampling technique

An alternative to the oversampling method explained before is the *synthetic minority oversampling technique* (SMOTE) introduced by Chawla et al. (2002). According to this algorithm, first N_{SMOTE} is needed to be defined which indicates the number of samples to be added to the minority class. Afterwards, for each sample x_i of the minority class (see Figure 8.3(a)), the k -nearest neighbors (k -NN) are looked for. Figure 8.3(b) shows these neighbors for $k = 3$. Finally, by randomly selecting one of the k -nearest neighbors \check{x}_i , a new synthetic sample x_{new} is added to the minority class by interpolation as follows

$$x_{\text{new}} = x_i + (\check{x}_i - x_i) \zeta, \quad (8.13)$$

where $\zeta \in [0, 1]$ is a random number. x_{new} is shown in Figure 8.3(c). In fact, new samples are generated with respect to the sample under investigation and some of its neighbors. In a multi-class problem with m classes the SMOTE is repeated $m - 1$ times to balance all classes.

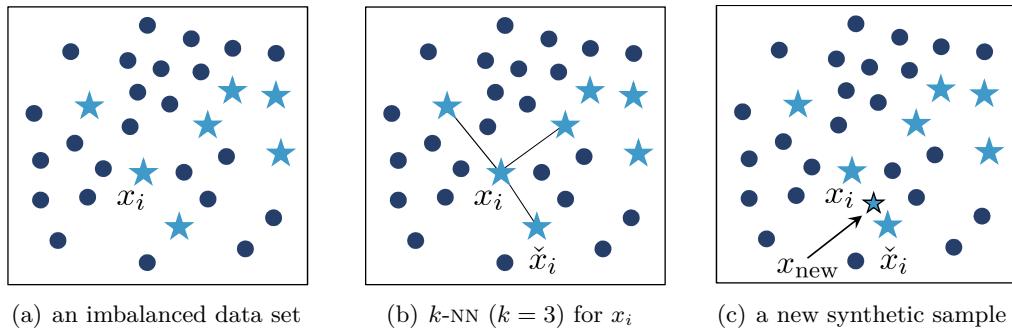


Figure 8.3.: Applying the SMOTE to an imbalanced data set

As explained before, adding new samples to the data set might lead to overlapping samples and consequently to the overfitting (He and Garcia, 2009). Moreover, depending on the classifier being used, the location of the new samples might affect the performance of the classifier. Therefore, it is recommended to apply *data cleaning* techniques afterwards. In fact, by removing some samples of the new data set, such cleaning techniques improve the separability of the available clusters and, in turn, the classifier performance.

A typical data cleaning method, which usually follows the SMOTE, is the *Tomek links* (Tomek, 1976). The Tomek links are the finite pair of samples to be removed and are defined as samples belonging to different classes, but located the closest to each other. In other words, the cleaning algorithm comprises finding and removing all Tomek link pairs, because either one of the samples of the pair is noise or both of them are close to the borderline (He and Garcia, 2009).

Mathematically, a pair (x_i, x_j) with x_i and x_j belonging to different classes and d_{x_i, x_j} as the distance between them is considered a Tomek link, if there exists no sample x_t of the opposite

class satisfying $d_{x_i,x_t} < d_{x_i,x_j}$ or $d_{x_j,x_t} < d_{x_i,x_j}$. The algorithm ends, if the nearest neighbor of each sample belongs to the same class. Figure 8.4 represents an example, where first, the SMOTE is applied to an imbalance data and then by removing the Tomek links, the clusters are easier to distinguish. This methods will be applied to our imbalanced data set in Section 8.2.3.

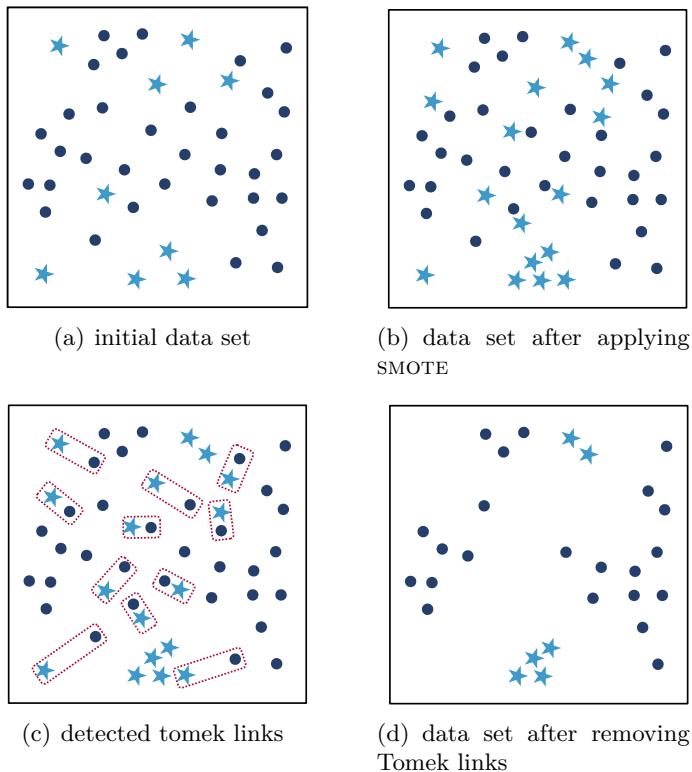


Figure 8.4.: Applying the SMOTE and the Tomek link cleaning technique to an imbalanced data

8.2. Artificial neural network classifier

The background theory of the artificial neural network classifier presented here is a summary taken from Jain et al. (1996), Zamora (2001), Uhlich (2006) and Duda et al. (2012) on this topic.

Inspired by the human nervous system and more specific the human brain, which is capable not only to learn and to generalize rules, but also to perform parallel tasks, the *artificial neural network* (ANN) also consists of elements called *neurons*. As a machine learning method, it is also capable to perform similar tasks. The first mathematical model of a simple neuron was originally introduced by McCulloch and Pitts in 1943 (McCulloch and Pitts, 1943). Eskandarian et al. (2007) and Friedrichs and Yang (2010a) also used this classifier for driver state classification.

Figure 8.5 shows the architecture of an ANN with three main layers: input layer, hidden layer and output layer. Since all connections are in one direction and there are also no feed-back loops between neurons, this architecture is called a *feed-forward* network. This network is memoryless, because the output to a specific input does not depend on the previous state of the network. Another variant of the network architecture discussed in Duda et al. (2012) is the *recurrent* or *feed-back* network.

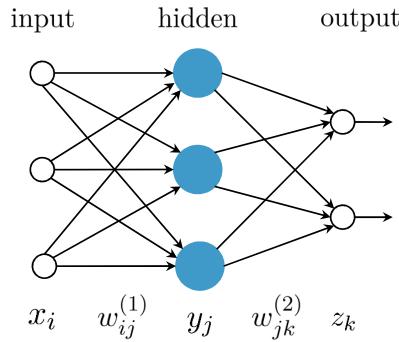


Figure 8.5.: Architecture of a feed-forward neural network with 3 inputs, 3 neurons in one hidden layer and 2 outputs

The number of features and the number of classes determine the number of inputs and outputs, respectively. Therefore, only the number of neurons and hidden layers are free parameters to be selected. Considering too many neurons or hidden layers leads to the overfitting of the network and consequently lack of generalization. On the contrary, too small numbers of them prevent the network from learning rules adequately. The impact of the number of neurons on the classification performance is discussed in the next sections.

8.2.1. Network's architecture

As shown in Figure 8.6, the input layer sends the input values x_i to the hidden layer without processing them. The hidden layer neurons calculate the weighted sum of the inputs called *net activation* (net). These calculated values are then fed to a non-linear *activation function* $f(\cdot)$ whose outputs y_j are the inputs to the next layer. Mathematically, we have

$$y_j = f(net_j) = f\left(\sum_{i=1}^D w_{ij}^{(1)} x_i + w_{0j}^{(1)}\right), \quad (8.14)$$

where index j refers to the j -th hidden neuron and $w_{ij}^{(1)}$ corresponds to the input-to-hidden neuron weights (see Figures 8.5 and 8.6).

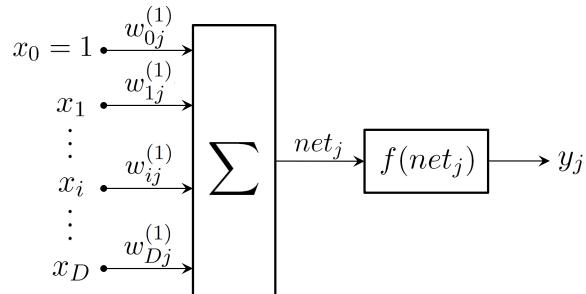


Figure 8.6.: Mathematical representation of the input-to-hidden layer of a network

Similarly, the output layer also calculates the net activation and the final result corresponds to the classifier output. Therefore, we have

$$z_k = f(net_k) = f\left(\sum_{j=1}^{N_h} w_{jk}^{(2)} y_j + w_{0k}^{(2)}\right). \quad (8.15)$$

In the above equation, index k denotes the k -th output unit (see Figure 8.5). N_h denotes the number of neurons in the hidden layer.

In the case of a multi-class classification problem with m classes, the class with the maximum value of z_k will be selected as the final classification result by the ANN classifier as follows

$$\hat{c} = \arg \max_{k=1, \dots, m} z_k. \quad (8.16)$$

The overall output of the introduced three-layer network in Figure 8.5 can be represented as

$$z_k = f\left(\sum_{j=0}^{N_h} w_{jk}^{(2)} f\left(\sum_{i=0}^D w_{ij}^{(1)} x_i\right)\right), \quad (8.17)$$

where z_k is given as the function of the input x_i by replacing (8.14) in (8.15) for y_j and $x_0 = y_0 = 1$. The generalization of (8.17) also allows considering other activation functions at the output layer in comparison to the hidden layers.

The non-linear activation function can be either a hard threshold function such as the *sign* function or a soft thresholding one such as the *sigmoid* function. The sigmoid function is popular for having the following properties as shown in Figure 8.7 for the tangent sigmoid $f(\text{net}) = \frac{2}{1+e^{-2\text{net}}}-1$:

- It is non-linear.
- It saturates which bounds the possible output values.
- It is continuous and differentiable.

It will be shown later that non-differentiable activation functions are, in general, not of interest. Since $f(\cdot)$ is a non-linear function, ANN is also a non-linear classifier and consequently can handle complex rules between features and classes.

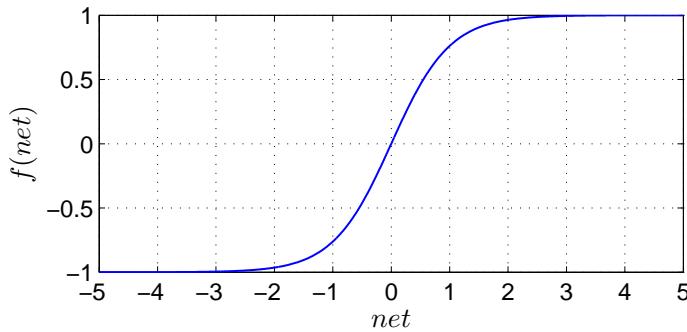


Figure 8.7.: Sigmoid activation function

8.2.2. Training of the network

After calculating the final outputs z of the network, which in our work correspond to the driver state, they are compared to the desired driver states c . Since these desired states are KSS values and are available, we are dealing with a supervised classification. Obviously, the goal is to minimize the difference between the estimated and the true states, i.e. the *error*, as shown in Figure 8.8. To this end, the *training error* J should be calculated in terms of mean squared

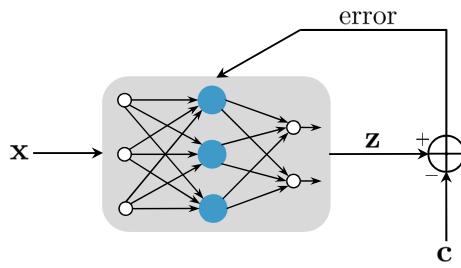


Figure 8.8.: Supervised classification by the ANN

error as follows

$$J = \frac{1}{2} \sum_{i=1}^m (c_i - z_i)^2 = \frac{1}{2} (\mathbf{c} - \mathbf{z})^T (\mathbf{c} - \mathbf{z}) \quad \text{given } \mathbf{x}. \quad (8.18)$$

The error minimization goal is achieved by updating the weights and calculating the new outputs several times. In fact, the network *learns* patterns of the training data, if weights are randomly initialized at the beginning of the algorithm and are updated iteratively based on an error minimization criterion. This kind of iterative learning is called the *back-propagation* algorithm and is performed by the *gradient descent* approach. In other words, the feed-forward property of the network sends the inputs from the input layer to the output layer and the back-propagation property updates the weights for achieving the most similar outputs to the desired ones. Therefore, mathematically, the partial derivative of the error with respect to the weights is calculated as follows

$$\Delta \mathbf{w} = -\eta \frac{\partial J}{\partial \mathbf{w}}. \quad (8.19)$$

The minus sign guarantees the reduction of error. η is called the *learning rate* and controls the relative change in weights for optimizing the error (Duda et al., 2012). If it is set too high, the final weights will be far from the optimal ones resulting in a poorly performing network. On the other hand, a very small value of η yields a very time-consuming training process. In Appendix G, it is explained how to train the network iteratively based on (8.19).

The iterative update rule for the τ -th iteration will be

$$\mathbf{w}(\tau + 1) = \mathbf{w}(\tau) + \Delta \mathbf{w}(\tau). \quad (8.20)$$

The initial values of weights are set randomly at the beginning. Depending on the number of samples available during the error minimization steps, different learning strategies are possible. In this work, all samples are considered at the same time which is called *batch learning*.

In addition to the above simple back-propagation method, there exists a variety of algorithms which differ in their speed for optimizing the error J and finding its global minimum instead of trapping in a local minimum. These algorithms are referred to as *second order methods* such as the *conjugate gradient*, *Newton* and *Levenberg-Marquardt*. They are all explained in detail in Duda et al. (2012). Unlike the gradient descent, these methods avoid a zigzag path towards the minimum. According to Zamora (2001), the conjugate gradient method is superior to the gradient descent optimization method in having a non-constant moving step towards the gradient in the negative direction. In other words, as long as the local or global minimum is not reached, the error J always decreases at each iteration (Bishop, 2006). In this work, we used the *scaled conjugate gradient* (Moller, 1993) due to its high optimization speed by requiring fewer iterations in optimizing the error J . Moreover, this method uses an approximation which avoids the full calculation of the Hessian in the conjugate gradient (Zamora, 2001).

Practical issues

Priddy and Keller (2005) and Duda et al. (2012) provided some practical issues for improving the training of the network which are summarized here.

- **Scaling the features:** Features, which are very different and far in their numerical values, will be handled differently by the network during the training phase, as if one feature would be more important than the other one. Duda et al. (2012) calls this phenomenon *non-uniform learning*. There are some approaches to solve this problem. In this work, in addition to the baselining step discussed in Section 7.1.3, we mapped the feature values into a $[-1, 1]$ range before feeding them to the ANN classifier as follows

$$\mathbf{x}_{\text{normalized}} = \frac{\mathbf{x} - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})} (\max_{\text{target}} - \min_{\text{target}}) + \min_{\text{target}}, \quad (8.21)$$

where $\max_{\text{target}} = 1$ and $\min_{\text{target}} = -1$. For other scaling functions see Priddy and Keller (2005).

- **Number of hidden layers:** Both of the mentioned references stated that one hidden layer is enough for learning any arbitrary function, given an enough number of neurons. As a result, in this work, we only use one hidden layer. Adding a second hidden layer did not improve the classification results.
- **Number of neurons:** There exist some rules of thumb for the selection of the number of neurons. In this work, however, we have selected it based on the classification performance for the training and test sets.

8.2.3. Classification results of subject-dependent data sets

This section discusses the classification results based on the ANN classifier for the subject-dependent data division. Moreover, results of different classification issues such as feature aggregation types and imbalanced data sets are discussed. Here, we have applied a feed-forward ANN classifier with scaled conjugate gradient back-propagation algorithm for adjusting the weights in one hidden layer. For hidden and output layers, the tangent sigmoid function similar to the example in Figure 8.7 and the linear transfer function were used, respectively. The results were generated using the Neural Network Toolbox™ R2010b of MATLAB.

In the following, it is shown that the most critical parameter of the ANN classifier, which directly impacts the classification performance, is the number of neurons N_h .

Results of KSS input-based features

In Section 7.1.1, we explained, how KSS input-based features are extracted. These features were applied to a binary subject-dependent ANN classifier, while considering different numbers of neurons. By considering 80% and 20% data division for the training and test sets, the sets contained 312 and 79 samples, respectively.

The ADR values for $N_h = 2, 3, 4, 5$ and 10 are shown in Figure 8.9 for both training and test sets. The error bars indicate the standard deviation with respect to all 100 permutations for selecting training and test sets. According to this figure, increasing N_h from 2 to 10 neurons improves the classification results of the training set as ADR increases. However, for the test set, the ADR

varies very slightly. Therefore, increasing the number of neurons does not improve the generality of the network. The corresponding confusion matrix for $N_h = 5$, as an example, is shown in Table 8.2. About 80% of the samples of each class are classified correctly. Overall, $3 \leq N_h \leq 5$ seems to be a sufficient number of neurons for the classification of the KSS input-based features. $N_h = 10$ increases the complexity of the network without improving the results.

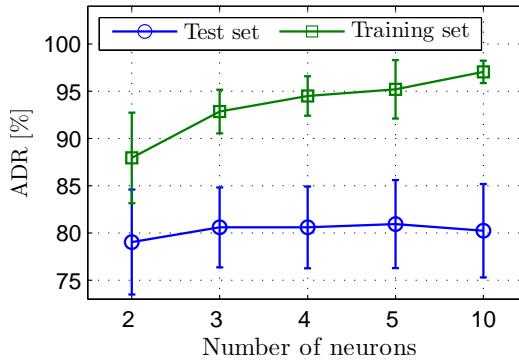


Figure 8.9.: ADR of the training and test sets of the binary subject-dependent ANN classifier for different numbers of neurons. Feature type: KSS input-based features. Bars refer to the standard deviation of permutations.

Table 8.2.: Confusion matrix of the binary subject-dependent ANN classifier ($N_h = 5$). Feature type: KSS input-based features

		predicted	
		driver state	drowsy
		awake	drowsy
given	awake	80.9%	19.1%
	drowsy	19.0%	81.0%

Results of the drive time-based features

Drive time-based features were defined in Section 7.1.2. As mentioned before, on the one hand, we expect to have poor classification results for these features due to the assumption that the KSS values do not change between two inputs. Moreover, we showed in previous chapter that the correlation values of these features with KSS values were lower in comparison to KSS input-based features. On the other hand, since a larger number of samples are available for the training task, the classifier learns the rules based on more available information which clearly improves the results.

Since, in total 4021 samples are available for drive time-based features as explained in Section 7.1.2, the training and test sets contain 3216 (80%) and 805 (20%) samples, respectively. Figure 8.10(a) depicts, how the variation of the number of neurons yields more precise classification performance. The comparison of the results of drive time-based features in Figure 8.10(a) with that of the KSS input-based ones shown in Figure 8.9 indicates that our assumption about constant KSS values between successive KSS inputs is justified. In fact, the larger number of available samples in the feature matrix counteracts the possible imprecision of the class labels. The reason is that a larger amount of information is provided to the classifier for learning the underlying relationship between KSS and feature values. Again, we emphasize that higher correlation values were found between KSS input-based features and KSS values. This, however,

did not lead to better classification results for them. Moreover, the ADR values shown in Figures 8.9 and 8.10(a) imply that for a larger number of samples in the feature matrix, a larger number of neurons is needed. In fact, the number of neurons is directly linked to the complexity of the training set. For KSS input-based features, we saw that $N_h \leq 5$ was a suitable value for the number of neurons. However, for drive time-based features at least 10 neurons are needed. From $N_h = 10$ to 25, despite better classification results for the training set, no improvement for the test set is evident. In other words, increasing the number of neurons leads the classifier to learn more complex rules which do not necessarily generalize the unseen data of the test set. The confusion matrix for the network with $N_h = 10$ is shown in Table 8.3. The awake class is detected much better in comparison to the case shown in Table 8.2 (87.9% vs. 80.9%) at a cost of 3% drop in the DR value of the drowsy class (78.2% vs. 81.0%).

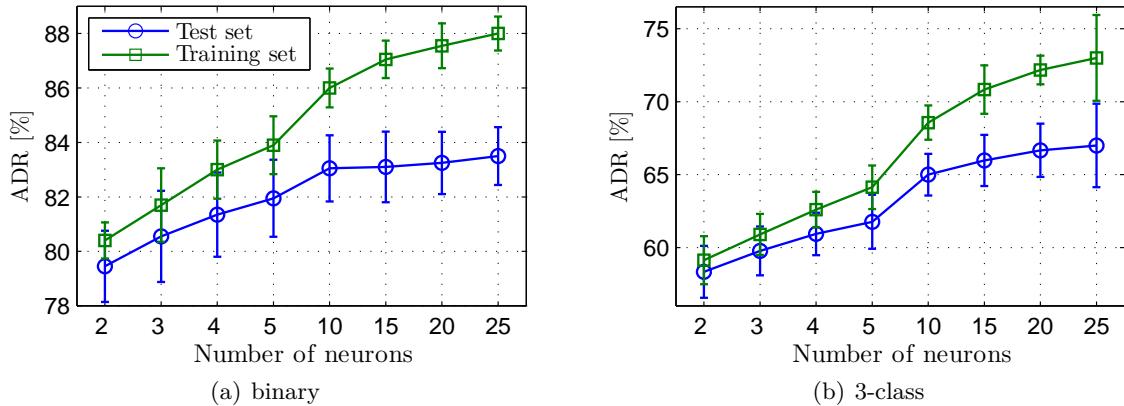


Figure 8.10.: ADR of the training and test sets of the binary and 3-class subject-dependent ANN classifier for different numbers of neurons. Feature type: drive time-based features. Bars refer to the standard deviation of permutations.

Table 8.3.: Confusion matrices of the subject-dependent ANN classifiers for 2-class ($N_h = 10$) and 3-class ($N_h = 20$) cases. Feature type: drive time-based features

		predicted			predicted			
		state	awake	drowsy	state	awake	medium	drowsy
given	state	awake	87.9%	12.1%	awake	84.2%	11.9%	3.9%
	awake	drowsy	21.8%	78.2%	medium	40.2%	42.8%	17.0%
	drowsy				drowsy	9.8%	17.2%	73.0%

Now, we consider a 3-class subject-dependent classification case with respect to the KSS boundaries explained in section 8.1.1 for dividing KSS values into 3 classes. ADR values for this case are shown in Figure 8.10(b). The values are not as good as those of the binary classification. Similar to Figure 8.10(a), initially, increasing the number of neurons gives better results (higher ADR values) whereas from $N_h = 20$, due to overfitting, the test set results have not improved. As mentioned before, in the 3-class classification problem, the classifier needs to learn more complex rules which consequently requires a larger number of neurons in comparison to the previously studied cases. Moreover, according to the confusion matrix for $N_h = 20$ shown in Table 8.3, as expected, the medium class is mixed up with the awake class most of the time (40.2%), while it is well distinguished from the drowsy class (17.0%). We showed in Figure 7.6 that the feature boxplots corresponding to different KSS values overlap each other to a large extent. As a result, the ANN classifier is also unable to find an acceptable rule for distinguishing them from each

other, especially for the awake versus medium class.

Results of the imbalanced KSS input-based features

In previous sections, the feature matrix under investigation was based on almost equally distributed classes, i.e. balanced data sets. Now, we consider the imbalanced data set introduced in Section 8.1.5 and resolve the consequences of imbalanced data issue. We mentioned that the training set was constructed by considering KSS input-based features collected in the driving simulator experiment and removing features of the real road drives from the initial feature matrix. The new feature matrix with 261 samples was divided to a balanced test set with 26 samples for each class and an imbalanced training set with 209 samples (24% awake and 76% drowsy samples).

We trained an ANN classifier with different numbers of neurons based on 100 imbalanced training sets. The corresponding ADR values are shown in Figure 8.11(a). Although we obtained ADR values of about 75% for test sets, this does not imply that both of the classes are similarly classified correctly. This figure only shows that increasing the number of neurons does not improve the ADR value for the test set and even deteriorates it. The confusion matrix for $N_h = 2$ is shown in the left part of Table 8.4. As expected, unlike the drowsy class, the awake class is classified close to random guessing due to lack of available information about it in the imbalanced training set. This is in agreement with the statement of He and Garcia (2009) regarding the drawbacks of an imbalanced data set.

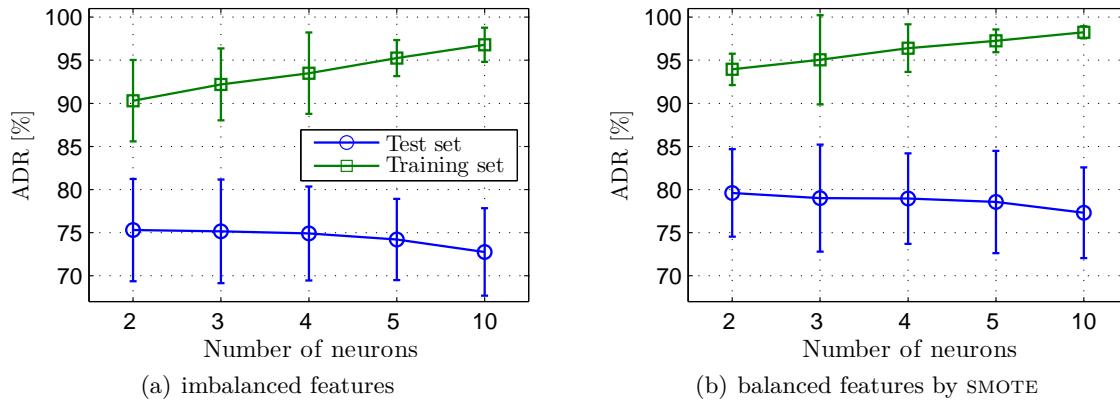


Figure 8.11.: ADR of the training and test sets of the binary subject-dependent ANN classifier for different numbers of neurons. Feature type: imbalanced and balanced by SMOTE KSS input-based features of driving simulator experiment. Bars refer to the standard deviation of permutations.

Table 8.4.: Confusion matrices of the binary subject-dependent ANN classifier for KSS input-based features of the driving simulator experiment. Left: imbalanced features ($N_h = 2$). Right: balanced features by SMOTE ($N_h = 2$)

		predicted					predicted		
		state	awake	drowsy	given	state	awake	drowsy	
given	awake	59.6%	40.4%	given	awake	75.3%	24.7%		
	drowsy	9.0%	91.0%		drowsy	16.1%	83.9%		

In Section 8.1.5, we introduced two known methods for dealing with imbalanced data sets. The

under-sampling method is not used here, because it results in a lower number of samples in the feature matrix which was shown to degrade the ANN classification results. Here, the SMOTE was applied by considering $k = 5$ neighbors for adding new samples to the training set. After cleaning the new training set based on the Tomek link approach, we obtained nearly balanced class distributions. The same balanced test sets as before were classified again based on the retrained network with balanced classes. The results of calculated ADR values are shown in Figure 8.11(b). Interestingly, increasing the number of neurons does not improve the results in this case either. The confusion matrix for $N_h = 2$ is shown in the right part of Table 8.4. The comparison of the both confusion matrices, i.e. before and after applying the SMOTE, indicates that SMOTE improved the classifier performance, such that the same awake samples of the test sets were classified correctly up to 75.3% (Table 8.4, right) instead of only 59.6% (Table 8.4, left). Clearly, this is obtained at a cost of 7% drop in the DR of the drowsy class (91.0% vs. 83.9%).

Now, the question is whether the retrained network based on the balanced data by SMOTE is able to classify unseen awake data correctly. In fact, the 75.3% DR for the awake class generated by SMOTE does not necessarily guarantee that true awake samples can be classified to the same extent correctly. The word *true* emphasizes that SMOTE adds *artificially* generated awake samples to the training set, not the measured true ones. On this account, we applied the removed samples of the real road experiment, which mainly belong to the awake class, as the test set to the trained network based on the SMOTE and driving simulator data. This clarifies how close the artificial awake samples are to the true ones. If we obtain high DR values, then we save time and cost by considering sleep-deprived and drowsy subjects for our experiments and balance the class distributions afterwards artificially.

Figure 8.12 shows the ADR values for the KSS input-based features of the real road experiment. It can be seen that increasing the number of neurons does not improve the classification results. It only leads to a better DR value for one of the classes at a cost of worse DR value for the other class as shown in Table 8.5 for two choices $N_h = 3$ and $N_h = 10$. The comparison of the confusion matrices clarifies why ADR values do not change in Figure 8.12. Increasing N_h from 2 to 10 improves the DR value of the drowsy class to the same extent that it degrades the DR value of the awake class, namely about 3%.

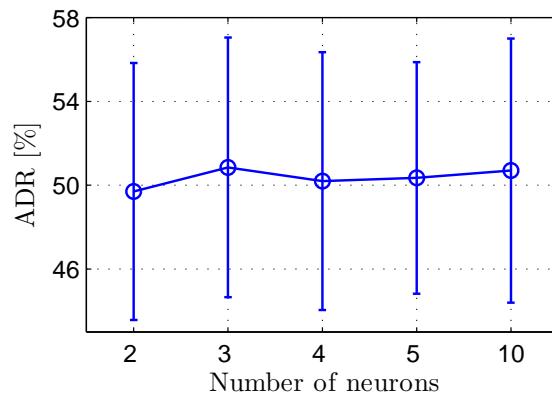


Figure 8.12.: ADR of the KSS input-based features for the real road experiment applied to the network trained based on the SMOTE. Bars refer to the standard deviation of permutations.

Regardless of the number of neurons, we conclude that adding artificial awake samples to the imbalanced training set of the driving simulator data improves the classification result of this class. Nevertheless, the retrained network results tend towards the awake class. The reason is that the classifier failed to classify most of the unseen drowsy samples correctly. In other words,

Table 8.5.: Confusion matrices of the binary subject-dependent ANN classifier for KSS input-based features of the real road experiment applied to the network trained based on the SMOTE. Left: $N_h = 3$. Right: $N_h = 10$.

		predicted		predicted	
		state	awake	drowsy	state
given	awake	63.5%	36.5%	awake	60.3%
	drowsy	61.8%	38.2%	drowsy	58.9%

		predicted		predicted	
		state	awake	drowsy	state
given	awake	60.3%	39.7%	awake	60.3%
	drowsy	58.9%	41.1%	drowsy	58.9%

the network is overfitted to the new samples added to the training set and does not generalize to the unseen data.

Based on the findings in this part, we suggest the collection of both awake and drowsy data during the experiment. This is due to the fact that artificially generated samples lead to overfitted classifiers with the tendency towards the minority class and lack of generalization.

8.2.4. Classification results of the subject-independent data sets

This section studies a binary subject-independent ANN classifier considering drive time-based features of all conducted experiments. As explained in Section 8.1.4, here, with a total of 43 subjects, the network was trained with 42 subjects and tested on the remaining subject who was excluded from the training set. The DR values are shown in Table 8.6 for $N_h = 2$. For other values of N_h , the network was overfitted. The small number of neurons for subject-independent data sets implies that the test set, namely the data of unseen subject, was to a large extent different from other 42 subjects. Consequently, a small number of neurons avoids overfitting and improves the generalization of the classifier.

Table 8.6.: Confusion matrix of the binary subject-independent ANN classifier for drive time-based features ($N_h = 2$)

		predicted		
		driver state	awake	drowsy
given	awake	80.8%	19.2%	
	drowsy	37.4%	62.6%	

Figure 8.13 compares these results with those of the subject-dependent case shown in Table 8.3. In this figure, DR refers to the DR of each class and $100\% - \text{DR}$ denotes the FNR and FPR as defined in (8.5) and (8.6). Clearly, the subject-dependent classifier performs better in the detection of both classes than the subject-independent one, because it has seen similar samples in the training set. As soon as totally new data is fed to the classifier, classification results degrade for at least 7%. This shows that individual properties of the subjects were not filtered out thoroughly during the baselining step discussed in Section 7.1.3. Moreover, since the DR of the drowsy class decreased more severely (about 15%, from 78.2% to 62.6%) in comparison to that of the awake class, we conclude that the effect of drowsiness represents different behaviors in different subjects. In other words, feature samples of each subject have certain characteristics which do not necessarily apply to all.

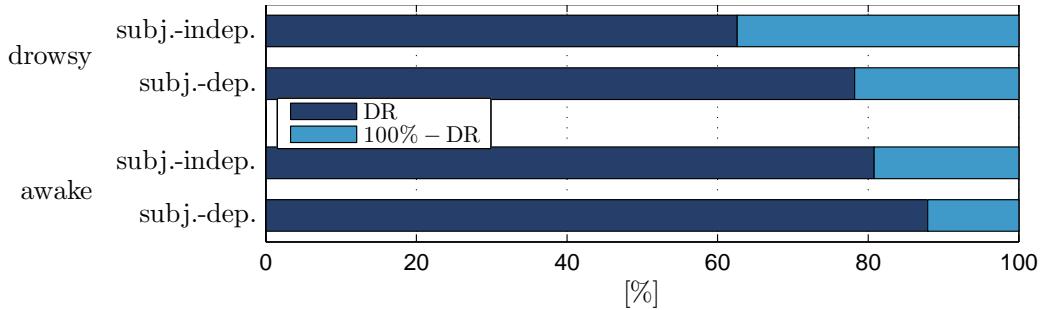


Figure 8.13.: Comparing confusion matrix of the binary subject-dependent ANN classifier with that of the subject-independent

8.3. Support vector machine classifier

The background theory of this section is based on a summary from Cristianini and Shawe-Taylor (2000), Schmieder (2009), Abe (2010), Duda et al. (2012) and Abdellaoui (2013).

Support vector machines (SVM) introduced by Vladimir Vapnik (Cortes and Vapnik, 1995) is another machine learning method with the capability of learning rules. In contrast to the ANN classifier which is sensitive to outliers and might suffer from getting trapped in multiple local minima, the SVM classifier is more robust to outliers with a unique solution (Olson and Delen, 2008; Yang, 2014). In addition, in applications with SVM, overfitting seems no to be a major issue (Olson and Delen, 2008). Hu and Zheng (2009) and Abdellaoui (2013) also used SVM classifier for driver state classification based on the blink features.

In the following sections, the basis of the SVM classifier is briefly introduced and it is explained how to train it by tuning the parameters.

8.3.1. Hard margin support vector machines

For a binary classification case with a linearly separable data set and class labels $y_i \in \{1, -1\}$ as shown in Figure 8.14, the decision function is defined by the hyperplane \mathcal{H} : $\mathbf{w}^T \mathbf{x} + b = 0$, where \mathbf{w} and \mathbf{x} refer to the weight and feature vectors, respectively. b denotes the bias. Therefore, the classes can be distinguished from each other by the following inequality

$$y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1. \quad (8.22)$$

Generally, the distance between sample \mathbf{x}_i of the training set and the separating hyperplane is called the *margin* γ_i as shown in Figure 8.14(a). It is defined as

$$\gamma_i = \frac{y_i (\mathbf{w}^T \mathbf{x}_i + b)}{\|\mathbf{w}\|}, \quad (8.23)$$

where $\|\mathbf{w}\|$ denotes the Euclidean norm of \mathbf{w} . The goal of support vector machine classifier is to find the hyperplane with the maximal margin for the training set among different possible separating hyperplanes as shown in Figures 8.14(a) and 8.14(b). This hyperplane is called the *optimal separating hyperplane* and is found by optimizing the value of γ_{\min} . γ_{\min} , which refers to the minimum distance between the separating hyperplane and all samples of the data set in

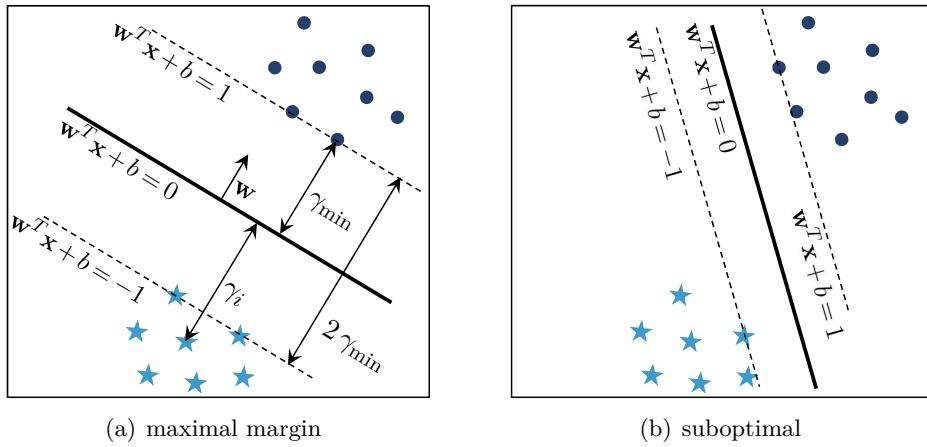


Figure 8.14.: Different separating hyperplanes

each class, is defined as follows

$$\gamma_{\min} = \frac{1}{\|\mathbf{w}\|}. \quad (8.24)$$

The largest value of γ_{\min} , called γ_{opt} , belongs to the optimal separating hyperplane and a classifier based on it is called *maximal margin classifier*. Thus, the goal is to find γ_{opt} , i.e.

$$\begin{aligned} \max_{\mathbf{w}, b} \gamma_{\min} &\implies \min_{\mathbf{w}, b} \|\mathbf{w}\| \\ \text{subject to } y_i (\mathbf{w}^T \mathbf{x}_i + b) &\geq 1 \quad i = 1, \dots, N \end{aligned} \quad (8.25)$$

which is equivalent to

$$\begin{aligned} \max_{\mathbf{w}, b} \gamma_{\min} &\implies \min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{subject to } y_i (\mathbf{w}^T \mathbf{x}_i + b) &\geq 1 \quad i = 1, \dots, N. \end{aligned} \quad (8.26)$$

In addition to (8.26), which is called the *primal form* of the optimization problem, there exists also an alternative *dual form* which in comparison to the primal form is much easier to solve. In fact, in the primal form, it is difficult to handle the inequality constraint.

The dual form of the optimization problem of (8.26) is based on the *Lagrangian function* \mathcal{L} as follows

$$\mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N \alpha_i (y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1), \quad (8.27)$$

where $\boldsymbol{\alpha} = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_N]^T$ is the non-negative *Lagrangian multipliers*. For optimizing the above equation, the extrema of $\mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha})$ should be calculated by differentiation with respect to b and \mathbf{w} . By substituting the differentiation results in (8.27), the dual problem is obtained as

follows

$$\begin{aligned} \max_{\alpha} \mathcal{L}_d(\alpha) &= \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{subject to } &\sum_{i=1}^N \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad i = 1, \dots, N \end{aligned} \quad (8.28)$$

where \mathcal{L}_d refers to the *dual Lagrangian function*. The above optimization is called *hard margin* SVM and is clearly independent of the weight vector \mathbf{w} .

A main drawback of the above optimization problem for the maximal margin classifier based on the hard margin is the requirement of a linearly separable data set. Clearly, such sets are not always the case in real life data collection. Moreover, since the goal of such a classifier is to classify the training data with no training error, it is inevitable to avoid overfitting. Consequently, low generalization ability is expected. This issue will be addressed in the next section.

8.3.2. Soft margin support vector machines

Since in real-world application not all data sets are linearly separable, the hard margin SVM should be revised before applying it to linearly inseparable data sets. A possible solution is, on the contrary to the previous approach, to tolerate misclassification of the training data to some extent as shown in Figure 8.15. It should be mentioned that it is the data set which might be linearly inseparable. The decision function is still a linear boundary.

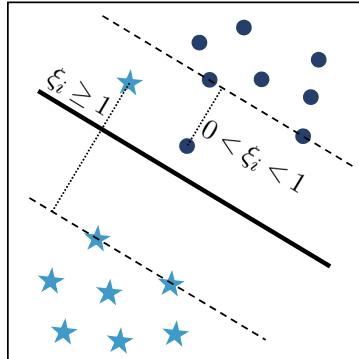


Figure 8.15.: Example of a linearly inseparable data

This goal is achieved by defining the *slack variables* $\xi_i \geq 0$ which modify (8.22) as follows

$$y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i. \quad (8.29)$$

In the case of $\xi_i = 0$, \mathbf{x}_i is classified correctly¹. For $0 < \xi_i < 1$, \mathbf{x}_i is classified correctly and is located within the selected margins, i.e. the selected margins are not the maximum ones (see Figure 8.15). However, for $\xi_i \geq 1$, \mathbf{x}_i is misclassified with respect to the selected optimal hyperplane (see Figure 8.15).

The hyperplane based on this approach is called the *soft margin* hyperplane. Accordingly, the classifier is called the soft margin SVM. The primal form of the optimization problem in (8.26)

¹ \mathbf{x}_i is not necessarily located on the boundary.

for finding this hyperplane is reformulated as follows

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} & \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{l} \sum_{i=1}^N \xi_i^l \\ \text{subject to } & y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i , \quad \xi_i \geq 0 , \quad i = 1, \dots, N \end{aligned} \quad (8.30)$$

where $\boldsymbol{\xi} = [\xi_1 \ \xi_2 \ \dots \ \xi_N]^T$ and the parameter C control the trade-off between minimizing the number of misclassified samples and maximizing the margin. In fact, parameter C is responsible for penalizing samples within the margin or those which are classified wrongly. l is usually 1 or 2 and accordingly the problem is called either L_1 -norm or L_2 -norm SVM. Although the soft margin solution does not suffer from the drawback of the hard margin in terms of linear separability of the data set, its performance depends on the choice of parameter C .

Similar to (8.26), there also exists an alternative dual form for (8.30) whose derivation is provided in Appendix H.2. For the L_1 -norm, we have

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \mathcal{L}_d(\boldsymbol{\alpha}) &= \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{subject to } & \sum_{i=1}^N \alpha_i y_i = 0 , \quad 0 \leq \alpha_i \leq C , \quad i = 1, \dots, N . \end{aligned} \quad (8.31)$$

In comparison to (8.28), now α_i has an upper bound in (8.31). Moreover, ξ_i is not directly involved in the dual form.

An important property in the resulting dual forms of the optimization problem for both hard and soft margin SVM ((8.28) and (8.31)) is that finding the optimal separating hyperplane never directly depends on the isolated values of the training data, but only on the inner product of the original feature vector \mathbf{x}_i^T and \mathbf{x}_j . It will be shown in the next section, how we benefit from this property.

8.3.3. Kernel trick

Although the previous section discussed the solution for handling linearly inseparable data sets, the resulting optimal hyperplane might still suffer from lack of generalization depending on the amount of non-linearity. We explained in Section 8.2 that ANN classifier benefits from non-linear transfer functions to deal with non-linearities. For SVM, however, another approach addresses this problem by analyzing the space \mathbb{R}^D on which the features are lying. The reason is that the more complex the original space is, the more difficult it is to learn the underlying patterns. This motivates *mapping* (\mapsto) the attributes to another space, called *feature space* \mathbb{F} , of a higher dimension by function Φ . As a result, the linear separation of a linearly inseparable data set becomes possible. Mathematically, the mapping is denoted as follows

$$\begin{aligned} \Phi: \mathbb{R}^D &\mapsto \mathbb{F} \\ \mathbf{x} &\mapsto \Phi(\mathbf{x}) . \end{aligned} \quad (8.32)$$

Figure 8.16 shows an example of such a mapping, where increasing the dimension of the original space with linearly inseparable data set leads to a linear classification problem in the feature space.

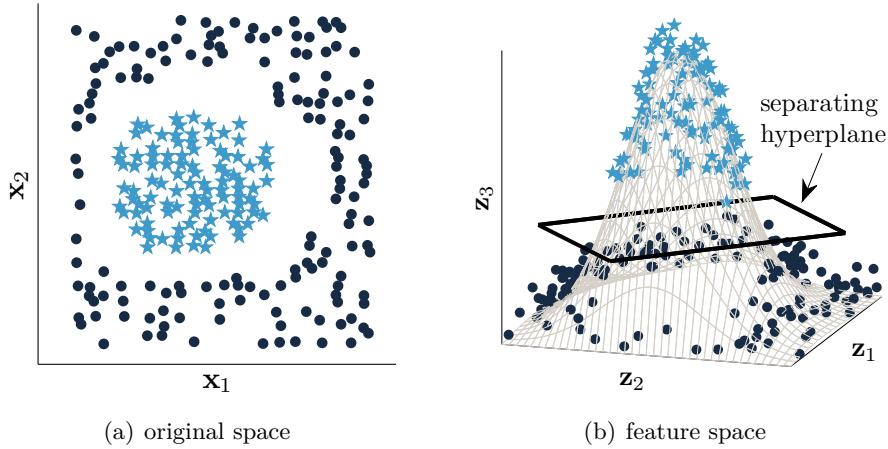


Figure 8.16.: An example of feature mapping for a linearly inseparable data set

Since after mapping, $\Phi(\mathbf{x})$ contains linearly separable values, the linear decision function introduced in (8.22) is reformulated as

$$\mathcal{H}: \mathbf{w}^T \Phi(\mathbf{x}_i) + b. \quad (8.33)$$

Accordingly, in (8.28) and (8.31), the inner product of $\mathbf{x}_i^T \mathbf{x}_j$ is replaced by the *kernel function* $K(\cdot)$, namely

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j). \quad (8.34)$$

The advantage of the kernel function is that it allows the calculation of the inner product of $\Phi(\mathbf{x}_i)^T$ and $\Phi(\mathbf{x}_j)$ without explicitly calculating the mapped values. Moreover, the dimension of the feature space does not play any role in the calculation of the kernel function. Consequently, even a feature space with a very large dimension does not increase the computational complexity of the classification problem. Other properties of the kernels are discussed in Cristianini and Shawe-Taylor (2000) and Schmieder (2009).

Some well-known kernel functions are

- *linear*: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$
- *polynomial*: $K(\mathbf{x}_i, \mathbf{x}_j) = (a + \mathbf{x}_i^T \mathbf{x}_j)^d$, $d \in \mathbb{N}$, $a \geq 0$
- *radial basis function (RBF)*: $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}$, $\gamma > 0$
- *sigmoid*: $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\kappa \mathbf{x}_i^T \mathbf{x}_j + r)$, $\kappa > 0$, $r < 0$

As it is shown, some kernel functions have also a parameter which should be tuned in addition to the parameter C of the SVM classifier. For example, the parameter γ of the RBF kernel, is responsible for controlling under and overfitting during the training phase (Asa and Weston, 2010).

In general, there is no known method which determines the type of the kernel function for a specific application. Thus, depending on the characteristic of the data set, different kernels might be appropriate or inappropriate. Since RBF kernel has only one parameter to be optimized, it is usually the first choice (Chang and Lin, 2011).

8.3.4. Model construction

We mentioned in the previous section that the training phase of the SVM classifier involves optimization of two types of parameters: the parameter C for controlling the trade-off between

the training error and the margin between classes and the kernel function parameter(s). Here, we choose the RBF kernel which means the parameter pair (C, γ) must be optimized. A common method for achieving this goal is applying the *grid search* method combined with the *cross validation* as explained in the following.

Grid search

In general, the grid search method comprises searching for the optimal parameters guided by a performance metric. For each parameter pair, a performance metric such as the accuracy is calculated and the pair with the highest accuracy will be selected for constructing the SVM model. It should be noted that here, we use accuracy as the performance metric to guide the grid search, because its classification results outperformed those of other metrics during the training phase. By choosing the RBF kernel, first, the range of C and γ needs to be defined. Since finding the best parameter pair is very time-consuming, the suggestion of Hsu et al. (2003) is adapted. According to that, first a *coarse* grid search is applied to roughly find the best values of C and γ , namely (C_0, γ_0) , such that

$$(C_0, \gamma_0) = \arg \max_{C, \gamma} \text{ACC}_{\text{train}}(\text{SVM}(C, \gamma)) \quad (8.35)$$

$$C = 2^{x_C}, \quad x_C \in \{-5, -3, \dots, 15\} \quad (8.36)$$

$$\gamma = 2^{x_\gamma}, \quad x_\gamma \in \{-15, -13, \dots, 3\}, \quad (8.37)$$

where $\text{SVM}(C, \gamma)$ denotes the SVM model constructed using parameters C and γ . $\text{ACC}_{\text{train}}$ refers to the accuracy during the training phase. We suppose that (C_0, γ_0) are calculated based on (x_{C_0}, x_{γ_0}) . Afterwards, a *fine* grid search is performed around (C_0, γ_0) to determine the optimal parameter pair $(C_{\text{opt}}, \gamma_{\text{opt}})$ based on a new range of values for x_C and x_γ as follows

$$(C_{\text{opt}}, \gamma_{\text{opt}}) = \arg \max_{C, \gamma} \text{ACC}_{\text{train}}(\text{SVM}(C, \gamma)) \quad (8.38)$$

$$x_C \in \{x_{C_0} - 2, x_{C_0} - 1.75, \dots, x_{C_0} + 1.75, x_C + 2\} \quad (8.39)$$

$$x_\gamma \in \{x_{\gamma_0} - 2, x_{\gamma_0} - 1.75, \dots, x_{\gamma_0} + 1.75, x_{\gamma_0} + 2\}. \quad (8.40)$$

Unlike the coarse search, where the step sizes of x_C and x_γ are set to 2, in fine search, they are set to 0.25. Figure 8.17 shows the results of the coarse and fine grid search with the highest accuracy of 83.5% and 84.0%, respectively. The selected parameters are $(C_0, \gamma_0) = (32, 2)$ and $(C_{\text{opt}}, \gamma_{\text{opt}}) = (16, 2)$.

Cross validation

Cross validation is a method for avoiding overfitting during the training phase. As mentioned in Section 8.1.1, prior to applying any classifier to the data set, a training and a test set are constructed. The cross validation method splits the training set into two sets for a second time. The first set of it is only used for the training, while the second one, called *validation set* $\mathcal{S}_{\text{validation}}$, is considered as the test set (see Figure 8.18). Repeating this division step j times is called *j-fold cross validation* which randomly divides the N_{train} samples of $\mathcal{S}_{\text{train}}$ into j subsets with the length of $\frac{N_{\text{train}}}{j}$ per subset. Each time, one of the j sets is used as $\mathcal{S}_{\text{validation}}$ and the remaining $j - 1$ sets are combined together as the new training set. Clearly, by repeating this procedure j -times, all sets appear one time as the validation set. A performance metric on $\mathcal{S}_{\text{validation}}$ is

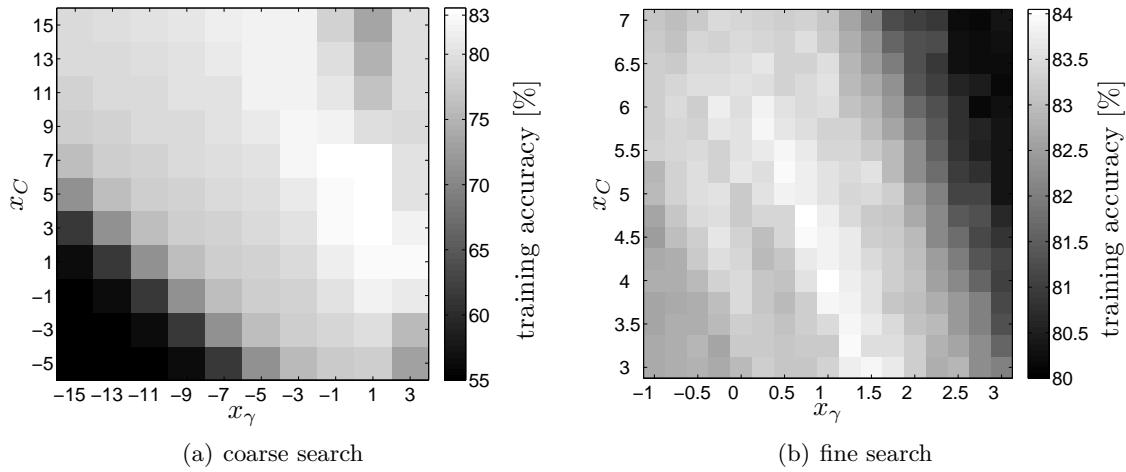


Figure 8.17.: An example of the grid search for finding (C_0, γ_0) and $(C_{\text{opt}}, \gamma_{\text{opt}})$

calculated each time. The overall performance of the training phase is the average over all j calculated performance metrics. Figure 8.18 pictorially shows this method. If additionally the samples of different classes are also equally distributed in the training and test sets, then the method is called *j -fold stratified cross validation*. After constructing the final model, it will be then applied to the initial test set which is totally new to the classifier.

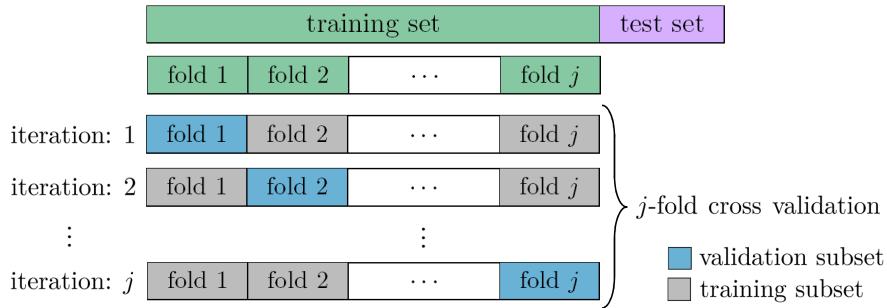


Figure 8.18.: j -fold cross validation method

8.3.5. Multi-class classification approaches

In order to extend the application of the SVM classifier for covering multi-class cases, different approaches exist which are reviewed in Abe (2010). Here, we only explain *One-Against-One* and *One-Against-All* strategies. Both of these strategies decompose the original problem into multiple binary sets and then apply the introduced SVM classifier to them as explained in the following.

One-Against-All

This approach decomposes the original m -class problem into m binary classification problems where sample \mathbf{x}_i belongs either to class $J \in \{1, 2, \dots, m\}$ or does not belong to this class. Clearly, a sample, which does not belong to class J , is the member of one of the other $m - 1$ classes. In order to cover all m classes, the decision function should be calculated m times.

According to (8.33), for the J -th decision function, which classifies the J -th class, we have

$$\mathbf{w}_J^T \Phi(\mathbf{x}_i) + b_J \begin{cases} \geq 1 & \mathbf{x}_i \text{ belongs to class } J \\ \leq -1 & \mathbf{x}_i \text{ belongs to the remaining classes} \end{cases},$$

where $\mathbf{w}_J^T \Phi(\mathbf{x}_i) + b_J = 0$ is the optimal separating hyperplane. The above decision function is a discrete one, since only its sign plays a role in the classification. A shortcoming of such a decision function is that a sample might be unclassifiable as shown in Figure 8.19(a) with shaded areas. In this figure, a sample is classified as belonging to class J , if $\mathbf{w}_J^T \Phi(\mathbf{x}_i) + b_J > 0$ (shown with an arrow). Clearly, a sample is unclassifiable,

- if it satisfies $\mathbf{w}_J^T \Phi(\mathbf{x}_i) + b_J > 0$ for different classes or
- if it does not satisfy $\mathbf{w}_J^T \Phi(\mathbf{x}_i) + b_J > 0$ for any class J .

Therefore, instead of a discrete decision function, a continuous one is used and the final predicted class is the one which maximizes the decision function as follow

$$\arg \max_{J=1, \dots, m} (\mathbf{w}_J^T \Phi(\mathbf{x}_i) + b_J). \quad (8.41)$$

In addition to the mentioned problem, another problem of the One-Against-All approach is that the distribution of the classes in the training set is imbalanced. This is due to classifying samples of one class against samples of all other classes. As an example, for five classes each with equal numbers of samples, this approach leads to the class distribution of 20% versus 80% in the training set. In Section 8.1.5, we discussed this issue.

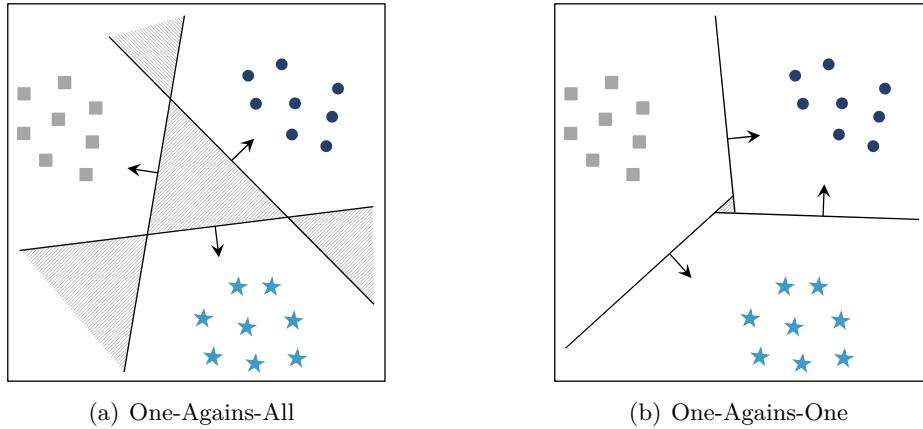


Figure 8.19.: An example of a 3-class classification with shaded areas as the unclassifiable. The arrows show the positive sides of the hyperplanes. Decision functions for the One-Against-All approach: $\mathcal{H}: \mathbf{w}_J^T \Phi(\mathbf{x}) + b_J = 0, J = 1, 2, 3$ and for the One-Against-One approach: $\mathcal{H}: \mathbf{w}_{IJ}^T \Phi(\mathbf{x}) + b_{IJ} = 0, I = 1, 2, 3, J = 1, 2, 3$ and $I \neq J$.

One-Against-One

On the contrary to the previous approach, the One-Against-One approach decomposes the original multi-class problem into $K = \frac{m(m-1)}{2}$ binary cases, where m refers to the number of available classes. By applying the SVM classifier to these binary problems, a sample \mathbf{x}_i will be then classified for K times based on K decision functions as either the member of class I or class J ,

where $I \neq J$. Consequently, this method performs the training phase with a fewer number of samples, namely only with those belonging to the classes under investigation. On the contrary, the One-Against-All method considers all samples together. By applying the conventional SVM classifier to the binary classes, we have

$$\mathbf{w}_{IJ}^T \Phi(\mathbf{x}_i) + b_{IJ} \begin{cases} \geq 1 & \mathbf{x}_i \text{ belongs to class } I \\ \leq -1 & \mathbf{x}_i \text{ belongs to class } J \end{cases},$$

where $I = 1, 2, \dots, m$, $J = 1, 2, \dots, m$, $I \neq J$ and $\mathbf{w}_{IJ}^T \Phi(\mathbf{x}_i) + b_{IJ} = 0$ is the optimal separating hyperplane. The final predicted class of sample \mathbf{x}_i corresponds to the class with the maximum number of votes after K times classification as follows

$$\arg \max_{I=1, \dots, m} \sum_{J=1, J \neq I}^m \operatorname{sgn}(\mathbf{w}_{IJ}^T \Phi(\mathbf{x}_i) + b_{IJ}). \quad (8.42)$$

In fact, a sample \mathbf{x}_i is classified as belonging to the I -th class, if the above equation is equal to $m - 1$ for the I -th class and a value smaller than $m - 1$ for the other class. If for none of the classes the value $m - 1$ is achieved, then \mathbf{x}_i is unclassifiable, because multiple classes satisfy (8.42).

Figure 8.19(b) shows an example of an unclassifiable area for this method with shaded area. In this figure, $\mathbf{w}_{IJ}^T \Phi(\mathbf{x}_i) + b_{IJ} > 0$ leads to the classification of \mathbf{x}_i as belonging to class I (shown with arrows) and otherwise as belonging to class J . According to this figure, the advantage of this method over the One-Against-All method in Figure 8.19(a) is that the unclassifiable area for the current method is much smaller. Therefore, in this work, this approach is applied.

8.3.6. Dealing with imbalanced data

As mentioned in Section 8.1.5, an imbalanced data set degrades the classifier performance. For the SVM classifier with imbalanced data, the optimal hyperplane tends more towards the minority class. The reason is that the more it gets closer to the samples of the minority class, the larger is the number of correctly classified samples of the majority class in comparison to the number of misclassified samples of the minority class. In other words, the classifier prefers to classify a large number of samples of the majority class correctly without violating the margins instead of violating the margins to correctly classify few samples of the minority class. As a result, most of the unseen samples of the test set will automatically be classified as belonging to the majority class due to the shifted optimal separating hyperplane.

We explained in Section 8.3.2 that the soft margin SVM tolerates wrong classifications by applying the parameter C in (8.30) which is also referred to as the *misclassification cost*. The second term in (8.30) penalizes the errors of both classes equally, which is not desired, if the available data set is imbalanced. Therefore, the following solution proposed by Veropoulos et al. (1999) is used for the L_1 -norm SVM which considers different values of C for each class, namely

$$C \sum_{i=1}^N \xi_i \longrightarrow C_+ \sum_{i \in N_+} \xi_i + C_- \sum_{i \in N_-} \xi_i. \quad (8.43)$$

In the above equation, N_+ and N_- refer to the number of samples in the majority and minority classes, respectively. Similarly, C_+ and C_- refer to different misclassification costs for each class.

Accordingly, the upper bound of α_i in (8.31) also becomes

$$0 \leq \alpha_{i+} \leq C_+ , \quad \mathbf{x}_i \in \mathcal{S}_+ \quad (8.44)$$

$$0 \leq \alpha_{i-} \leq C_- , \quad \mathbf{x}_i \in \mathcal{S}_- . \quad (8.45)$$

To reduce the negative effect of an imbalanced data set, misclassified samples of the minority class must be penalized to a larger extent than those of the majority class, i.e. $C_+ < C_-$. Akbani et al. (2004) empirically found good results by selecting the following ratio between C_+ and C_-

$$\frac{C_-}{C_+} = \frac{N_+}{N_-} \implies C_- = C_+ \frac{N_+}{N_-} . \quad (8.46)$$

Finally, this leads to

$$C_- = C \quad (8.47)$$

$$C_+ = C \frac{N_-}{N_+} . \quad (8.48)$$

However, Schmieder (2009) suggested following values

$$C_- = \frac{C}{2N_-} \quad (8.49)$$

$$C_+ = \frac{C}{2N_+} . \quad (8.50)$$

In this work, we use the ratio in (8.46).

8.3.7. Classification results of subject-dependent data sets

In this work, the soft margin SVM classifier was applied using the LIBSVM library (Chang and Lin, 2011). As mentioned before, this classifier has two parameters to be tuned: the kernel parameter(s) and the classification error weight C for controlling the trade-off between the training error and the margin between classes. The RBF kernel was selected here because of its good classification results within the shortest simulation runtime. The γ parameter of the RBF and C were optimized by the grid search method and the 5-fold cross-validation as explained in Section 8.3.4. The range of C and γ were also defined according to (8.36) and (8.37). The performance metric for guiding the search algorithm was the accuracy as defined in (8.8), because its classification results outperformed those of other metrics during the training phase.

Results of the KSS input-based features

As mentioned before, the balanced test and training sets were defined by 100 permutations. Due to this fact, 100 binary SVM models were trained with the parameters shown in Figure 8.20 regarding the KSS input-based features with balanced class distributions. The corresponding training and test accuracies are also shown in this figure. The confusion matrix, which is calculated based on the average over all permutations, is also shown in Table 8.7. It can be seen that both classes are classified with very close DR values. The comparison of these results with those of the ANN classifier in Table 8.2 indicates that the SVM classifier has achieved slightly better DR values, although the differences are only about 4%.

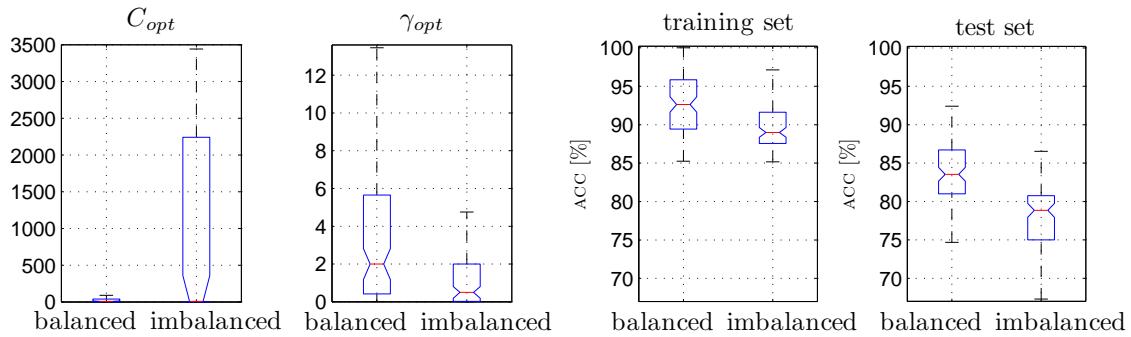


Figure 8.20.: Boxplot of C , γ , training and test accuracies for the balanced and imbalanced 2-class subject-dependent classification with the SVM for all 100 permutations. Feature type: kss input-based features

Table 8.7.: Confusion matrix of the binary subject-dependent SVM classifier. Feature type: kss input-based features

		predicted	
		driver state	awake drowsy
given	awake	84.2%	15.8%
	drowsy	16.4%	83.6%

Results of the drive time-based features

Similar to the ANN classifier, 2-class and 3-class SVM classifiers with drive time-based features were trained. We expect to achieve better results in comparison to the results of the kss input-based features due to having a feature matrix with a larger number of samples and a larger amount of information.

Figure 8.21 shows the trained parameters C and γ for both subject-dependent SVM classifiers. Comparing these parameters with those of the kss input-based features does not reveal a large difference between them. Moreover, the parameters of the 2- and 3-class cases are also not very far from each other. Therefore, we conclude that in spite of different feature aggregation approaches, the optimization parameters do not differ to a large extent. The corresponding confusion matrices are also shown in Table 8.8. Similar to the ANN classifier, using drive time-based features improves the DR of the awake class by about 5% (89.0% vs. 84.2%). However, the DR of the drowsy class drops to a larger extent, namely by about 6% from 83.6% to 77.6%. The comparison of the performance of the SVM with that of the ANN in Table 8.3 indicates that for both binary and 3-class cases, the type of the classifier did not affect the classification results. As an example, the medium class is also confused with the awake class by the SVM classifier. This finding emphasizes two facts. First, the class labels might be imprecise. As a result, regardless of the type of the classifiers applied to the features, some classes are not distinguishable such as awake versus medium class. Second, the features might not be informative enough for our task which limits the result improvement. Consequently, both classifiers performed similarly. Besides the mentioned points, we also conclude that the underlying driver state information in the extracted features is interpreted similarly by both classifiers. Since this is valid independent of the classifiers' type, we consider it as the strong point of the features.

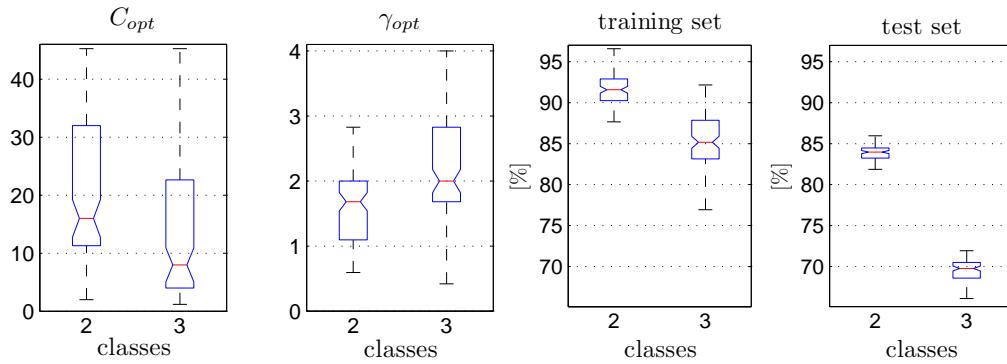


Figure 8.21.: Boxplot of C , γ , training and test accuracies for the 2-class and 3-class subject-dependent classification with the SVM for all 100 permutations. Feature type: drive time-based features

Table 8.8.: Confusion matrices of the subject-dependent SVM classifiers for the 2-class and 3-class cases. Feature type: drive time-based features

			predicted			
			state	awake	drowsy	
given			awake	89.0%	11.0%	
			drowsy	22.4%	77.6%	

			predicted			
			state	awake	medium	drowsy
given			awake	85.3%	9.8%	4.9%
			medium	40.2%	41.7%	18.1%
			drowsy	10.5%	14.9%	74.6%

Results of the imbalanced KSS input-based features

In this section, we study whether the SVM classifier is sensitive to the distribution of classes in the training set. Therefore, similar to Section 8.2.3, we fed imbalanced KSS input-based features of the driving simulator experiment to the binary subject-dependent SVM classifier. The values of C , γ and the corresponding accuracies are depicted in Figure 8.20. Interestingly, in comparison to balanced case, the value of C has a larger range. The calculated confusion matrix is shown on the left part of Table 8.9. The awake class is classified less correctly (63.0%) in comparison to the drowsy class (93.7%) which underscores the tendency of the classifier towards the majority class, namely the drowsy class. We explained in Section 8.3.6 that this happens due to equally penalizing the misclassified samples of both classes during the training phase. There, we also mentioned that this problem can be solved by considering different misclassification costs for each class. According to (8.46), we need the ratio between the number of samples in the minority and majority classes which were $N_- = 50$ and $N_+ = 159$. The confusion matrix of the binary subject-dependent SVM classifier with different misclassification costs is shown in the right part of Table 8.9.

Table 8.9.: Confusion matrices of the binary subject-dependent SVM classifiers for KSS input-based features of driving simulator experiment. Left: imbalanced features, right: balanced features by considering different misclassification costs

			predicted		
			state	awake	drowsy
given			awake	63.0%	37.0%
			drowsy	6.3%	93.7%

			predicted		
			state	awake	drowsy
given			awake	73.3%	26.7%
			drowsy	7.5%	92.5%

According to the listed values, the classification of the awake class improves by about 10%. However, the DR value of the drowsy class decreases by about 1%. This is unlike the ANN

classifier where the DR of the awake class improved to a larger extent by reducing the DR of the drowsy class (see Table 8.4). As a result, the ANN classifier was more sensitive to imbalanced data set.

Overall, both classifiers perform almost similarly in detection of the awake class. However, the SVM outperforms the ANN in the detection of the drowsy class by about 8%. This might be due to the fact that different solutions for handling the imbalanced data were applied. In fact, totally different training sets were fed to these classifiers. The ANN classifier was trained with the new artificially generated awake samples based on the SMOTE, while the SVM classifier received only the same imbalanced training set. In other words, the ANN classifier was retrained by changing the input values and not the structure of the training phase. On the contrary, the SVM classifier was retrained by the fixed input samples and a new structure of the training phase, i.e. different misclassification costs. In addition, the SMOTE is a general approach which is independent of the classifier type. However, the approach used for the SVM directly affects the classifier itself. Consequently, it has yielded better classification results.

In order to know whether the newly trained SVM also classifies unseen awake samples correctly, we applied the features of the real road drives to it. According to the corresponding confusion matrix shown in Table 8.10, both of the classes are detected completely randomly, i.e. DR values are close to 50%. Consequently, similar to the ANN (compare with Table 8.5), an SVM classifier retrained with different misclassification costs is also still unable to classify unseen awake samples. It even fails to classify drowsy samples correctly. In fact, the new classifier is overfitted and is far from a generalized model. Therefore, we conclude that even by applying the SVM classifier and its approach for handling imbalanced classes, the collection of both awake and drowsy data during the experiment is essential.

Table 8.10.: Confusion matrix of the binary subject-dependent SVM classifier for KSS input-based features of the real road experiment applied to the model trained by considering different misclassification costs

		predicted	
		driver state	drowsy
given	awake	44.4%	55.6%
	drowsy	47.6%	52.4%

8.3.8. Classification results of the subject-independent data sets

Similar to Section 8.2.4, we also studied the SVM while considering the subject-independent classification approach. The confusion matrix shown in Table 8.11 does not differ from that of the ANN classifier listed in Table 8.6. Therefore, the type of the classifier seems not to be an issue for solving the problem of unseen data given the drive time-based features. However, Figure 8.22, which compares Table 8.11 with the similar confusion matrix for the subject-dependent classification by the SVM (Table 8.8), indicates that classification of the awake class is less problematic (10% drop of the DR from 89.0% to 79.5%). On the contrary, for the drowsy class, the DR decreases more severely (16%, from 77.6% to 61.5%).

Table 8.11.: Confusion matrix of the binary subject-independent SVM classifier for drive time-based features

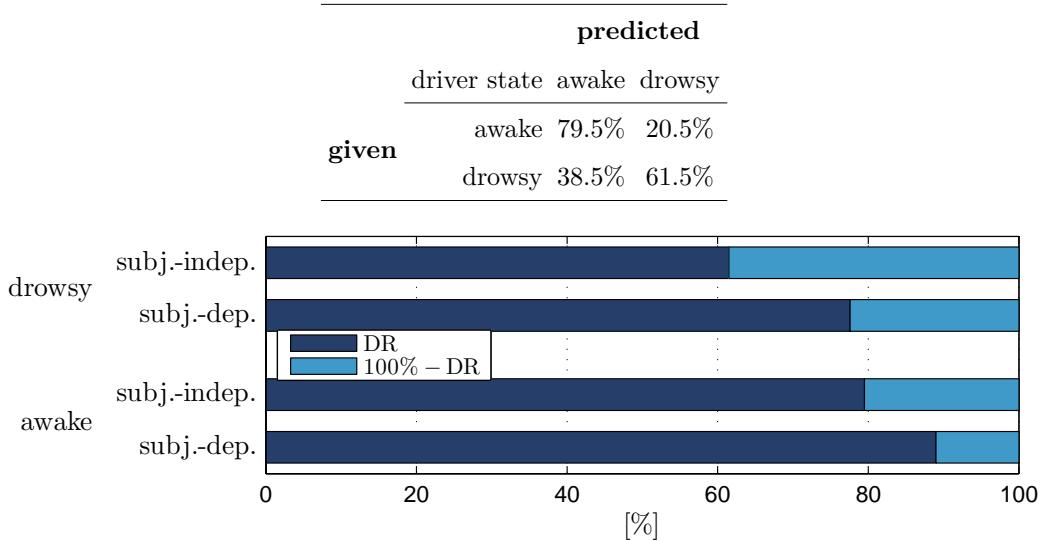


Figure 8.22.: Comparing confusion matrix of the binary subject-dependent SVM classifier with that of the subject-independent

8.4. *k*-nearest neighbors classifier

The classifiers, which were studied in Sections 8.2 and 8.3, are both based on complex ideas and optimization solutions. The ANN classifier works based on neurons and hidden layers and the SVM looks for the optimal separating hyperplane. In this section, another classifier based on a very simple idea, namely *nearest neighbors*, is studied to show how complex a classifier is required to be in our application.

8.4.1. Background theory

The *k*-nearest neighbor classifier is a nonparametric classification method which does not need any information about the underlying distribution of the data set. It is also very famous for its simplicity. As its name shows, this classifier determines the class of each sample depending on the majority class of the *k* nearest adjacent samples. According to Cover and Hart (1967), the number of *k* should not be set very large, otherwise outliers of other classes influence the true class. Moreover, by selecting odd numbers of *k* ambiguous decisions are avoided. However, the special case of *k* = 1 should not be chosen. It is subject to high amount of noise and unreliable results, because it always leads to overfitting.

This method strongly depends on the *distance* between the sample under investigation and samples of the training set. Thus, different metrics are defined for calculating it such as the general metric of *Minkowski* $L_p(\mathbf{x}, \mathbf{y})$

$$L_p(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^N |x_i - y_i|^p \right)^{\frac{1}{p}}. \quad (8.51)$$

In the case of $p = 1$, (8.51) is called *Manhattan* distance or L_1 -norm and for $p = 2$, it is the conventional *Euclidean* distance or the L_2 -norm. As mentioned in Duda et al. (2012), a main

drawback of the Euclidean distance is its sensitivity to the features' scaling, i.e. a large disparity in the range of features. This fact underlines the importance of feature normalization prior to the classification. Another alternative distance metric, which does not suffer from the mentioned problem, is the *Mahalanobis* distance defined as

$$L_{\text{maha}}(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{y})}, \quad (8.52)$$

where \mathbf{S} refers to the covariance matrix. In addition, a linear transformation of \mathbf{x} does not affect the value of L_{maha} (Yang, 2014).

8.4.2. Classification results of the subject-dependent data sets

Similar to other classifiers, in this section, we study the k -NN classifier under subject-dependent and subject-independent, 2-class and 3-class cases and finally its sensitivity to the imbalanced data.

Results of the KSS input-based features

Here, we study the performance of the k -NN classifier with regard to the KSS input-based features for different values of k , namely $k = 3, 5, 7$ and 9 . The Mahalanobis distance was used as the distance metric for finding the nearest neighbors. Figure 8.23 shows the calculated ADR values. The results seem very similar for all values of k with the maximum ADR achieved at $k = 5$. The confusion matrix for $k = 5$ is also listed in Table 8.12. In comparison to the ANN classifier (Table 8.2), this classifier detects the awake class by about 2% better (83.2% vs. 80.9%) which makes it comparable with the SVM classifier (83.2% vs. 84.2%) (Table 8.7). In detection of the drowsy class, however, the SVM classifier is superior to both of them (SVM: 83.6%, ANN: 81.0% and k -NN: 79.7%). Overall, these results show that relying on the class of the nearest neighbors is a good solution for defining the class of the sample under investigation given the KSS input-based features.

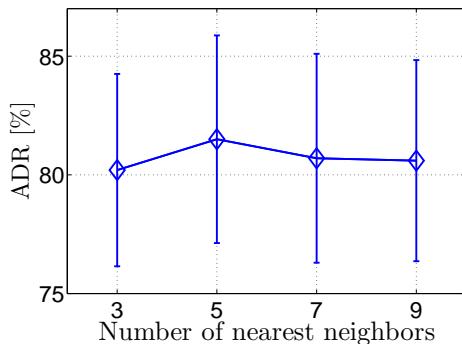


Figure 8.23.: ADR of the test sets of the binary subject-dependent k -NN classifier for different numbers of neighbors. Feature type: KSS input-based features. Bars refer to the standard deviation of permutations.

Results of the drive time-based features

The drive time-based features were also fed into the k -NN classifier for the binary and 3-class cases. The corresponding values of the ADR are shown in Figure 8.24. For both binary and 3-class

Table 8.12.: Confusion matrix of the binary subject-dependent k -NN classifier for $k = 5$. Feature type: KSS input-based features

		predicted		
		driver state	awake	drowsy
given		awake	83.2%	16.8%
		drowsy	20.3%	79.7%

cases, increasing the number of nearest neighbors k slightly improves the classification results. The confusion matrices for $k = 7$ are also listed in Table 8.13. In comparison to the results for the KSS input-based features in Table 8.12, the DR value for the awake class increases by about 3% (86.8% vs. 83.2%), while the detection of the drowsy class shows no improvement (79.2% vs. 79.7%). Comparison of the binary k -NN classifier with the binary ANN and SVM classifiers regarding drive time-based features indicates that, overall, all classifiers perform similarly and the difference between the DR values of each class is only about 2%.

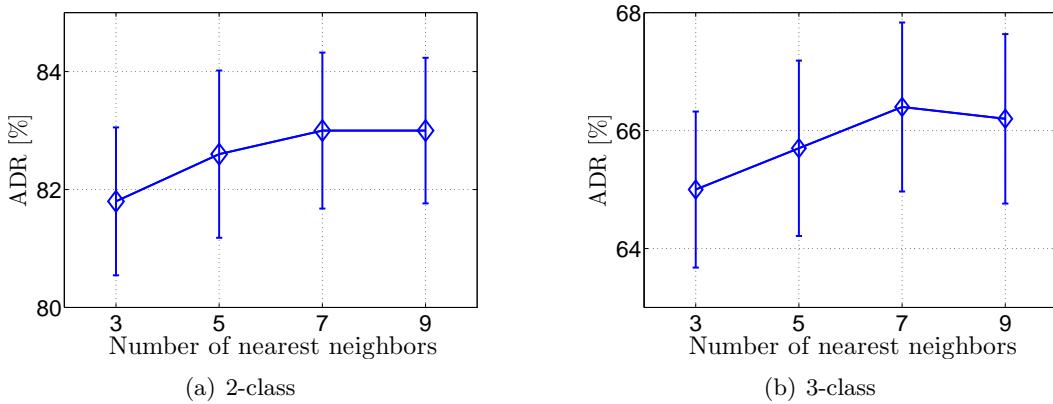


Figure 8.24.: ADR of the 2-class and 3-class subject-dependent k -NN classifier for different numbers of neighbors. Feature type: drive time-based features. Bars refer to the standard deviation of permutations.

Table 8.13.: Confusion matrices of the subject-dependent k -NN classifier ($k = 7$) for the 2-class and 3-class cases. Feature type: drive time-based features.

		predicted			predicted			
		state	awake	drowsy	state	awake	medium	drowsy
given		awake	86.8%	13.2%	awake	82.0%	14.4%	3.6%
		drowsy	20.8%	79.2%	medium	36.8%	47.6%	15.6%
					drowsy	11.0%	19.5%	69.5%

For the 3-class case, the detection of the medium class slightly improves in comparison to that of the ANN and SVM classifiers (k -NN: 47.6%, ANN: 42.8% and SVM: 41.7%). However, this is followed by the degradation of the DR value for the awake and drowsy classes by this classifier.

Results of the imbalanced KSS input-based features

The sensitivity of the k -NN classifier to an imbalanced training set based on the KSS input-based features is studied in this section. According to Figure 8.25, increasing the number of nearest

neighbors k from 3 to 7 slightly improves the DR values. However, $k = 9$ yields similar results to $k = 7$. The confusion matrix for $k = 7$ is listed in Table 8.14. Similar to the ANN and SVM classifiers (Tables 8.4 and 8.9), this classifier also fails to classify the samples of the minority class as correct as the samples of the majority class with about 30% difference between DR values of each class (60.2% vs. 89.2%).

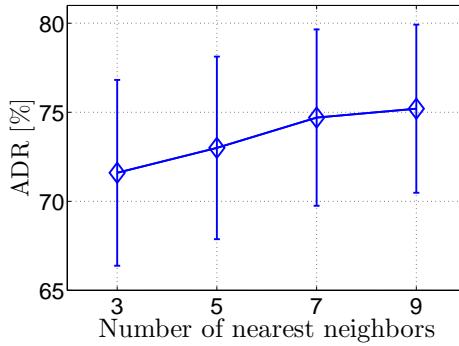


Figure 8.25.: ADR of the test sets of the binary subject-dependent k -NN classifier for different numbers of neighbors. Feature type: imbalanced KSS input-based features of driving simulator experiment. Bars refer to the standard deviation of permutations.

Table 8.14.: Confusion matrix of the binary subject-dependent k -NN classifier ($k = 7$). Feature type: imbalanced KSS input-based features of driving simulator experiment

		predicted	
		driver state	drowsy
given	awake	60.2%	39.8%
	drowsy	10.8%	89.2%

8.4.3. Classification results of the subject-independent data sets

The results of the subject-independent classification for the binary k -NN classifier based on drive time-based features are shown in Table 8.15 for $k = 9$. We showed in Sections 8.2.4 and 8.3.8 that the performance of the ANN and SVM classifiers degraded in the case of subject-independent classification. Nevertheless, the drowsy class was classified by them up to 61% correctly. However, it seems that the k -NN classifier is less suitable for the classification of unseen data in comparison to the other classifiers due to the smaller DR value for the drowsy class (57.4%). By varying the number of k , the results do not improve.

Table 8.15.: Confusion matrix of the binary subject-independent k -NN classifier for drive time-based features ($k = 9$)

		predicted	
		driver state	drowsy
given	awake	80.0%	20.0%
	drowsy	42.6%	57.4%

Figure 8.26 compares the metrics of the confusion matrix for the subject-dependent and subject-independent k -NN classifiers. It can be seen that the DR of the drowsy class drops by about 22%

from 79.2% to 57.4%. For the awake class, however, the DR decreases only by about 7% from 86.8% to 80.0%.

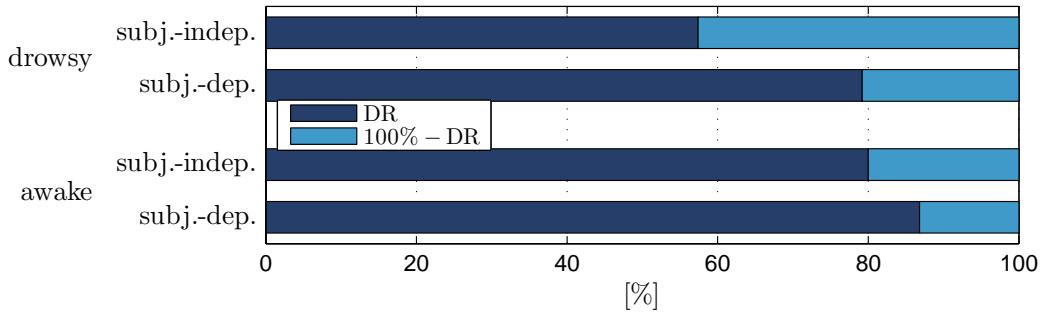


Figure 8.26.: Comparing confusion matrix of the binary subject-dependent k -NN classifier with that of the subject-independent

8.5. Comparison of the supervised classifiers for driver state classification

In previous sections, three classifiers were introduced and their classification results were discussed. In this section, we review them in terms of different aspects such as the performance, subject-dependent versus subject-independent classification, simulation runtime, etc.

Figure 8.27 summarizes the confusion matrices of all binary classifiers for the subject-dependent and subject-independent classifications given the drive time-based features. In this figure, all DR and 100% – DR values as listed in a confusion matrix are provided. $100\% - DR_{awake}$ and $100\% - DR_{drowsy}$ correspond to the FPR and FNR values, respectively. Overall, none of the classifiers outperforms the other ones. If a classifier detects one class with a higher DR value, it is usually at a cost of degraded classification result for the other class. As an example, the binary subject-dependent k -NN classifier classifies the drowsy class slightly better than the SVM and ANN classifiers, while it achieves the smallest value of the DR for the classification of the awake class among other classifiers. Therefore, it seems that any of the classifiers can be applied to the drive time-based features for the subject-dependent classification.

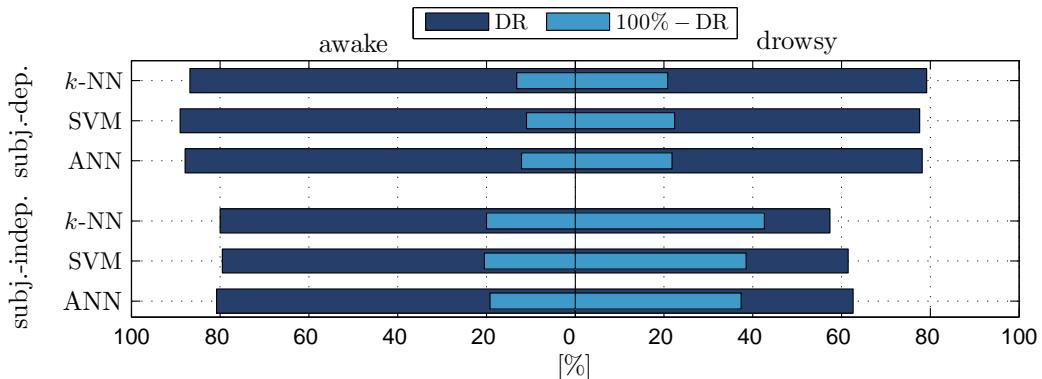


Figure 8.27.: Comparing confusion matrices of the binary ANN, SVM and k -NN classifiers for the subject-independent and subject-dependent classifications. Feature type: drive time-based features

Regarding subject-independent classification, the performance of all classifiers degrade as expected. The k -NN method also seems not to be a suitable classifier for a data set with large

between-subject differences due to its poorer performance in the classification of the drowsy class. The ANN and SVM classifiers interpret the features in a similar way in this case. Nevertheless, a more effective feature baselining method might improve the results of subject-independent classification.

For the 3-class classification, all confusion matrices are summarized in Figure 8.28. All in all, the DR values of all classes and all classifiers are not as good as those of the binary cases. The SVM and ANN classifiers detect awake and drowsy classes with higher DR values in comparison to the *k*-NN classifier. However, the medium class is best detected by the *k*-NN.

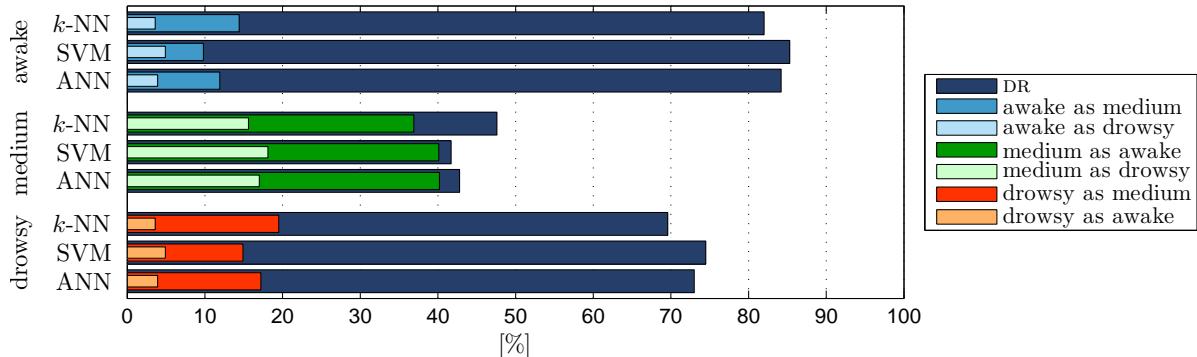


Figure 8.28.: Comparing confusion matrices of the 3-class ANN, SVM and *k*-NN classifiers for the subject-dependent classification. Feature type: drive time-based features

Since the performance of a supervised classifier is highly influenced by the preciseness of the labels, all results provided here also fully depend on the accuracy of the collected KSS values. Apart from that, in this study, all binary classifiers perform well with the ADR of 82% which is 32% better than random guessing. It is, hence, concluded that apart from the informativeness of the features, the subjects were also able to properly distinguish between awake and drowsy states leading to accurate labels. Therefore, for the 3-class problem also all classifiers seldom confuse the awake and drowsy classes with each other, while the medium class is more often misclassified by all of them, especially as the awake class. This might be due to the fact that the transition from the awake to the medium state was not very distinguishable for the subjects themselves and, consequently, they interpreted their states inconsistently. Nevertheless, an average DR of 66% for the 3-class classification is still twice more effective than a random classifier.

Figure 8.29 shows all results for the KSS input-based features. The first three bars from the top compare the DR values for the balanced case where all features of all experiments, namely real road and driving simulator experiments, were included in the training and test sets. All classifiers perform almost similarly. SVM achieves the maximum DR values for both of classes.

It should be mentioned that the number of samples for KSS input-based features are about 10 times smaller than that of the drive time-based ones. Hence, comparing the classification results based on each feature aggregation method should be done with care. Taking the large difference between the number of samples in the training sets into consideration, the classification results of the KSS input-based features are even more satisfactory. In fact, the smaller DR values for the KSS input-based features do not necessarily imply that these features are not as informative as the drive time-based features. Nevertheless, adding a larger number of samples to the training sets of KSS input-based features might improve the results.

Regarding imbalanced training sets, also shown in Figure 8.29, we conclude that the SVM classifier with the standard misclassification cost is more robust against imbalanced distributions of the

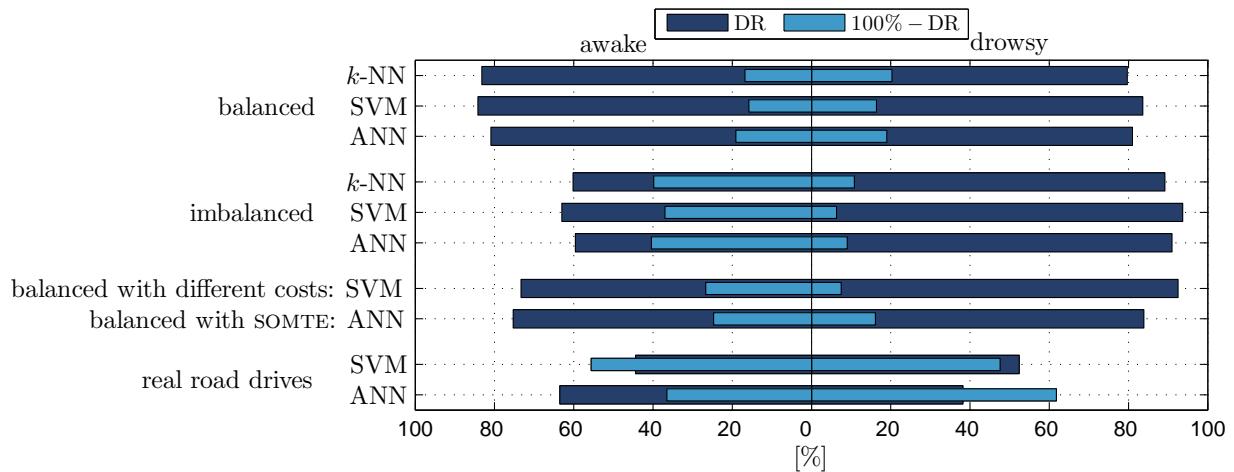


Figure 8.29.: Comparing confusion matrices of the binary subject-dependent ANN, SVM and *k*-NN classifiers for different KSS input-based features

classes, since it achieves the highest values of the DR for both classes among other classifiers. By applying two methods, we tried to improve the imbalance issue of the training set, namely SMOTE and different misclassification costs. The former was combined with the ANN classifier and we artificially added new samples to the minority class. The latter, however, took properties of the SVM classifier into account. For both methods, the classification results improve, but the SVM classifier is superior to the ANN in this case. This might be due to the fact that the SMOTE solves the imbalanced data problem regardless of the classifier, while different misclassification costs approach solves this problem by directly manipulating the classifier itself.

Although we could solve the problem with the lack of awake samples in the training sets, the resulting classifiers are still inefficient in classifying unseen awake samples. The ANN classifier classifies these samples up to about 63% correctly at a cost of misclassifying most of the unseen drowsy samples. The SVM, however, performs not better than random guessing. Therefore, it can be inferred that both of the methods lead to overfitted classifiers, such that they cannot be generalized to other unseen data. This occurred despite the approaches which we applied during the training phase, such as the cross validation, to avoid overfitting. This finding might also be related to the overall difference between driving simulator and real driving conditions.

As mentioned before, these results confirm that for driver state classification, always two types of events should be available in the training sets, namely both awake and drowsy events. In fact, a warning system, which always warns all drowsy subjects correctly, is not necessarily a reliable system, if it is unable not to warn awake subjects. On this account, in this study, part of the data set was collected during the night as the drowsy data and the other part during the day as the awake data, because we showed that they cannot be artificially substituted in the data set in the case of imbalanced class distributions.

In addition to the classification performance, classifiers can also be compared with each other in terms of the simulation runtime for achieving the mentioned results. All of the results in this work were generated by MATLAB. The *k*-NN classifier is the fastest, with the runtime of less than 10 s for classifying drive time-based features. The runtime for the ANN classifier depends highly on the selected number of neurons. On average, training a network with $N_h = 2$ neurons is very fast (< 1 min), while for $N_h = 10$, it takes about 5 min. The highest runtime is needed by the SVM with about 45 min. All of these runtime durations are based on one iteration out of 100 permutations given the drive time-based feature. Considering both the performance of

the classifier, namely all achieved results, and the computational complexity factor, the ANN classifier is the best solution for the driver state classification based on blink features in this work.

8.6. Features of the driving simulator versus real road driving

As mentioned before, due to safety concerns, driving simulators are required for data collection with higher drowsy-related characteristics. However, this makes the collected data less applicable for the comparison with real driving conditions. In this context, Hallvig et al. (2013) reported longer blink duration in the driving simulator compared to real driving and believes that due to safety aspects of driving simulators higher level of drowsiness is generally achieved in them. Philip et al. (2005), who also compared real driving with driving in simulators, reported slower reaction time and higher KSS values in the driving simulator. Therefore, all good classification results based on driving simulator data might suffer from the fact that very deep phases of drowsiness are included in the data set which sharpens the discrimination of classes. In general, under real driving conditions, drowsiness should be detected at a time that the warning of the corresponding assistance system can still be perceived by the driver for a timely correcting reaction.

To address the mentioned issues, two approaches are considered here. The first approach generalizes the driving simulator to real driving conditions by discarding very drowsy parts of the drives. Unlike the first approach, the second approach uses all valuable features collected in both driving simulator and under real driving conditions to investigate whether unseen drowsy data collected under real driving conditions can be classified correctly.

8.6.1. Generalization of the simulator data to real road driving

In this section, we want to investigate, if driver drowsiness in the driving simulator is representative of the same effect under real driving conditions. To this end, we removed all parts of the drives after the first KSS = 9 or after the second KSS = 8 from the data set. This procedure is referred to as the *generalization of simulator to real driving* (GSRD). The reason that we applied these conditions to the data is that as mentioned in Section 4.2.3, exactly these rules have been followed for collecting drowsy data on the real road.

The new feature matrix contains 3070 samples of all 19 drive time-based features (19×3070) based on real road and driving simulator experiments. After removing approximately 950 samples from the drowsy class, the classes are distributed as awake = 72% and drowsy = 28%. As we discussed in previous sections, imbalanced class distributions degrade the performance of the classifiers. Therefore, we randomly undersampled the features, i.e. repeatedly removing random samples of the majority class to obtain balanced distributed classes (50% vs. 50%) as explained in Section 8.1.5. The ANN, SVM and k -NN classifiers were again applied to the new feature matrix including both subject-dependent and subject-independent data sets, but only for the binary case. The KSS input-based features were not studied due to a smaller number of available samples after removing the samples of the drowsy class according to the above procedure.

Results of the ANN classifier

Figure 8.30 depicts the ADR values for different numbers of neurons of the GSRD subject-dependent ANN classifier. The comparison of this figure with Figure 8.10(a) indicates that for all values of N_h the ADR of the test sets decreases by approximately 4–7% with slightly larger standard deviation values. Moreover, increasing the number of neurons N_h from 10 to 25 only improves the ADR of the training set. The corresponding values of the confusion matrix for $N_h = 10$ are shown in Figure 8.31. The DR value for the awake class is more affected in comparison to Figure 8.27 (about 10% drop from 87.9% to 78.4%).

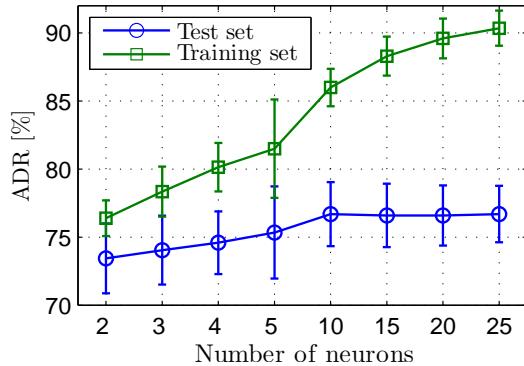


Figure 8.30.: ADR of the training and test sets of the binary subject-dependent ANN classifier for different numbers of neurons based on GSRD case. Feature type: drive time-based features. Bars refer to the standard deviation of permutations.

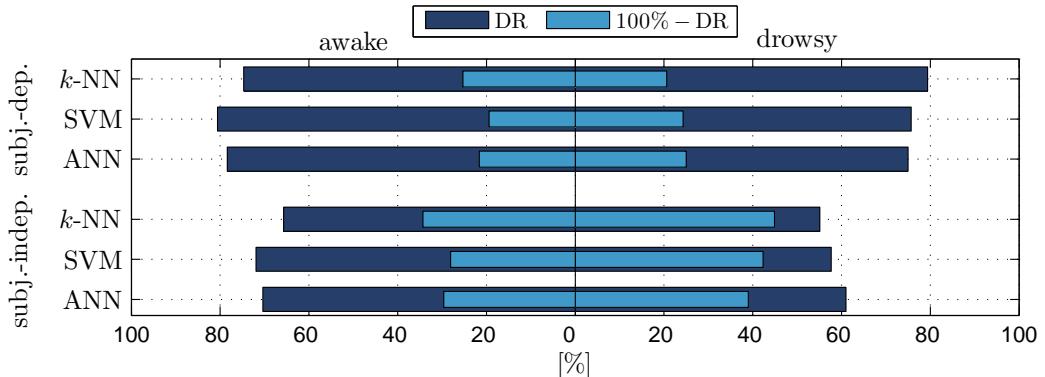


Figure 8.31.: Comparing confusion matrices of the binary ANN ($N_h = 10$), SVM and k -NN ($k = 7$) classifiers for the subject-dependent and subject-independent classifications of the GSRD case. Feature type: drive time-based features.

Figure 8.31 also shows the result of the subject-independent classification which should be cautiously compared with the corresponding results in Figure 8.27. In GSRD, only parts of the drowsy phase per subject were removed, leading to smaller numbers of drowsy samples and smaller values of FP+TN. In the worst case, if TP, FP and FN values of the confusion matrix remain unchanged after applying the GSRD procedure, the decreased number of TN results in extremely increased FPR in (8.6) (or $100\% - \text{DR}$) as occurred here. Surprisingly, for the subject-independent case, it is still the correct classification of the drowsy class which is more problematic (DR of the drowsy class $<$ DR of the awake class). Despite this fact, applying the GSRD procedure degrades the classification of the awake class to a larger extent, i.e. 10% drop of the DR value from 80.0% to 70.4%.

Results of the SVM classifier

The parameters for training the SVM classifier under the GSRD case are shown in Figure 8.32. These parameters are slightly different from those shown in Figure 8.21. In particular, the interquartile ranges are larger for the GSRD case. The accuracy of the test sets also decreases by about 6%. Figure 8.31 also depicts the results of subject-dependent and subject-independent classifications of the GSRD for the SVM. Similar to the results of the ANN classifier, these results are not as good as the previous results shown in Figure 8.27, with a 5% drop of the ADR for the subject-dependent case (from 83.3% to 78.1%) which is mostly due to 8% degradation in the correct classification of the awake class (from 89.0% to 80.6%). For the subject-independent case, a similar explanation holds.

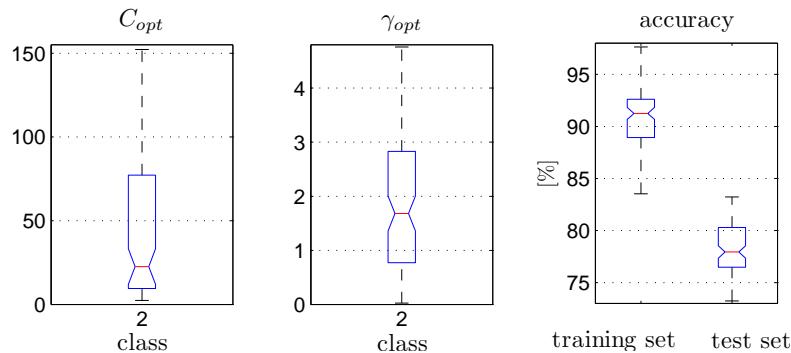


Figure 8.32.: Boxplot of C , γ , training and test accuracies for the 2-class subject-dependent classification of the GSRD case with SVM for all 100 permutations. Feature type: drive time-based features

Results of the k -NN classifier

The results for different numbers of neighbors k under the GSRD procedure are shown in Figure 8.33. In comparison to Figure 8.24(a), the GSRD procedure impairs all metrics by approximately 6%. The metrics of the confusion matrices for both classes under subject-dependent and subject-independent cases are also shown in Figure 8.31. By comparison with the corresponding DR values of the ANN and SVM classifiers, the k -NN subject-dependent classifier classifies the drowsy class with the highest DR value (k -NN: 79.4%, SVM: 75.7% and ANN: 75.0%), although it is not the best classifier for classification of the awake class. As a subject-independent classifier, k -NN has the poorest performance with the lowest DR values for both classes. Therefore, the k -NN classifier is the least suitable classifier, if the unseen data differs a lot from the training set.

Conclusion

Considering the results of the GSRD approach provided by all classifiers for the subject-dependent data division, removing samples from the very drowsy parts of the drive degrades the performance of the classifiers. Nevertheless, it is still possible to correctly detect both classes over 70%. An interesting finding is that regardless of the data division type, namely subject-dependent versus subject-independent, the removed drowsy samples are crucial not only for the correct classification of the drowsy samples, but also for the correct classification of the awake samples. The DR value of the drowsy class by the subject-independent k -NN classifier, as an example, varies only 2% (57.4% vs. 55.1%), while for the awake class, it drops by about 15% (80% vs. 65.7%).

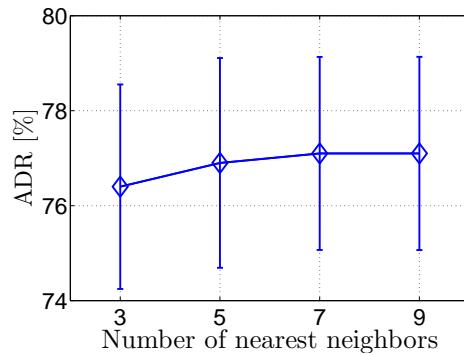


Figure 8.33.: ADR of the 2-class subject-dependent k -NN classifier of the GSRD case for different numbers of neighbors. Feature type: KSS input-based features. Bars refer to the standard deviation of permutations.

8.6.2. Classification of drop-outs under real road driving conditions

On the contrary to the previous approach, where parts of the drowsy data were removed, here, we used the entire feature matrix based on all conducted experiments as the training set. This set was then fed to the ANN, SVM and k -NN classifiers. For the test set we used new unseen data which was collected nighttime under real driving conditions as explained in Section 4.2.3. This test set comprised solely the data from subjects who aborted the real driving experiment due to severe drowsiness according to their own subjective assessment or that of the investigator. Simon et al. (2011) called such subjects “*drop-outs*” and believed that this condition is the “*most objective fatigue criterion available*”. They also found a larger variation of the EEG features for these subjects in comparison to non-drop-outs who completed the experiment to the end. Following the plausible idea of Simon et al. (2011), we used the data of the drop-outs participated in our experiment as the test set and repeated the classification task by the trained network or model which is, in fact, a subject-independent case.

Figure 8.34 shows the resulting confusion matrices for the ANN, SVM and k -NN classifiers. The parameters of the classifiers are $N_h = 10$ for the ANN, $C = 90.5$ and $\gamma = 1.4$ for the SVM and $k = 7$ for the k -NN. These parameters lead to the best classification results compared to other parameter values. It seems that overall, the ANN classifier outperforms other classifiers. In classifying the awake class, all of the classifiers perform similarly, while the drowsy class is classified more correctly by the ANN classifier.

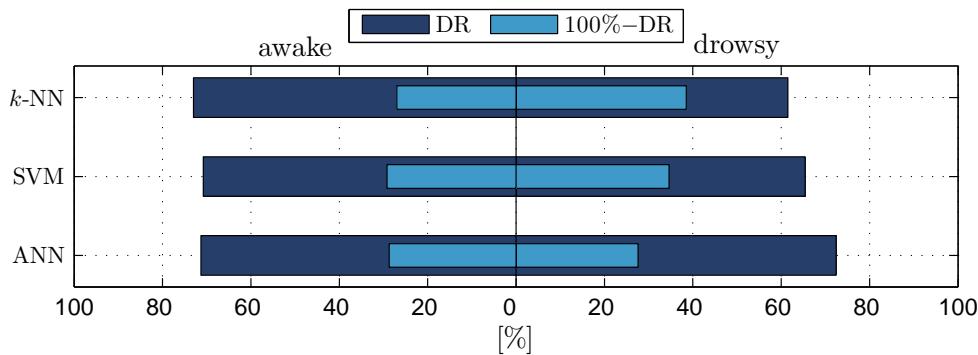


Figure 8.34.: Comparing confusion matrices of the binary subject-independent ANN ($N_h = 10$), SVM ($C = 90.5$, $\gamma = 1.4$) and k -NN ($k = 7$) classifiers for unseen real road drives of drop-outs. Feature type: drive time-based features.

For drive time-based features, all classifiers achieve an accuracy of about 70%. Therefore, we conclude that the collected drowsy events in the driving simulator are not far from reality. The remaining 30% wrong classifications in each group might occur due to the followings:

- Getting drowsy in the driving simulator is in some unknown senses different from getting drowsy under real driving conditions. As a result, the drowsy events collected in the driving simulator do not represent all drowsy events of the real drives.
- Large between-subject differences in blink features leads to individual characteristics in different subjects which might be sorted out by including more subjects in the training set in the future.

8.7. Feature dimension reduction

In the previous sections, we discussed the classification results with regard to 19 features in the feature matrix. Regardless of the performance of the classifiers, it is not clear which single feature or which feature subset has more contribution in generating the results. The reason is that features were applied *all together* to the classifier in order to complement each other. In this section, we explore whether all of the 19 features are needed or whether it is possible to reduce the feature matrix dimension and achieve the same results with a smaller number of features. This is an important issue for in-vehicle warning systems as we explain in the following.

Imagine that the quality of an extracted feature or a feature subset deteriorates as soon as the EOG is replaced with a driver observation camera due to image processing problems and lower frame rate. If we show that the existence of this feature or this feature subset is not crucial for having a reliable driver state classification, then the degraded feature quality is not an issue any more. Apart from that, for most of in-vehicle systems, processing time and memory storage are serious concerns. Consequently, a feature matrix with a reduced dimensionality is desired and preferred.

The application of feature dimension reduction is also motivated by the *curse of dimensionality* which requires the number of unknown parameters of a classifier to be at least 5 to 10 times less than the number of samples available in the training set (Yang, 2014). In other words, by increasing the dimension of the feature matrix, the number of training samples should increase exponentially, otherwise overfitting is inevitable (Bishop, 2006). This point plays an important role in applications with small numbers of available training samples in comparison to the number of extracted features. Therefore, this is not an issue in this work. Feature dimension reduction also helps to avoid redundancy by keeping the most uncorrelated features for further analysis.

According to Yang (2014), there exist two types of approaches for reducing the dimension of the feature matrix. The first method, called *feature selection*, selects a subset of features with the desired dimension \tilde{D} out of available D features, where $\tilde{D} < D$. On the contrary, the *feature transform* method transforms the features either linearly or non-linearly to reduce D . Both of these methods are guided either by a classifier-dependent metric, e.g. accuracy of the classification result or by a classifier-independent criterion such as the correlation. The former, which uses the learning algorithm itself, is a *wrapper* approach discussed in Sections 8.7.1 and 8.7.2. The latter is a *filter* approach which is based on the intrinsic properties of the data and is discussed in Section 8.7.3.

8.7.1. Sequential floating forward selection

Sequential floating forward selection (SFFS), a feature selection method introduced by Pudil et al. (1994), is a combination of both *sequential forward selection* (SFS) and *sequential backward selection* (SBS). The former achieves the desired number of features by adding the best feature combination to an empty feature set, while the latter achieves this goal by removing the worst feature combination from the full feature set. According to Pudil et al. (1994), both of these feature selection methods suffer from wrong decisions, called *nesting effect*, during adding or removing a feature. This is due to the fact that no correction steps are considered in their algorithms. Hence, combining these methods with each other results in a more dynamic feature selection method. Based on the considered criterion for evaluating the selected features (e.g. classification performance), inclusion and exclusion steps are then applied. In other words, after adding a new feature, some backward steps are performed as long as the new subset outperforms the previous one. If this is not the case, the backward step is disregarded.

In the following, the algorithm is clarified by a numerical example. The mathematical notation of the SFFS algorithm is taken from Lugger (2011).

We consider the original feature set \mathcal{M} with 7 features, namely $\mathcal{M} = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$ and the desired number of features $\check{D} = 4$. The SFFS algorithm starts with an empty feature set $\mathcal{Y}_k = \emptyset, k = 0$ and selects the best single feature as follows

$$\begin{aligned} y_i &= \arg \max_{y \in \mathcal{M} \setminus \mathcal{Y}_k} J(\mathcal{Y}_k \cup y) \\ \mathcal{Y}_{k+1} &= \mathcal{Y}_k \cup y_i \\ k &= k + 1, \end{aligned} \tag{8.53}$$

e.g. $\mathcal{Y}_1 = \{y_4\}$. J denotes the performance function. Then the remaining 6 features will be added to \mathcal{Y}_1 separately and the best pair is selected based on the criterion in (8.53) with $k = 1$, e.g. $\mathcal{Y}_2 = \{y_4, y_2\}$ out of $\mathcal{Y}_1 \cup y = \{y_4, y_1\}, \{y_4, y_2\}, \dots, \{y_4, y_7\}$. This forward step is repeated for the third time leading to the best 3-feature combination, e.g. $\mathcal{Y}_3 = \{y_4, y_2, y_7\}$. Now, the backward step is applied for controlling the redundancy in the new feature set (Uhlich, 2006). If a recently added feature contains similar information, which was already taken into consideration, it can then be excluded. Consequently, the SFFS algorithm analyzes the following subsets of the current 3-element feature set: $\{\mathcal{Y}_3 \setminus y, y \in \mathcal{Y}_3\} = \{y_2, y_7\}, \{y_4, y_7\}, \{y_4, y_2\}$ and selects the best subset, e.g. $y_j = 4, \mathcal{Y}_3 \setminus y_j = \{y_2, y_7\}$, where y_j is defined as

$$y_j = \arg \max_{y \in \mathcal{Y}_k} J(\mathcal{Y}_k \setminus y). \tag{8.54}$$

If the performance of this new subset, namely $\mathcal{Y}_3 \setminus y_j = \{y_2, y_7\}$, is better than that of $\mathcal{Y}_3 = \{y_4, y_2, y_7\}$, then y_4 is excluded. This leads to the new definition of \mathcal{Y}_2 , i.e. $\mathcal{Y}_2 = \{y_2, y_7\}$. Mathematically, we have

$$\text{if } J(\mathcal{Y}_k \setminus y_j) > J(\mathcal{Y}_{k-1}), \quad \mathcal{Y}_{k-1} = \mathcal{Y}_k \setminus y_j, \text{ repeat (8.54) with } k = k - 1. \tag{8.55}$$

Otherwise, a new feature will be added to the current feature set based on (8.53) as follows

$$\text{if } J(\mathcal{Y}_k \setminus y_j) \leq J(\mathcal{Y}_{k-1}), \quad \text{repeat (8.53)} \tag{8.56}$$

By these steps, the SFFS algorithm looks for the best feature combination. Hence, a feature,

which does not improve the performance of the current feature combination, is then discarded similar to y_4 in our example. This is valid regardless of the previous good results provided by that feature. Clearly, SFFS is a classification-dependent feature selection method.

We applied the SFFS algorithm combined with the ANN classifier ($N_h = 10$) to select the best 10-feature combination out of 19 ($D = 19$ and $\check{D} = 10$) drive time-based features. Accuracy was used to guide inclusion/exclusion steps of this algorithm as shown in Figure 8.35. Selected features in each of the 10 steps are listed in Table 8.16. At the first glance, it seems that 4 features are enough for the driver state classification task, since increasing the number of features does not improve the classification accuracy. The right part of Table 8.17 shows the confusion matrix for the best selected 4 features by the ANN classifier ($N_h = 10$). A similar confusion matrix with the full 19 features (Table 8.3) is also shown on the left of Table 8.17. Comparison of the DR values in both cases indicates that the detection of the awake class is apparently possible with only 4 features with the same performance as that of the 19 features (87.3% vs. 87.9%). However, the remaining 15 features are responsible for improving the DR value of the drowsy class by about 4% (74.3% vs. 78.2%). In other words, the 4% increase of the DR for the drowsy class underscores the fact that some of the underlying information in the feature space can be covered only if more than the 4 mentioned features are included in the feature space.

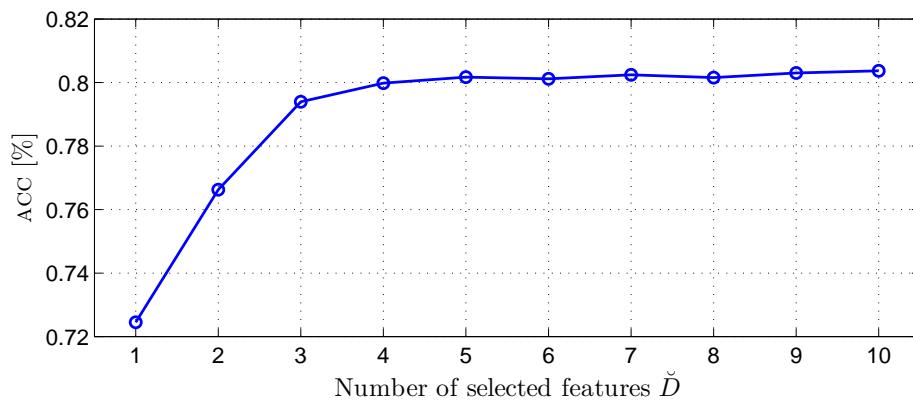


Figure 8.35.: The ANN classification accuracy of the best selected features by the SFFS algorithm from 1 to 10-feature combination. Feature type: drive time-based features.

Table 8.16.: Best selected feature combination set by the SFFS and ANN classifier from 1 to 10 features. Feature type: drive time-based features

\check{D}	selected features								accuracy		
1	T_{50}								72.4%		
2	T_{50}	F							76.6%		
3	E	F	MCV						79.4%		
4	E	F	MCV	A					80.0%		
5	E	F	MCV	A	T_{ro}				80.2%		
6	E	F	MCV	A	T_{ro}	$PERCLOS$			80.1%		
7	E	F	MCV	A	T_{ro}	$PERCLOS$	ACV		80.2%		
8	E	F	MCV	A	T_{ro}	$PERCLOS$	ACV	T_o	80.2%		
9	E	F	MCV	A	T_{ro}	AOV	ACV	T_o	MOV	80.3%	
10	E	F	MCV	A	T_{ro}	AOV	ACV	T_o	MOV	$T_{cl,2}$	80.4%

Additionally, we applied the trained ANN classifier with 4 features to the unseen features of the drop-outs introduced in Section 8.6.2. This helps to investigate the generalization aspect of the classifier based on 4 features for unseen data. Right part of Table 8.18 shows the confusion matrix

Table 8.17.: Confusion matrices of the binary subject-dependent ANN classifier ($N_h = 10$) for drive time-based features. Left: classification with 19 features. Right: classification with 4 features

			predicted						predicted				
			state	awake	drowsy				state	awake	drowsy		
			given	awake	87.9%	12.1%				given	awake	87.3%	12.7%
				drowsy	21.8%	78.2%					drowsy	25.7%	74.3%

for the binary subject-independent ANN classifier for drive time-based features of drop-outs with $\bar{D} = 4$ and $N_h = 10$. On the left part of this table, the confusion matrix of the corresponding subjects with 19 features (shown in Figure 8.34) is listed. Interestingly, the DR of the awake class improves by about 5% (from 71.3% to 77.0%) after removing 15 features from the data set. For the drowsy class, however, the DR drops by about 7% (from 72.4% to 65.2%). This emphasizes the role of other features for the correct classification of the drowsy unseen data.

Table 8.18.: Confusion matrices of the binary subject-dependent ANN classifier ($N_h = 10$) for drive time-based features of drop-outs. Left: classification with 19 features. Right: classification with 4 features

			predicted						predicted				
			state	awake	drowsy				state	awake	drowsy		
			given	awake	71.3%	28.7%				given	awake	77.0%	23.0%
				drowsy	27.6%	72.4%					drowsy	34.8%	65.2%

The 4 selected features are E , F , MCV and A , with the Spearman's correlation coefficient calculated between them and KSS values as listed in Table 7.11: 0.11, 0.36, -0.41 and -0.23. As mentioned before, it is the complementary property of a feature which determines its contribution to the classification task, not its efficiency and informativity as a single feature. Moreover, according to Table 7.21, except for the feature pair (A, MCV) , which is linearly up to 0.78 correlated with each other, for other pairs we have $\rho_p < 0.5$. As a result, the low correlation of the features with each other seems to positively impact their complementary properties.

8.7.2. Margin influence analysis

Margin influence analysis (MIA) (Li et al., 2011a) is a feature selection method especially designed for the SVM classifier. As mentioned in Section 8.3, the performance of the SVM classifier thoroughly depends on the size of the margin between classes. Consequently, SVM models with larger margins are expected to perform better in classification. Regarding this fact, Li et al. (2011a) suggested the MIA method which evaluates a feature based on its contribution to and influence on the classifier's margin. In other words, a feature is considered as informative, if its consideration leads to an SVM model with a larger margin.

According to Li et al. (2011a), the MIA algorithm includes the following steps:

1. Define the number of features \bar{D} to be sampled by the *Monte Carlo sampling* (MCS) (Lemieux, 2009) which is equivalent to the desired dimensionality of the feature matrix.
2. By applying N_{MCS} Monte Carlo sampling, select \bar{D} features randomly out of all existing D features in each sampling (e.g. $N_{MCS} = 10000$ (Li et al., 2011a)). This leads to N_{MCS} subsets each containing \bar{D} features.
3. Train an SVM for all N_{MCS} feature subsets which results in N_{MCS} margins.

4. For a specific feature i , find all subsets \mathcal{A}_i which include this feature and all subsets \mathcal{B}_i which exclude it.
5. Compute the mean value of margins in subsets \mathcal{A}_i and \mathcal{B}_i , namely $\overline{\gamma_{\mathcal{A}_i}}$ and $\overline{\gamma_{\mathcal{B}_i}}$.
6. Define feature i as informative based on the difference between the computed mean values, namely $\Delta\gamma_i = \overline{\gamma_{\mathcal{A}_i}} - \overline{\gamma_{\mathcal{B}_i}}$, as follows
 - $\Delta\gamma_i > 0$: feature i is informative and increases the margin of the SVM classifier.
 - $\Delta\gamma_i \leq 0$: feature i is not considered informative and consequently will decrease the margin of the SVM classifier.
7. Remove all features not considered informative.
8. Determine whether informative features lead to a significant increase of the margin by applying the *Mann-Whitney U test* (Field, 2007) and analyzing the calculated p -value as follows
 - p -value $\leq \alpha$: feature i significantly increases the margin and is informative.
 - p -value $> \alpha$: feature i does not increase the margin significantly.

According to Field (2007), the Mann-Whitney U test is a non-parametric test for comparing mean values.

Table 8.19 shows the calculated $\Delta\gamma$ values for each drive time-based feature with $\check{D} = 4$. The corresponding p -value of the Mann-Whitney U test is also shown. It can be seen that except for A , MOV , A/MCV , A/MOV , ACV , AOV , F , T_o and $T_{c1,1}$ with $\Delta\gamma > 0$, the other features cannot be proved informative. According to the p -values, only A/MCV has passed the significance test at $\alpha = 0.05$. In comparison to the features selected by the SFFS combined with the ANN classifier, only A was selected as an informative feature also by the MIA and the SVM classifier.

Table 8.19.: Values of $\Delta\gamma$ calculated based on the MIA method for drive time-based features ($\check{D} = 4$)

Feature	$\Delta\gamma$	p -value	Feature	$\Delta\gamma$	p -value
A	0.002	0.24	T_c	-0.000	1.31
E	-0.002	1.78	T_o	0.001	0.84
MCV	-0.001	1.88	$T_{cl,1}$	0.005	0.12
MOV	0.002	0.65	$T_{cl,2}$	-0.008	1.00
A/MCV	0.006	< 0.001	T_{ro}	-0.002	1.99
A/MOV	0.003	0.55	$PERCLOS$	-0.002	1.24
ACV	0.000	0.71	T_{50}	-0.001	1.69
AOV	0.003	0.07	T_{80}	-0.004	1.09
F	0.003	0.08	T_{90}	-0.007	1.00
T	-0.002	1.13			

Overall, this method is a very time-consuming approach, depending on the number of features \check{D} to be selected. We showed in Section 8.5 that the SVM is a difficult classifier in terms of runtime. Moreover, for a feature matrix with small dimensions, e.g. $\check{D} \leq 5$ instead of $D = 19$, it is more time-consuming to construct and train an optimized model. Therefore, in this respect, the MIA approach is impractical.

8.7.3. Correlation-based feature selection

Another feature selection method introduced by Hall (1999) is the *correlation-based feature selection* (CFS). This method is, on the contrary to the previous approaches, a classifier-independent

feature selection method. According to Hall (1999), the term *correlation* used in the name of the method does not necessarily refer to the classical linear correlation. It should be interpreted as any measure quantifying the amount of relationship and dependency between two features.

The CFS method first generates subsets \mathcal{R} of features with k_{CFS} numbers of features in each subset out of all available D features. As an example, for $D = 19$ and $k_{\text{CFS}} = 3$ in our study, this corresponds to 3-feature combination subsets of all 19 features, namely 969 subsets. Afterwards, the subsets are ranked based on the following evaluation metric $M_{\mathcal{R}}$ for a subset \mathcal{R}

$$M_{\mathcal{R}} = \frac{k_{\text{CFS}} \bar{r}_{\text{cf}}}{\sqrt{k_{\text{CFS}} + k_{\text{CFS}}(k_{\text{CFS}} - 1) \bar{r}_{\text{ff}}}}, \quad (8.57)$$

where \bar{r}_{cf} and \bar{r}_{ff} denote the average correlation (in its general meaning) between classes and features, i.e. inter-correlation, and between features, i.e. intra-correlation. Obviously, considering both correlations in the evaluation function (8.57) covers both redundancy and relevance issues of features simultaneously.

Calculating $M_{\mathcal{R}}$ for all $2^D - 1$ feature combination subsets makes this method, however, very time-consuming for very large values of D . Hall (1999) suggested forward or backward selection search instead where a feature is added only if any improvement is seen. Moreover, he studied the so-called *relief* (Kira and Rendell, 1992a,b) and the *minimum description length* (MDL) (Rissanen, 1978) as a correlation measure. In a later study (Hall, 2000), however, he suggested the Pearson correlation coefficient for continuous attributes.

Table 8.20 shows the results of the CFS method for different numbers of drive time-based features in a subset using the Pearson correlation coefficient for calculating \bar{r}_{cf} and \bar{r}_{ff} . For each value of k_{CFS} , the best feature subset based on the maximum value of $M_{\mathcal{R}}$ is listed. Similarly, Table 8.21 shows the $M_{\mathcal{R}}$ values using the Spearman's rank correlation coefficient as a measure of relevance and dependency of features. The features, which are shown in red, were selected by both Pearson and Spearman's rank correlation coefficients.

Table 8.20.: Best selected drive time-based features based on the CFS method and Pearson correlation coefficient (Red features were also selected by the Spearman's rank correlation coefficient in Table 8.21.)

k_{CFS}	selected features								$M_{\mathcal{R}}$	
1	<i>A/MOV</i>								0.478	
2	<i>F</i>	<i>ACV</i>							0.553	
3	<i>F</i>	<i>MOV</i>	<i>T_{ro}</i>						0.595	
4	<i>F</i>	<i>MOV</i>	<i>T_{ro}</i>	<i>T_c</i>					0.594	
5	<i>F</i>	<i>MOV</i>	<i>A/MOV</i>	<i>PERCLOS</i>	<i>T₉₀</i>				0.592	
6	<i>F</i>	<i>MOV</i>	<i>A/MOV</i>	<i>PERCLOS</i>	<i>T₉₀</i>	<i>T_{ro}</i>			0.589	
7	<i>F</i>	<i>MOV</i>	<i>A/MOV</i>	<i>PERCLOS</i>	<i>T₉₀</i>	<i>T_{ro}</i>	<i>MCV</i>		0.584	
8	<i>F</i>	<i>MOV</i>	<i>A/MOV</i>	<i>PERCLOS</i>	<i>T₉₀</i>	<i>T_{ro}</i>	<i>MCV</i>	<i>T</i>	0.575	
9	<i>F</i>	<i>MOV</i>	<i>A/MOV</i>	<i>PERCLOS</i>	<i>T₉₀</i>	<i>T_{ro}</i>	<i>MCV</i>	<i>T_o</i>	<i>ACV</i>	0.570
10	<i>F</i>	<i>MOV</i>	<i>A/MOV</i>	<i>PERCLOS</i>	<i>T₉₀</i>	<i>T_{ro}</i>	<i>MCV</i>	<i>T_o</i>	<i>ACV</i>	<i>T₈₀</i> 0.565

The calculated $M_{\mathcal{R}}$ values for both correlation coefficients are shown in Figure 8.36. According to this figure, the $M_{\mathcal{R}}$ values for $k_{\text{CFS}} = 2$ to 5 are almost the same, regarding both correlation coefficients. From $k_{\text{CFS}} = 6$, however, the values deviate from each other to a larger extent. This might be due to the non-linear relationship between features and kss labels, since $M_{\mathcal{R}}$ values using Spearman's rank correlation coefficient are larger. In general, CFS with both correlation

coefficients selects almost the same feature sets, except for $T_{cl,2}$ and T_{90} which were selected only by one of the correlation coefficients repeatedly.

Table 8.21.: Best selected drive time-based features based on the CFS method and Spearman's rank correlation coefficient (Red features were selected by the Pearson correlation coefficient in Table 8.20.)

k_{CFS}	selected features								M_R		
1	T_{ro}								0.510		
2	F	ACV							0.546		
3	F	MOV	T_{ro}						0.589		
4	F	MOV	$PERCLOS$	T					0.595		
5	F	MOV	$PERCLOS$	A/MOV	$T_{cl,2}$				0.602		
6	F	MOV	$PERCLOS$	A/MOV	$T_{cl,2}$	T_{ro}			0.610		
7	F	MOV	$PERCLOS$	A/MOV	$T_{cl,2}$	T_{ro}	ACV		0.602		
8	F	MOV	$PERCLOS$	A/MOV	$T_{cl,2}$	T_{ro}	T	MCV	0.599		
9	F	MOV	$PERCLOS$	A/MOV	$T_{cl,2}$	T_{ro}	T	MCV	ACV	0.589	
10	F	MOV	$PERCLOS$	A/MOV	$T_{cl,2}$	T_{ro}	T	MCV	ACV	T_c	0.583

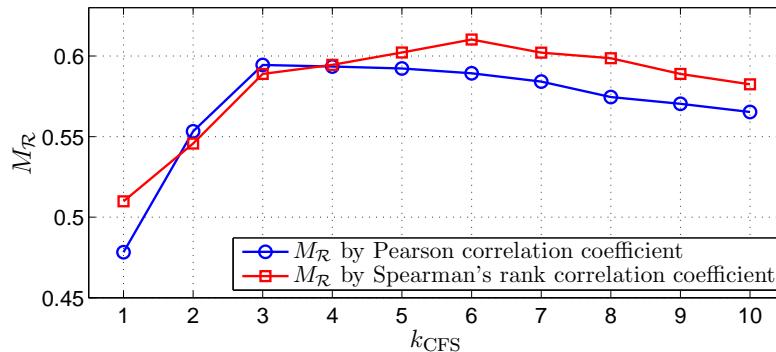


Figure 8.36.: M_R values of the best 10-feature combinations calculated based on the Pearson and Spearman's rank correlation coefficients

In order to be able to compare the results of the CFS with those of the SFFS method, we applied a binary subject-dependent ANN classifier with $N_h = 10$ to $k_{\text{CFS}} = 4$ drive time-based features selected using both Pearson and Spearman's rank correlation coefficients. The corresponding confusion matrices are shown in Table 8.22. Comparing these results with those of the SFFS method (Table 8.17) indicates degraded classification results for the drowsy class by at least 4% (70.2% and 68.6% vs. 74.3%). Overall, the selected feature combination by the filter approach is different from the subset selected by the SFFS method as a wrapper approach. The results of the SFFS approach outperform those of the CFS clearly due to the fact that in the SFFS approach, the classifier is directly involved in the selection of the best 4-feature combination. As a result, a poor performance for all classifier-independent methods can be expected due to their generality or underfitting.

Regardless of the value of k_{CFS} , Tables 8.23 and 8.24 show the best 10-feature combinations with respect to M_R out of all possible combinations for both correlation coefficients. Interestingly, the CFS method based on the Pearson correlation coefficient selected on average 4 features. However, the Spearman's rank correlation coefficient achieved the highest score by selecting 6 features on average. F and MOV are the only features which were selected in all sets of best feature combinations regarding M_R . Other selected feature subsets differ to some extent.

Table 8.22.: Confusion matrices of the binary subject-dependent ANN classifier ($N_h = 10$) for drive time-based features and $k_{\text{CFS}} = 4$. Left: Pearson correlation coefficient. Right: Spearman's rank correlation.

predicted			predicted				
	state	awake	drowsy		state		
given	awake	85.2%	14.8%	given	awake	86.5%	13.5%
	drowsy	29.8%	70.2%		drowsy	31.4%	68.6%

The only feature, which was selected by all of the introduced methods, is F . Hence, we conclude that F is a feature which, on the one hand, is associated with the KSS values and on the other hand, it is to a lesser extent correlated with other features. In addition, since it was also selected by the SFFS, it has a strong complementary property.

Table 8.23.: Best selected drive time-based features based on the CFS method using the Pearson correlation coefficient regardless of the number of features

rank	selected features					M_R
1	F	MOV	T_{ro}			0.5945
2	F	MOV	T_c	T_{ro}		0.5935
3	F	ACV	A/MOV	T_{ro}		0.5926
4	F	MCV	A/MOV	T_{ro}		0.5926
5	F	MOV	A/MOV	$PERCLOS$	T_{90}	0.5923
6	F	MOV	A/MOV	$PERCLOS$	T_{ro}	0.5914
7	F	MOV	T_c			0.5910
8	F	MOV	T_{90}	T_{ro}		0.5905
9	F	MOV	A/MOV	$PERCLOS$	T_{80}	0.5905
10	F	MOV	A/MOV	T_{ro}		0.5903

Table 8.24.: Best selected drive time-based features based on the CFS method using the Spearman's rank correlation coefficient regardless of the number of features

rank	selected features						M_R
1	F	A/MOV	$T_{cl,2}$	T_{ro}	$PERCLOS$	MOV	0.6103
2	F	A/MOV	$T_{cl,2}$	T_{ro}	$PERCLOS$	MCV	0.6026
3	F	A/MOV	$T_{cl,2}$	T_{ro}	$PERCLOS$	ACV	0.6023
4	F	A/MOV	$T_{cl,2}$	MOV	$PERCLOS$		0.6022
5	F	A/MOV	$T_{cl,2}$	T_{ro}	$PERCLOS$	MOV	0.6021
6	F	A/MOV	$T_{cl,2}$	T_{ro}	$PERCLOS$	AOV	0.6019
7	F	A/MOV	$T_{cl,2}$	T_{ro}	$PERCLOS$	MOV	0.6014
8	F	A/MOV	$T_{cl,2}$	T_{ro}	ACV	MOV	0.6011
9	F	A/MOV	$T_{cl,2}$	T_{ro}	$PERCLOS$	MOV	0.6003
10	F	A/MOV	$T_{cl,2}$	T_{ro}	MOV	T	0.6002

9. Summary, conclusion and future work

9.1. Summary and conclusion

Timely detection of a drowsy driver and warning him to make him aware of his low vigilance state plays an important role in improving the traffic safety. In this work, we addressed driver state classification based on blink features collected by the electrooculography as a reference measurement system.

In Chapter 1, we discussed different terminologies for defining drowsiness along with defining driver distraction and inattention. We also reviewed drowsiness countermeasures during driving such as conversing, rumble strips, etc. which, however, do not have a long-lasting effect on the vigilance. This chapter also provided an overview of drowsiness detection systems on the market.

Chapter 2 discussed objective and subjective methods for measuring driver state. The objective measures are either driving performance measures which monitor the driver indirectly based on the lane keeping behavior or steering wheel movements or a fusion of them. We reviewed that such measures suffer from external factors such as the quality of the lane marking, road condition, etc. In addition, their efficiency is restricted to the situations that the driver performance is not improved by other assistance systems. We also introduced driver physiological measures which are the result of direct monitoring of the driver, such as EEG, ECG, etc. We proposed the idea of removing phases of the driver, where the driver is visually distracted, to improve the association between an EEG-based measure and the drowsiness. At the end of this chapter, we also introduced different subjective measures and the concerns about their interpretation and reliability.

In Chapter 3, the human visual system was introduced. There, we mentioned the concepts of *what* and *where* which describe the visual attention. Further, the structure of the human eye and relevant types of eye movements during driving were defined. We also categorized eye movements into two groups with regard to their velocity, namely slow and fast eye movements, and showed that blinks can belong to both of these groups depending on the driver's vigilance state.

The robustness and reliability of the EOG measuring system for collecting eye movements during driving was tested in a pilot study in Chapter 4. There, we studied the relationship between driver eye movements and different real driving scenarios independent of driver's vigilance state in a fully controlled experiment conducted on a proving ground. All in all, it can be concluded that ground excitation and large amplitude bumps add an extra pattern to the EOG signals. On the other hand, monitoring driver eye movements seems to be undisturbed by a single small amplitude bump. Moreover, it is clear that the inevitable sawtooth pattern due to curve negotiation is not related to the driver's inattention or drowsiness. Therefore, we suggested the exclusion of tortuous road sections for further investigation of driver eye movements.

Since the capability of the EOG as a robust and reliable reference measuring system for eye movement monitoring even under real driving conditions was acknowledged by the pilot study, we

conducted daytime and nighttime experiments under real road and simulated driving conditions using EOG to collect eye movement data for the rest of this work as described in Chapter 4.

In chapter 5, we addressed the detection approaches of blinks and saccades in the raw EOG signals. We showed that the median filter-based method as the most conventional blink detection approach was only suitable for the detection of blinks during the awake phase of driving. As soon as, the shape of blinks changed due to drowsiness, this method either missed an event or detected only part of it. In addition, we showed that the median filter-based method was not suitable for detecting saccades and slow eye movements. As a result, we proposed a method based on the derivative of the EOG signal for detecting saccades in addition to the blinks. It was shown how to detect vertical saccades and blinks simultaneously in vertical EOG signal. In addition, a 3-means clustering algorithm was recommended to distinguish between saccades and blinks in those applications where the data of both awake and drowsy phases are available. This helped to prevent confusing a driver's decreased amplitude blinks with saccades or other eye movements. Moreover, blinks with long eye closure and microsleep events, whose patterns deviated from those during the awake phase, were detected and distinguished from saccades based on the statistical distribution of the amplitude. This method, however, was shown to perform poorly in the detection of slow eye movements. Therefore, we introduced the wavelet transform method which was superior to the Fourier transform in providing time localization information. In addition to the continuous wavelet transform for detection of both fast and slow eye movements, we applied the discrete wavelet transform as a suitable method for preprocessing of the EOG signal, namely drift and noise removal. Finally, comparison of the detection methods showed that the proposed derivative-based algorithm outperformed the method based on median filtering in detection of fast eye movements. Although the wavelet transform method performed best in the correct detection of both fast and slow eye movements, it suffered from high false detection rates. Consequently, we combined its detected events with those of the derivative-based algorithm to balance false detections.

In Chapter 6, we studied blinking behavior under distracted and undistracted driving. In the first experiment, during which the subjects performed a secondary visuomotor task in addition to the driving task, we showed that saccades and gaze shifts induced the occurrence of blinks. However, we observed two different behaviors among subjects, direction dependent and direction-independent gaze shift-induced blinks. For the former group, performing the secondary task (either visuomotor or auditory) did not alter the blink rate in comparison to the undistracted driving. However, for the latter, the blink rate changed to a large extent due to distraction. In addition, we showed that visual distraction led to a blinking time interval synchronous with the occurrence of the gaze shift. In a second experiment, during which the subjects were not distracted, the results represented that the amount of gaze shift was positively correlated with the occurrence of a simultaneous blink, i.e. the higher the amplitude of the gaze shift, the larger was the probability of the blink occurrence. Based on these results, we suggested those, who consider blink rate as an indicator of drowsiness, to handle gaze shift-induced blinks differently to spontaneous ones, particularly if the driver is visually or cognitively distracted. In fact, since such blinks are situation-dependent, they locally change the blinking behavior, especially the blink frequency.

Based on the detected events of Chapter 5, in Chapter 7, we extracted 19 different features for each event. These features were aggregated by considering two strategies, namely KSS input-based and drive time-based approaches. Unlike the latter, the former sacrificed the available number of samples for more reliable class labels. In addition, feature baselining was addressed to improve the classification results in Chapter 8. Further, in this chapter, based on the scatter

plots and correlation coefficients between features and KSS values, we showed whether features were positively or negatively associated with the driver state. Interestingly, for some of the features, such as the blink amplitude, different trends were observed. Thus, we conclude that a warning system, which relies only on a single feature for its decision strategy, is prone to high false alarm rates. This chapter also discussed the variation of each feature shortly before the occurrence of the first safety-critical event, namely a lane departure and a microsleep, in comparison to the beginning of the drive. The results showed an important finding regarding driving performance measures. For the lane departure event, overall, a larger variation of the features shortly before the event was found in comparison to the microsleep event. This proves for our data set that a drowsy driver experiences a microsleep event without necessarily departing the lane or degraded driving performance measures. In other words, this finding acknowledges the fact that lane departure might be related to a deeper drowsiness phase in comparison to the microsleep. As a result, from this aspect, driver physiological measures are superior to driving performance measures for the early driver drowsiness detection. Finally, in this chapter we degraded the sampling frequency of the raw EOG signal to make it similar to the raw signal provided by the driver observation cameras on the market. The goal was to study the effect of sampling frequency on the feature quality. According to the results, we conclude that velocity-based features are at high risk of quality degradation.

Finally, in Chapter 8, we classified the driver state by three classifiers, namely ANN, SVM and k -NN classifiers based on the extracted features in Chapter 7. The feature matrix was divided either by a subject-dependent or a subject-independent approach. We also addressed the issue of imbalanced data based on classifier-dependent approach. According to the results, for the binary subject-dependent classification, all classifiers performed similarly regarding drive time-based features. This was also valid for binary subject-independent classifiers. In the binary subject-dependent case, we obtained at least about 80% correctly classified samples in each class regardless of the selected classifier. The binary subject-independent classifiers, however, performed poorly in the classification of drowsy samples. In fact, the classification of unseen drowsy samples seems to be more challenging. In the 3-class classification, the ANN and SVM classifiers performed poorer in the detection of the medium class in comparison to the k -NN. We believe that this is due to imprecise class labels and that the subjects were not good at rating medium levels of drowsiness.

For KSS input-based features, we achieved slightly better classification results by the binary SVM classifier in comparison to the ANN and k -NN. The high detection rate of both classes (each around 80%) by this aggregation approach also underlines that in self-rating, subjects most likely take the time interval shortly before the KSS inquiry into account when rating themselves.

For imbalanced class distributions, it was shown that all classifiers performed poorly to the same extent in the classification of the minority class. We solved the issue of imbalanced data with two approaches. The first one, as a classifier-independent method, was the SMOTE which artificially generated additional samples similar to those of the minority class. We combined it with the ANN classifier and obtained improved classification results. The retrained ANN, however, was not applicable to unseen data. For the SVM classifier, we applied a classifier-dependent approach where the misclassification cost was tuned with respect to the number of samples in the minority and majority classes. Again, despite improved results with imbalanced data, the constructed model was poor on unseen data. Therefore, we conclude that imbalanced class distributions in the task of driver state classification do not lead to a generalized classifier and should never be considered as a substitution for the minority class data collection. In other words, the results of driver state classification are reliable only if the features of both awake and drowsy phases of

the drive are collected under similar circumstances and are included in the feature matrix in a balanced manner.

Chapter 8 also discussed the generalization aspects of the data collected in the driving simulator to the real road conditions. There, we showed that by removing very drowsy parts of the drive, which can only be collected in simulated driving, still both classes were detectable to 70% given the subject-dependent binary classification with drive time-based features. As soon as the subject-independent data division was applied, the results degraded. The k -NN classifier was most affected due to dominant within-subject differences. We also applied the unseen features of the drop-out subjects, who aborted the real nighttime driving experiment due to severe drowsiness, to all classifiers. They were classified with acceptable results given the drive time-based features (accuracy = 70%). Therefore, we conclude that the drowsiness behavior in the driving simulator is to an acceptable extent representative of the same behavior on the real road. Overall, the between-subject differences also have a significant contribution on the degradation of the classification results.

Finally, we discussed approaches for feature dimension reduction in order to address the issues of an in-vehicle warning system. According to the SFFS method fused with the ANN classifier, four features were determined to be sufficient. The trained ANN classifier, however, did not perform as good as a classifier trained with all 19 features in the detection of the drowsy class. As a result, we conclude that for the correct detection of the drowsy class, which seems overall to be more challenging, more than four features are needed.

9.2. Perspective of future work

We showed that blink features based on the EOG are a promising approach for the driver state detection. Nevertheless, in this section, we suggest possible directions for the future work.

The first issue is the EOG as a reference measuring system which should be substituted with a driver observation camera for having an in-vehicle warning system based on eye movements. It should be investigated, if similar results can be achieved by the camera even with low frame rates. Since cameras also measure the eyelid gap, an improvement of classification results is expected despite degraded quality for some features. In addition, after replacing EOG with camera, new problems arise which degrade the eye tracking process due to other factors. Examples of such issues are varying light conditions and reflections arising from wearing glasses.

In this work, we used 1 min of the EOG data for aggregating features. It is required to investigate how the variation of this time interval improves or degrades the relationship between features and driver state.

A third issue is concerned with the poor detection rate results of the medium class in the 3-class classification as shown in Chapter 8. In future, it should be scrutinized whether imprecise self-rating by subjects is responsible for poor classification results or whether the eye movement features themselves cannot reflect the evolution of drowsiness at a lower level.

The forth issue is the fusion of the introduced blink features with other features such as the saccade features or features based on the driving performance measures. Moreover, features like traffic density or monotony, time of day and time-on-task can be integrated to contribute to the classification task. Further, the combination of saccade occurrence with the traffic density as a new feature appears to be promising for driver state classification in terms of the short term variation or detection of the driver distraction.

Finally, the similarity of findings of this work and its extension to autonomous car driving needs to be studied. In partially automated driving, it is assumed that the vehicle performs steering and lane keeping activities, while being fully observed by the driver for a timely intervention. In this case, in addition to driver drowsiness, the level of driver attention or distraction is indeed crucial. In highly automated driving, driver distraction and attention detection is even more essential, because the driver is allowed to be distracted by turning his attention to other activities. In complex situations, however, the driver must still be able to take over the driving task after receiving a warning. Therefore, on top of the blink behavior studied in distracted driving in this work, new features such as the gaze direction and proper gaze shifts to the road ahead should be extracted and explored. Moreover, new experiments and analyses should be conducted to quantify the amount of workload for investigation of driver distraction and attention detection.

A. Derivation of sawtooth occurrence frequency during curve negotiation

Figure A.1 geometrically represents the scenario during which the vehicle (subject) moves from position A to B while tracking TP_A and TP_B , successively.

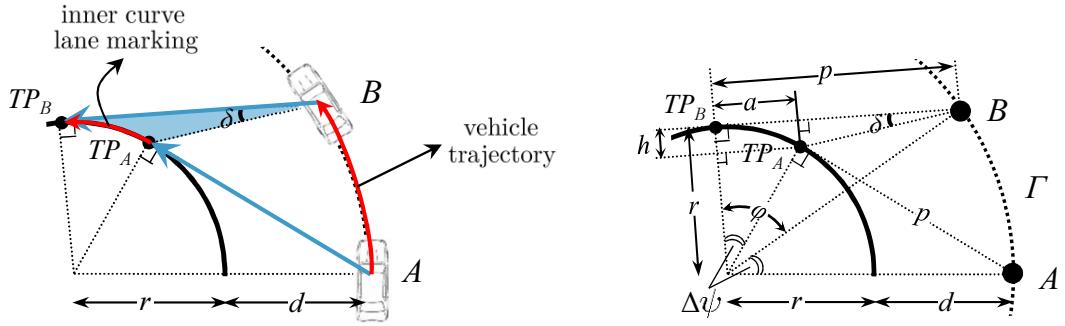


Figure A.1.: Geometrical representation of tracking two successive TPs during a curve negotiation

According to both plots of Figure 4.8, the measured time interval Δt_m between successive sawtooth patterns is very short (on average < 1 s). Thus, it can be assumed that the vehicle's lateral distance d to the inner curve lane marking in Figure A.1 remains constant during tracking two successive TPs, namely TP_A and TP_B . We assumed $d = 1.5$ m. For the same reason, we considered the radius r of the curve and the distance p between subject and the momentary TP to be constant.

Since the longitudinal acceleration is assumed negligible during tracking two successive TPs, Δt_c can be calculated from the velocity v and the displacement Γ between A and B as follows

$$\Delta t_c = \frac{\Gamma}{v} = \frac{(r + d) \Delta \psi}{v}, \quad (\text{A.1})$$

where $\Delta\psi$ is the yaw angle corresponding to the displacement Γ . r has been calculated out of road curvature κ ($\kappa = 1/r$) which is a function of measured v and yaw angle rate $\dot{\psi}$ ($\kappa = \dot{\psi}/v$). Since $\dot{\psi}$ was not equal for all subjects, the calculated value of κ and consequently r will be different for the left curve of Figure 4.1. Therefore, r is assumed to be the mean over all calculated radii for all subjects which corresponds to 52 m.

Based on our assumptions and the geometrical modeling of Figure A.1, the unknown value of angular displacement of subject's position $\Delta\psi$ leads to the angular displacement of the eyes δ . According to Figure A.1, the angular displacement of eyes δ is as follows

$$\delta = \arctan\left(\frac{h}{p - a}\right), \quad (\text{A.2})$$

where

$$h = r - r \cos(\Delta\psi) \quad (\text{A.3})$$

$$a = r \sin(\Delta\psi) \quad (\text{A.4})$$

and the distance p between the driver and the momentary TP is

$$p = \text{const.} = r \tan \varphi = r \tan \left(\arccos \left(\frac{r}{r+d} \right) \right). \quad (\text{A.5})$$

By substituting (A.3), (A.4) and (A.5) in (A.2), the angular displacement of eyes δ can be described as a function of yaw angle $\Delta\psi$ as the following

$$\delta = \arctan \left(\frac{1 - \cos(\Delta\psi)}{\tan \varphi - \sin(\Delta\psi)} \right). \quad (\text{A.6})$$

$\Delta\psi$ can be derived from (A.6) by knowing δ . Then, Δt_c can be approximated using the calculated $\Delta\psi$ in (A.1).

B. Description of a boxplot

Figure B.1 shows the information included in the boxplot representation. In this work, all boxplots are shown for $w = 1.5$.

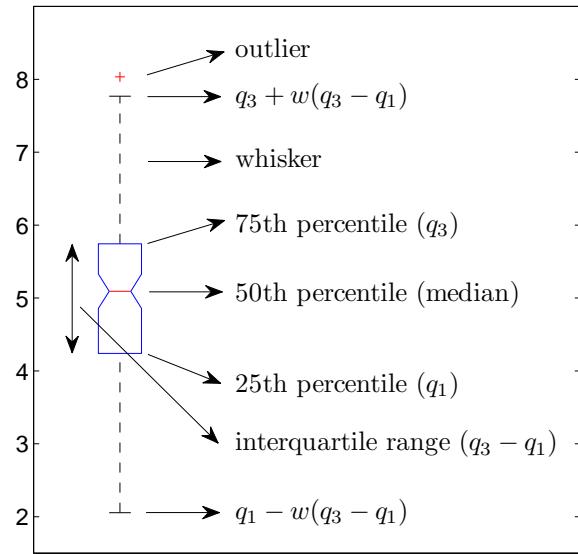


Figure B.1.: Description of a boxplot representation

C. k -means clustering

The material is taken from Bishop (2006).

The k -means clustering method aims to assign samples of a data set to different groups or *clusters*. Therefore, N samples of a matrix \mathbf{X} with D dimensions are assigned to a given number of clusters k . As a rule, the samples assigned to a cluster have a small inter-sample distance in comparison to the samples outside a cluster. We define $\boldsymbol{\mu}_i$, where $i \in \{1, \dots, k\}$, as the center of k clusters. The rule, which assigns samples of \mathbf{X} to a specific cluster, is based on the minimization of sum of squares of the distances between a sample \mathbf{x}_n , where $n \in \{1, \dots, N\}$, and the nearest $\boldsymbol{\mu}_i$. For each sample \mathbf{x}_n , we define the binary value $r_{ni} \in \{1, 0\}$. It shows which cluster i the sample \mathbf{x}_n is assigned to. If \mathbf{x}_n is assigned to the i -th cluster, then we have $r_{ni} = 1$ and $r_{nj} = 0$, where $i \neq j$. The cost function, called *distortion measure*, for guiding the cluster assignment is defined as follows

$$J = \sum_{n=1}^N \sum_{i=1}^k r_{ni} \|\mathbf{x}_n - \boldsymbol{\mu}_i\|^2. \quad (\text{C.1})$$

The goal is to minimize J with respect to r_{ni} and $\boldsymbol{\mu}_i$. This goal is achieved in an iterative manner comprising the following two successive steps:

- Step 1: minimizing J with respect to r_{ni} , while $\boldsymbol{\mu}_i$ is held fixed.
- Step 2: minimizing J with respect to $\boldsymbol{\mu}_i$, while r_{ni} is held fixed.

These steps correspond to the Expectation Maximization (EM) algorithm such that step 1 and 2 refer to the expectation (E) and maximization (M) step, respectively.

In step 1, where $\boldsymbol{\mu}_i$ is held fixed, the relationship between J and r_{ni} in (C.1) is a linear one. Therefore, by assigning the sample \mathbf{x}_n to the nearest cluster i , J is minimized. Mathematically, we have

$$r_{ni} = \begin{cases} 1 & \text{if } i = \arg \min_j \|\mathbf{x}_n - \boldsymbol{\mu}_j\|^2 \\ 0 & \text{otherwise.} \end{cases} \quad (\text{C.2})$$

In step 2, however, with r_{ni} held fixed, the relationship between J and $\boldsymbol{\mu}_i$ in (C.1) is quadratic. Therefore, for minimizing J , the derivative of J with respect to $\boldsymbol{\mu}_i$ is set to zero as follows

$$2 \sum_{n=1}^N r_{ni} (\mathbf{x}_n - \boldsymbol{\mu}_i) = 0 \implies \boldsymbol{\mu}_i = \frac{\sum_{n=1}^N r_{ni} \mathbf{x}_n}{\sum_{n=1}^N r_{ni}}. \quad (\text{C.3})$$

The denominator in previous equation denotes the number of samples assigned to the i -th cluster. Therefore, according to (C.3), the center of the i -the cluster $\boldsymbol{\mu}_i$ is defined as the mean of all samples \mathbf{x}_n assigned to that cluster. To this account, the method is called k -means clustering.

The aforementioned steps are repeated as long as the assignment of samples to clusters changes or a maximum number of iterations is reached. Since after each iteration, J is minimized, the cost function definitely converges. A drawback of the k -means clustering, however, is that

it might wrongly converge to a local minimum of J instead of the global one. Moreover, as we mentioned, this method uses the squared Euclidean distance for quantifying the amount of dissimilarity between samples and μ_i . Consequently, the Euclidean distance between all N samples and all cluster centers have to be calculated which leads to a slow convergence of this method. For further optimization ideas with respect to the convergence rate of the k -means clustering see Bishop (2006).

Obviously, before performing step 1, μ_i have to be initialized and the chosen initial values have a direct impact on the convergence rate of the algorithm. Bishop (2006) suggests setting μ_i randomly to a subset of k samples of \mathbf{X} .

D. Statistical tests

The material provided here are taken from Gosling (1995), Montgomery and Rungar (2006) and Field (2007).

D.1. Paired-sample t -test

The *Paired-sample t -test* or *dependent t -test* is a suitable test for comparing the mean of two groups, if the samples are measured as pairs. In other words, the samples should be collected under “*homogeneous conditions*” (Montgomery and Rungar, 2006). Mathematically, the pairs are $(x_{1,i}, x_{2,i})$ with $i = 1, 2, \dots, N$ samples. In the case that the test is applied to the samples collected in an experiment with different participants, similar participant should have been participated in both groups. This means that samples $x_{1,i}$ and $x_{2,i}$ belong to one participant, otherwise *pooled t -test* or *independent t -test* should be applied. Moreover, the results of the t -test are reliable, if its assumption is fulfilled which is the normal distribution of samples. Since t -test analyzes the difference between groups, i.e. $\Delta\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$, the samples of $\Delta\mathbf{x}$ should be normally distributed.

The hypotheses of the test are as follows:

$$\begin{aligned} H_0: \mu_1 &= \mu_2 \\ H_1: \mu_1 &\neq \mu_2, \end{aligned}$$

where μ_1 and μ_2 refer to the average of \mathbf{x}_1 and \mathbf{x}_2 , respectively.

The test statistic is found as follows

$$t_0 = \frac{\mu_{\Delta\mathbf{x}}}{\frac{\sigma_{\Delta\mathbf{x}}}{\sqrt{N}}}, \quad (\text{D.1})$$

where $\mu_{\Delta\mathbf{x}}$ and $\sigma_{\Delta\mathbf{x}}$ denote the mean and the standard deviation of the difference values in $\Delta\mathbf{x}$. The decision about the rejection of H_0 is made based on the value of the confidence level α and the degrees of freedom ν , where $\nu = N - 1$. The critical value of the test, namely $t_{\alpha,\nu}$, is found in the table of the Student’s t -distribution in statistic books such as Montgomery and Rungar (2006). Consequently, we have

$$\begin{aligned} \text{reject } H_0: & \text{ if } |t_0| > t_{\alpha,\nu} \implies \mu_1 \text{ differs significantly from } \mu_2. \\ \text{fail to reject } H_0: & \text{ if } |t_0| \leq t_{\alpha,\nu}. \end{aligned}$$

The results can also be reported by the p -value based on the Student’s t cumulative distribution function. Accordingly, p -values $< \alpha$ yields the rejection of H_0 .

D.2. Normal distribution test: Lilliefors test

In order to know whether the distribution of the data under investigation is normal, *Lilliefors test* is applied which is a *goodness of fit test*. This test determines the normality of the current data based on fitting a normal distribution to it and evaluating the difference between them. Therefore, for data \mathbf{x} with N values, the hypotheses are defined as follows

$$H_0: \mathbf{x} \text{ is normally distributed.}$$

$$H_1: \mathbf{x} \text{ is not normally distributed.}$$

First, the empirical cumulative distribution function $\hat{F}_N(x)$ of \mathbf{x} is calculated as follows

$$\hat{F}_N(x) = \frac{\text{number of samples in } \mathbf{x} \leq x}{N}. \quad (\text{D.2})$$

The goal is to assess the agreement between $\hat{F}_N(x)$ and the theoretical distribution function $F(x)$. $F(x)$ is normally distributed with regard to the mean and standard deviation of \mathbf{x} , namely μ and σ . As a result, the standard values of \mathbf{x} are needed as follows

$$z = \frac{x - \mu}{\sigma}. \quad (\text{D.3})$$

The test statistic is defined as

$$t_0 = \max(|F(z) - \hat{F}_N(z)|). \quad (\text{D.4})$$

By comparing t_0 with $t_{N,\alpha}$ denoting the critical value of the test, it can be decided whether to reject the H_0 or not, i.e.

reject H_0 : if $t_0 > t_{N,\alpha} \implies \mathbf{x}$ is not normally distributed.

fail to reject H_0 : if $t_0 \leq t_{N,\alpha}$,

where α is the confidence level. The values of $t_{N,\alpha}$ are listed in Gosling (1995).

D.3. Test of significance for the Pearson correlation coefficient

Based on a hypothesis test, namely the t -test, it is possible to show whether the Pearson product-moment correlation coefficient ρ_p between two variables is significantly different from zero. If $\rho_p = 0$, then we conclude that there is no linear relationship between the variables under investigation. Therefore, the hypotheses are defined as follows

$$H_0: \rho_p = 0 \implies \text{There is no linear relationship between variables.}$$

$$H_1: \rho_p \neq 0.$$

According to Field (2007), the test static t_{ρ_p} with $N - 2$ degrees of freedom for N samples of the variables is calculated as follows

$$t_{\rho_p} = \rho_p \sqrt{\frac{N - 2}{1 - \rho_p^2}}. \quad (\text{D.5})$$

By comparing t_{ρ_p} with $t_{\alpha,\nu}$, which denotes the critical value of the test based on the Student's t -distribution for $\nu = N - 2$ degrees of freedom, it can be decided whether to reject the H_0 or not, i.e.

reject H_0 : if $|t_{\rho_p}| > t_{\alpha,\nu} \implies \rho_p \neq 0$ and ρ_p is significantly different from zero.
fail to reject H_0 : if $|t_{\rho_p}| \leq t_{\alpha,\nu}$,

where α is the confidence level. The values of $t_{\alpha,\nu}$ are listed in Gosling (1995).

Accordingly, based on the calculated t_{ρ_p} value, the p -value of the test can be reported as well. p -value $< \alpha$ also yields the rejection of H_0 .

D.4. Comparison of two Pearson correlation coefficients

According to Field (2007), based on the t -statistic, it is possible to assess whether two Pearson correlation coefficients are significantly different from each other or not. Mathematically, three variables are available, namely \mathbf{x} , \mathbf{y} and \mathbf{z} , and the relationship between two pairs, namely $\rho_p(\mathbf{x}, \mathbf{y})$ and $\rho_p(\mathbf{z}, \mathbf{y})$, are of interest. For the ease of notation, we replace $\rho_p(\mathbf{x}, \mathbf{y})$ with $\rho_{\mathbf{xy}}$, since just the Pearson correlation coefficient is studied here.

The hypotheses of the test are as follows

$$H_0: \rho_{\mathbf{xy}} \text{ is not different from } \rho_{\mathbf{zy}}. \quad (\text{D.6})$$

$$H_1: \rho_{\mathbf{xy}} \text{ is different from } \rho_{\mathbf{zy}}. \quad (\text{D.7})$$

The test significance is calculated as follows

$$t_{\text{difference}} = (\rho_{\mathbf{xy}} - \rho_{\mathbf{zy}}) \sqrt{\frac{(N - 3)(1 + \rho_{\mathbf{xz}})}{2(1 - \rho_{\mathbf{xy}}^2 - \rho_{\mathbf{xz}}^2 - \rho_{\mathbf{zy}}^2 + 2\rho_{\mathbf{xy}}\rho_{\mathbf{zy}}\rho_{\mathbf{xz}})}}. \quad (\text{D.8})$$

By comparing the value of $t_{\text{difference}}$ with the critical test value provided in the table of t -distribution for $\nu = N - 3$ degrees of freedom and the confidence level of α , namely $t_{\alpha,\nu}$, following decisions are made, which are similar to the paired-sample t -test in Appendix D.1,

reject H_0 : if $|t_{\text{difference}}| > t_{\alpha,\nu} \implies \rho_{\mathbf{xy}}$ differs significantly from $\rho_{\mathbf{zy}}$.
fail to reject H_0 : if $|t_{\text{difference}}| \leq t_{\alpha,\nu}$.

The result can also be reported by the p -value based on the Student's t cumulative distribution function. Accordingly, p -values $< \alpha$ yields the rejection of H_0 .

D.5. One-way repeated measures ANOVA

The analysis of variance (ANOVA) is a technique for comparing means of different groups. If the measurements of the groups are related to different participants, it is then called *one-way independent* ANOVA. If, however, similar participants are available, i.e. several measurements for each subject, then *one-way repeated measures* ANOVA is used instead. This method concentrates on the within-subject differences and is, in fact, an extension of the paired-sample t -test explained

in Appendix D.1. On the contrary, one-way independent ANOVA focuses on the between-group differences.

Before applying the ANOVA for repeated measurements, following assumptions are required to be fulfilled:

- Normally distributed group differences.
- Sphericity, which is equivalent to the homogeneity of the variance. The variances of group differences should almost be the same. According to Field (2007), by applying the *Mauchly's test*, this assumption can be checked.
- Independent samples.

For a data set with G groups and N subjects, which fulfills all aforementioned assumptions, the hypotheses of the ANOVA are as follows

$$H_0: \mu_1 = \mu_2 = \dots = \mu_G$$

H_1 : at least one μ is different from the other ones.

μ_i refers to the mean of the i -th group. It is clear that, in the case of H_1 , the test does not provide information about the group or groups with different mean values. For the ease of notation, we consider the data set listed in Table D.1.

Subjects	Groups				
	1	2	...	G	mean
1	x_{11}	x_{12}	...	x_{1G}	\bar{x}_1
2	x_{21}	x_{22}	...	x_{2G}	\bar{x}_2
:	:	:	:	:	:
N	x_{N1}	x_{N2}	...	x_{NG}	\bar{x}_N
mean	μ_1	μ_2	...	μ_G	$\bar{\mu}$

Table D.1.: Typical data set of one-way ANOVA

First, the *groups variability*, called $SS_{between}$, is calculated as the summed square of the deviation of the group mean μ_i from the overall mean $\bar{\mu}$ of the data set

$$SS_{between} = \sum_{i=1}^G N(\mu_i - \bar{\mu})^2. \quad (\text{D.9})$$

The next step is the calculation of *within-subject variation* called SS_{within}

$$SS_{within} = \sum_{j=1}^G \sum_{i=1}^N (x_{ij} - \mu_j)^2. \quad (\text{D.10})$$

In the case of repeated measurements, each subject must also be considered separately. Therefore, the summed square of the deviation of the subject's mean \bar{x}_i from the overall mean $\bar{\mu}$ is needed, i.e.

$$SS_{subjects} = \sum_{i=1}^N G(\bar{x}_i - \bar{\mu})^2. \quad (\text{D.11})$$

\bar{x}_i refers to the mean of the samples for the i -th subject, namely

$$\bar{x}_i = \frac{1}{G} \sum_{j=1}^G x_{ij}. \quad (\text{D.12})$$

SS_{within} also includes $SS_{subjects}$. Hence, the error is defined as follows

$$SS_{error} = SS_{within} - SS_{subjects}. \quad (\text{D.13})$$

Now, the means of sum of squares are calculated by considering the degrees of freedom ν_1 and ν_2 as follows

$$MS_{between} = \frac{SS_{between}}{\nu_1} \quad (\text{D.14})$$

$$MS_{error} = \frac{SS_{error}}{\nu_2} \quad (\text{D.15})$$

$$\nu_1 = G - 1 \quad (\text{D.16})$$

$$\nu_2 = (N - 1)(G - 1). \quad (\text{D.17})$$

Finally, the test statistic F_0 is calculated as

$$F_0 = \frac{MS_{between}}{MS_{error}}. \quad (\text{D.18})$$

By comparing F_0 with the critical value of F -distribution with respect to ν_1 and ν_2 , namely F_{α, ν_1, ν_2} , following decision is made:

reject H_0 : if $F_0 > F_{\alpha, \nu_1, \nu_2} \implies$ At least one μ is significantly different.

fail to reject H_0 : if $F_0 \leq F_{\alpha, \nu_1, \nu_2}$.

α refers to the confidence level.

D.6. Homogeneity of variance: Levene's test

The homogeneity of variance is an assumption of the ANOVA which investigates whether the variance and the spread of values in each group are in the same range or not. This is done based on the *Levene's test*. We use the same notation as in Appendix D.5.

The hypotheses of the test are as follows:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_G^2$$

$$H_1: \text{at least one } \sigma^2 \text{ is different from the other ones.}$$

σ_i^2 denotes the variance of the i -th group.

We define the following variables

$$\begin{aligned} z_{ij} &= |x_{ij} - \mu_j| \\ \mu_{z_j} &= \frac{1}{N} \sum_{i=1}^N z_{ij} \\ \mu_z &= \frac{1}{NG} \sum_{i=1}^N \sum_{j=1}^G z_{ij}. \end{aligned}$$

Based on these definitions the test statistic is defined which is equivalent to applying one-way ANOVA to z_{ij} with i going from 1 to N samples and j going from 1 to G groups. Therefore, we have

$$F_0 = \frac{G(N-1)}{G-1} \frac{\sum_{j=1}^G N(\mu_{z_j} - \mu_z)^2}{\sum_{j=1}^G \sum_{i=1}^N (z_{ij} - \mu_{z_j})^2}. \quad (\text{D.19})$$

The critical value of the test and the conditions for rejecting/not rejecting H_0 are similar to the one-way repeated measures ANOVA with the degrees of freedom ν_1 and ν_2

$$\nu_1 = G - 1 \quad (\text{D.20})$$

$$\nu_2 = G(N-1). \quad (\text{D.21})$$

Therefore, we have

reject H_0 : if $F_0 > F_{\alpha, \nu_1, \nu_2} \implies$ At least one σ^2 is significantly different.

fail to reject H_0 : if $F_0 \leq F_{\alpha, \nu_1, \nu_2}$.

α refers to the confidence level.

D.7. Wilcoxon signed-rank test

As a non-parametric statistical test, the *Wilcoxon signed-rank test* analyzes whether the means of two paired samples are significantly different from each other. This test is mainly applied, if the assumption of the paired-sample t -test explained in Appendix D.1, namely the normal distribution of samples, is violated. This assumption Therefore, it is important to have similar participants in both groups, since this test studies within-subject differences.

For the paired samples $(x_{1,i}, x_{2,i})$ with $i = 1, 2, \dots, N$ and the corresponding mean values of μ_1 and μ_2 , H_0 and H_1 hypotheses are as follows:

$$\begin{aligned} H_0: \mu_1 &= \mu_2 \\ H_1: \mu_1 &\neq \mu_2. \end{aligned}$$

Test procedure:

1. Rank the absolute value of the difference between pairs $|\Delta x_i|$, where $\Delta x_i = x_{1,i} - x_{2,i}$, in the ascending order.

2. Give each rank the same sign as the corresponding Δx_i . For samples with similar difference values, the average of the ranks should be considered.
3. Find $W = \min(W_+, W_-)$, where W_+ and W_- denote the sum of positive and negative ranks, respectively, while considering the absolute value of each rank in the summation.
4. Calculate the following mean and variance values, namely \bar{W} and $V_{\bar{W}}$,

$$\bar{W} = \frac{N(N+1)}{4}$$

$$V_{\bar{W}} = \sqrt{\frac{N(N+1)(2N+1)}{24}}.$$

5. Calculate the z -score of W

$$z_0 = \frac{W - \bar{W}}{V_{\bar{W}}} . \quad (\text{D.22})$$

6. The test result is as follows

reject H_0 : if $|z_0| > w_\alpha^*$ $\implies \mu_1$ differs significantly from μ_2 .
fail to reject H_0 : if $|z_0| \leq w_\alpha^*$.

w_α^* with α as the confidence level is the critical value of the test which is listed in statistic books such as in Montgomery and Runger (2006).

Alternatively, the p -value of the test can be calculated based on the normal cumulative distribution function. Thus, H_0 is rejected, if p -value $< \alpha$.

D.8. Pearson's chi-square test

The *Pearson's chi-square test* is a statistical test for analyzing the relationship between observed categorical data with respect to the *chi-squared* (χ^2) distribution. If the categories interact with each other, then we conclude that they are dependent, because the occurrence of one event leads to the occurrence of the other one. Categorical data can be represented by the *contingency table* which summarizes the scores with respect to their membership in each category, as shown in Table D.2 for two categories.

Table D.2.: Contingency table with two categories

		condition A		total
		events	no	
condition B	yes	x_{11}	x_{12}	$x_{11} + x_{12}$
	no	x_{21}	x_{22}	$x_{21} + x_{22}$
	total	$x_{11} + x_{21}$	$x_{12} + x_{22}$	N

The test hypotheses are as follows

H_0 : The observed samples are statistically independent.

H_1 : The observed samples are statistically dependent.

The test statistic χ_0^2 is calculated as follows

$$\chi_0^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(x_{ij} - E_{ij})^2}{E_{ij}}, \quad (\text{D.23})$$

where x_{ij} is the observation summarized in Table D.2 with c columns and r rows. E_{ij} denotes the expected frequency for each member of the contingency table under the independency assumption and is calculated as

$$E_{ij} = \frac{\left(\sum_{k=1}^c x_{ik}\right)\left(\sum_{q=1}^r x_{qj}\right)}{N}. \quad (\text{D.24})$$

N refers to the total number of scores.

By comparing χ_0^2 with the critical value $\chi_{\alpha,\nu}^2$ with $\nu = (r-1)(c-1)$ degrees of freedom and confidence level of α , it is decided whether to reject the H_0 as follows

- reject H_0 : if $\chi_0^2 > \chi_{\alpha,\nu}^2 \implies$ Events occur independently.
- fail to reject H_0 : if $\chi_0^2 \leq \chi_{\alpha,\nu}^2$.

The result can also be reported by the p -value based on the χ^2 cumulative distribution function. Accordingly, p -values $< \alpha$ yields the rejection of H_0 .

E. Mother wavelets

Figure E.1 shows the scaling and wavelet functions of Haar and db4.

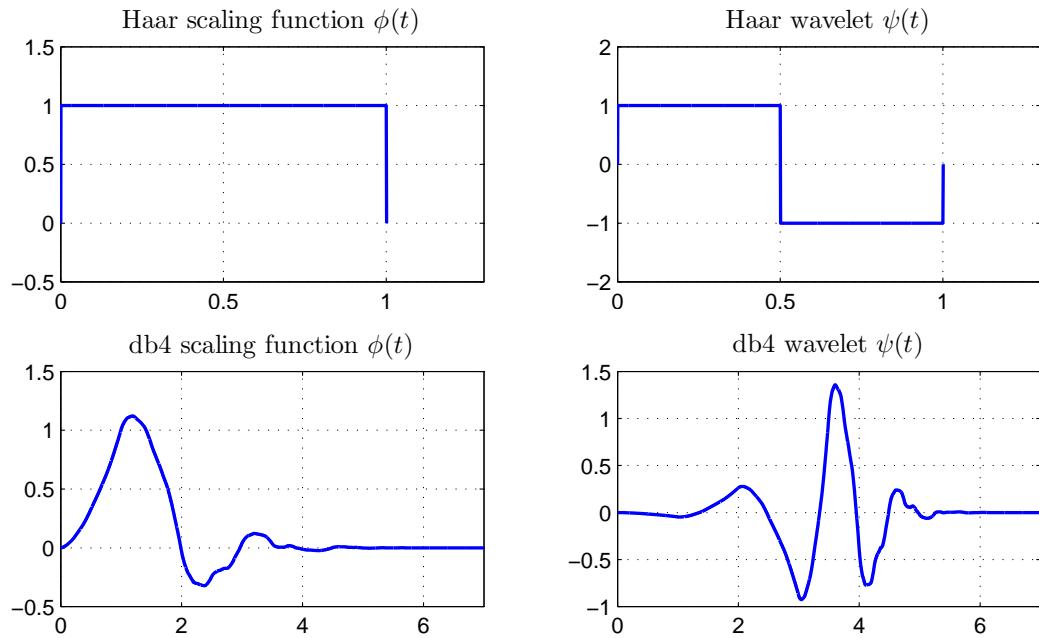


Figure E.1.: Scaling and wavelet functions of two mother wavelets

F. Additional results

F.1. Analysis of statistical measures of features

Figure F.1 shows the Spearman's rank correlation coefficient ρ_s between the statistical metrics of 18 drive time-based features and KSS values. Since frequency F is not affected by different metrics, it is not included. In addition to the mean, which was used in this work, following statistical metrics were calculated: standard deviation (std), median, minimum (min), maximum (max), range, defined as max – min, and root mean square (RMS), namely

$$\text{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}, \quad (\text{F.1})$$

where n denotes the number of events in an extraction window. All of these metrics were calculated for the events detected within an extraction window. Moreover, they were baselined afterwards to filter out individual differences. The missing bars could not be calculated. ρ_s values were not significant ($p\text{-value} > 0.001$) for: std, max and range for A , range for MCV and AOV and min for T_{80} .

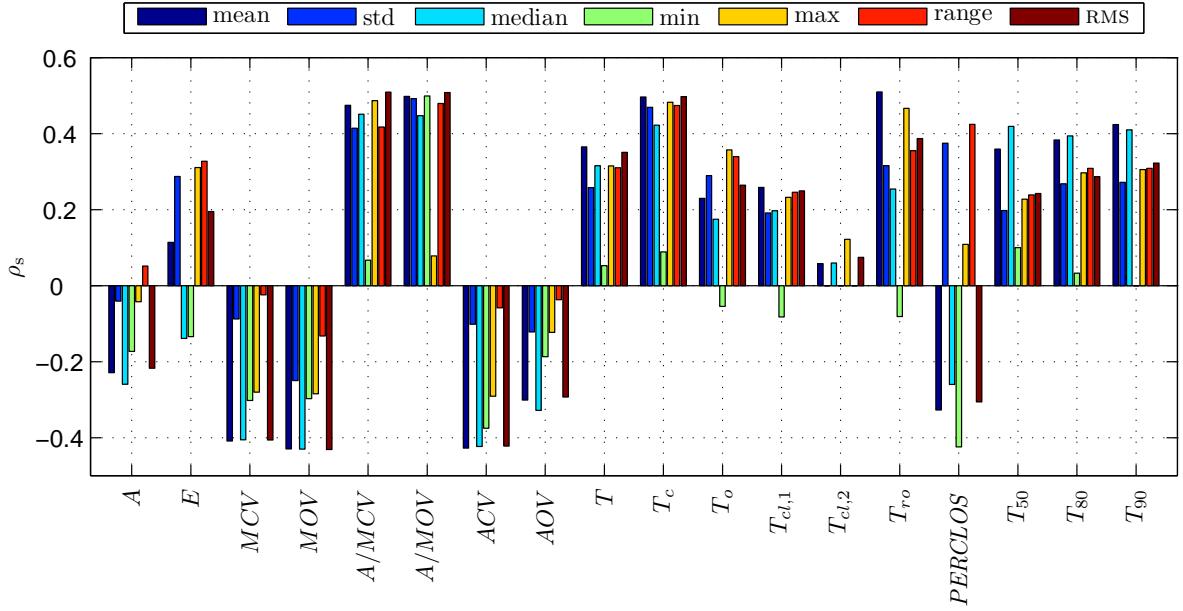


Figure F.1.: Comparison of Spearman's rank correlation coefficient between statistical metrics of features and KSS values. Feature type: drive time-based features

Interestingly, some of the metrics are associated with drowsiness in different directions. As an example, *PERCLOS* is positively correlated with the KSS regarding std and RMS, while for the mean and min, ρ_s values are negative.

Overall, mean, median and RMS are more consistent in terms of trends of all features in comparison to other metrics.

F.2. Boxplot of drive time-based features versus KSS values

Following figures show boxplots of drive time-based features versus KSS values for 42 subjects separately. For subject S2, which is not included in the following figures, all features are shown together in Figure F.40.

- A : Figures F.2 and F.3
- E : Figures F.4 and F.5
- MCV : Figures F.6 and F.7
- MOV : Figures F.8 and F.9
- A/MCV : Figures F.10 and F.11
- A/MOV : Figures F.12 and F.13
- ACV : Figures F.14 and F.15
- AOV : Figures F.16 and F.17
- F : Figures F.18 and F.19
- T : Figures F.20 and F.21
- T_c : Figures F.22 and F.23
- T_o : Figures F.24 and F.25
- $T_{d,1}$: Figures F.26 and F.27
- $T_{d,2}$: Figures F.28 and F.29
- T_{ro} : Figures F.30 and F.31
- $PERCLOS$: Figures F.32 and F.33
- T_{50} : Figures F.34 and F.35
- T_{80} : Figures F.36 and F.37
- T_{90} : Figures F.38 and F.39

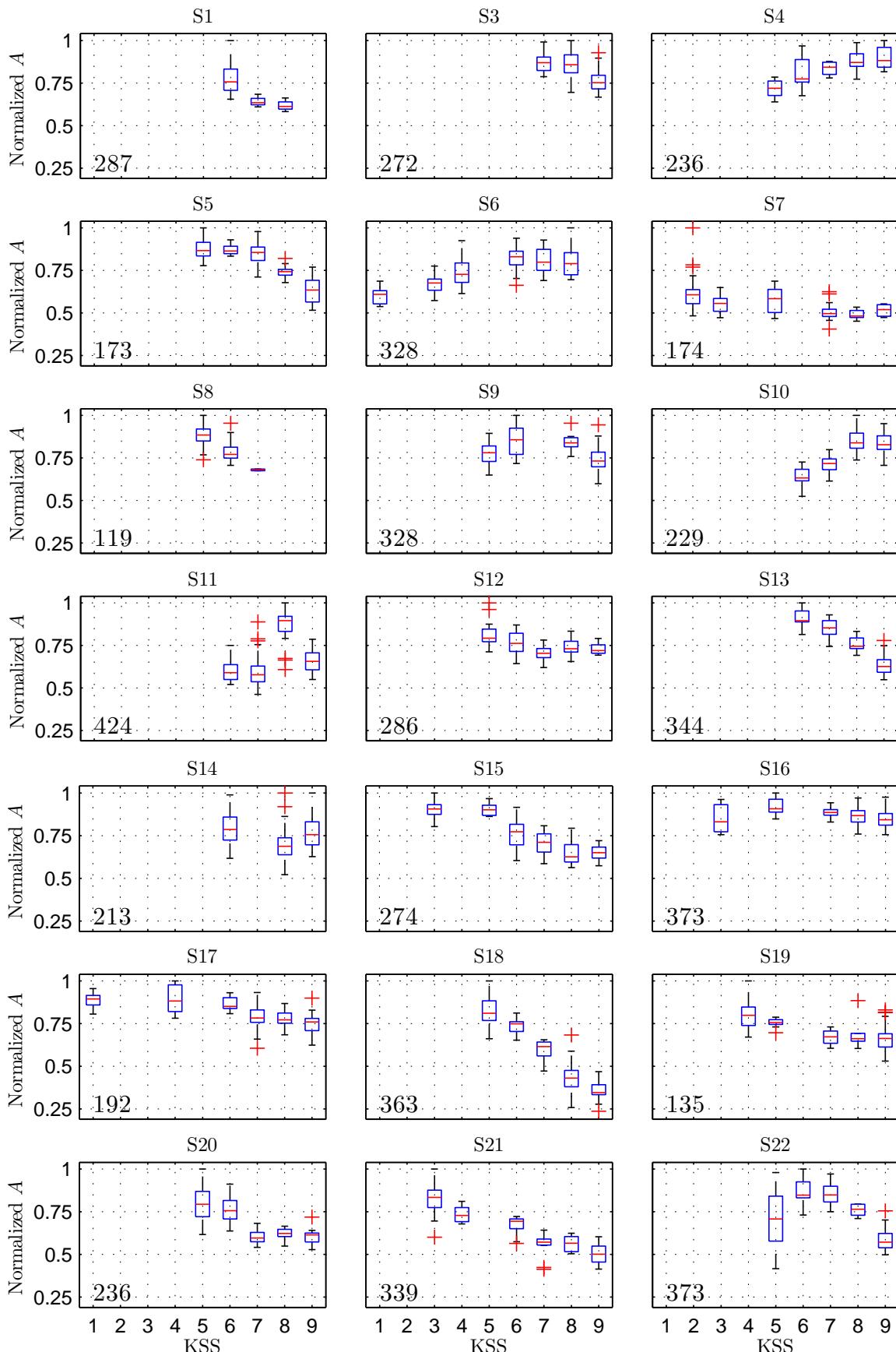


Figure F.2.: Boxplot of normalized feature A versus KSS for subjects S1 to S22 (except for subject S2).
The values on the bottom left show the maximum of A [μV] for each subject.

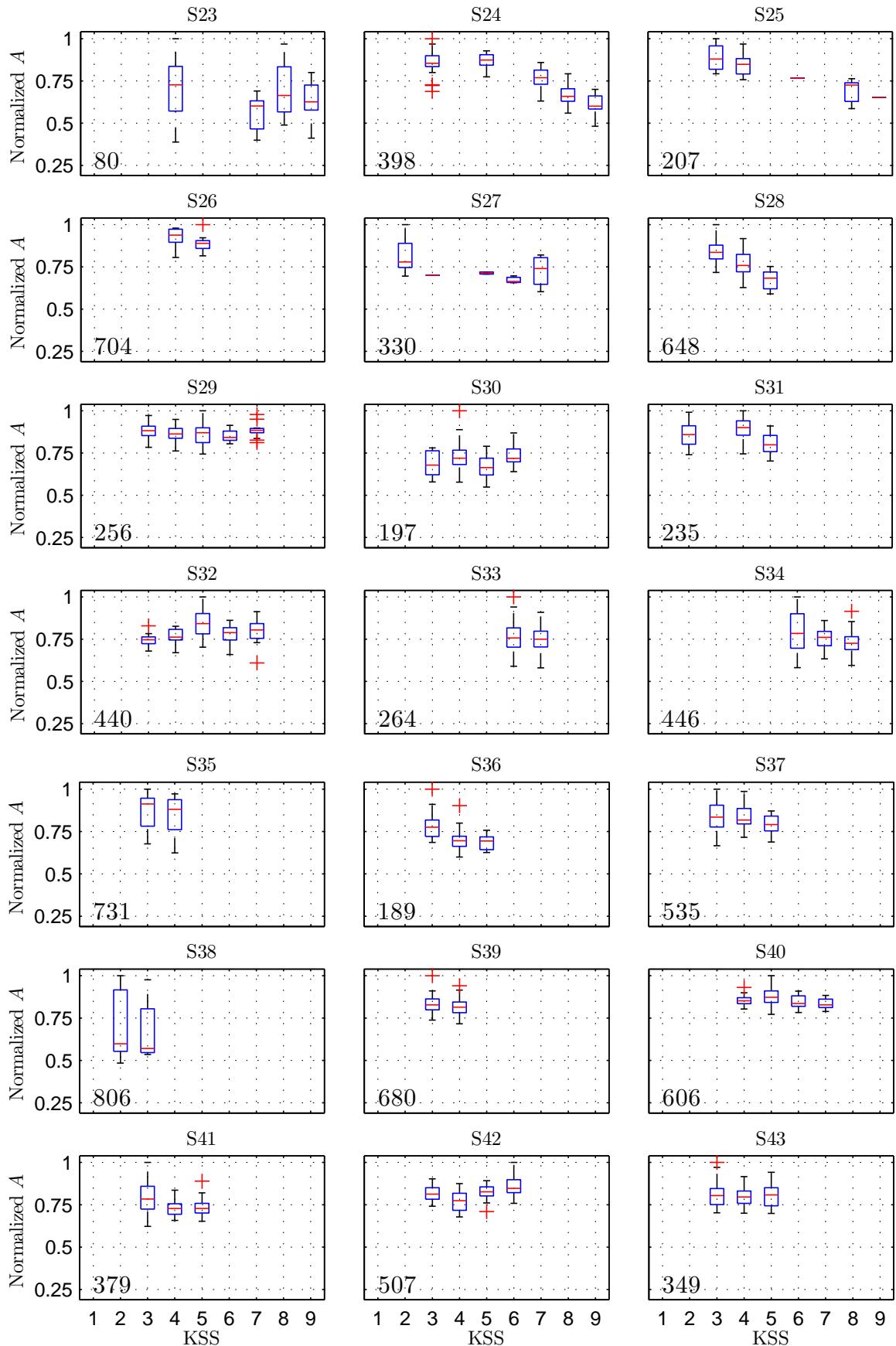


Figure F.3.: Boxplot of normalized feature A versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of A [μV] for each subject.

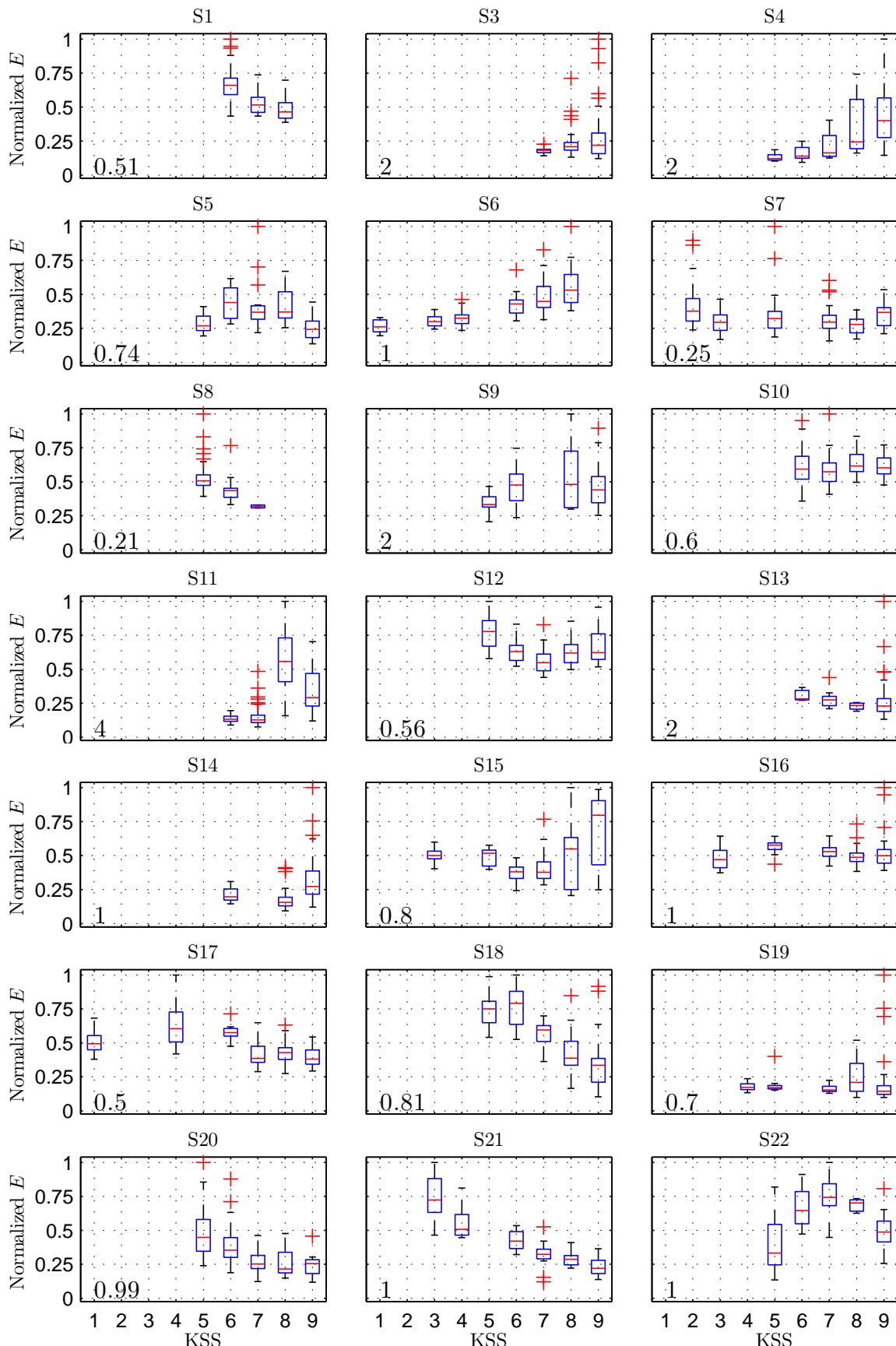


Figure F.4.: Boxplot of normalized feature E versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of E [$(\text{mV})^2$] for each subject.

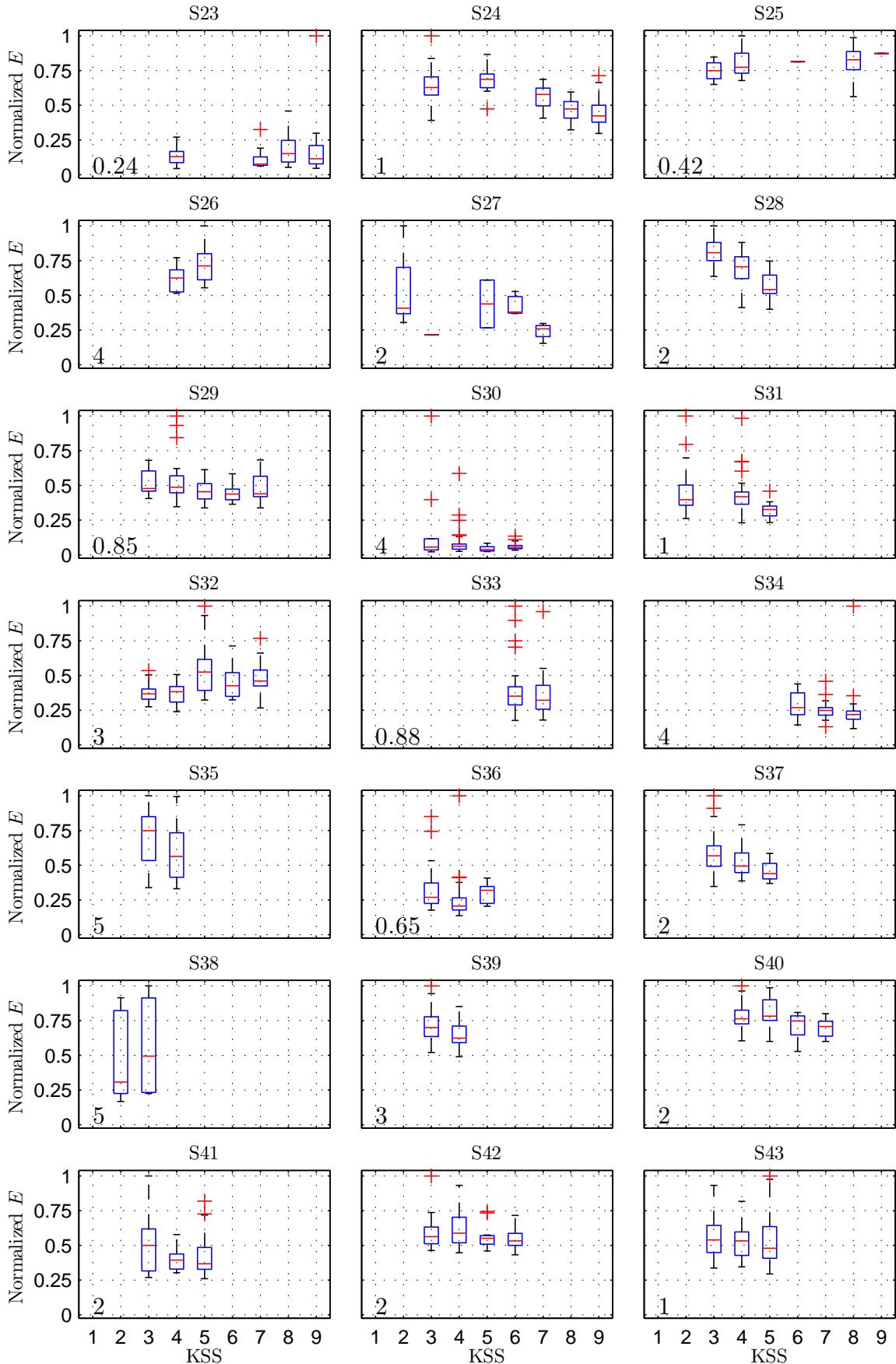


Figure F.5.: Boxplot of normalized feature E versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of E [$(\text{mV})^2$] for each subject.

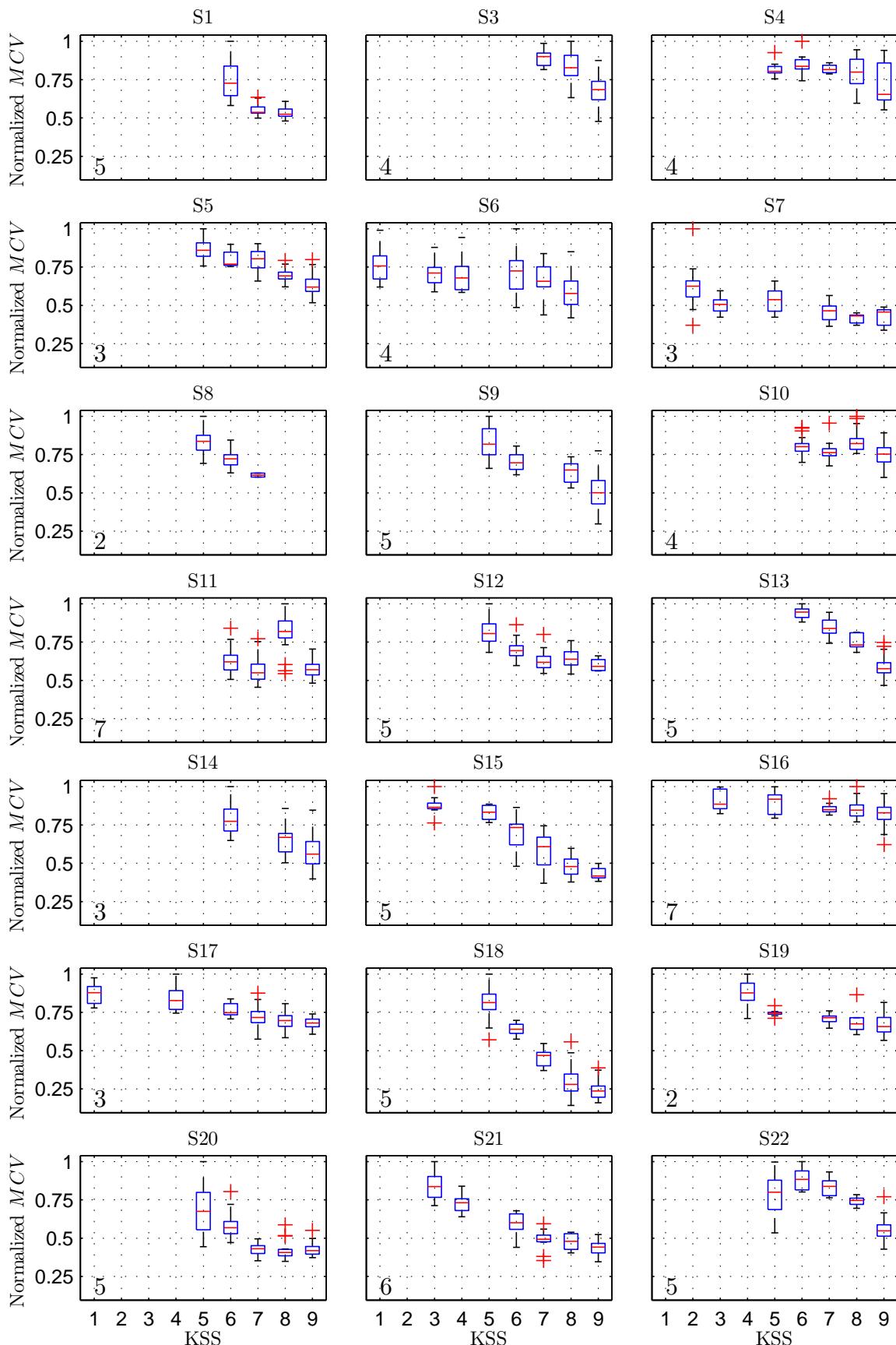


Figure F.6.: Boxplot of normalized feature MCV versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of MCV [mV/s] for each subject.

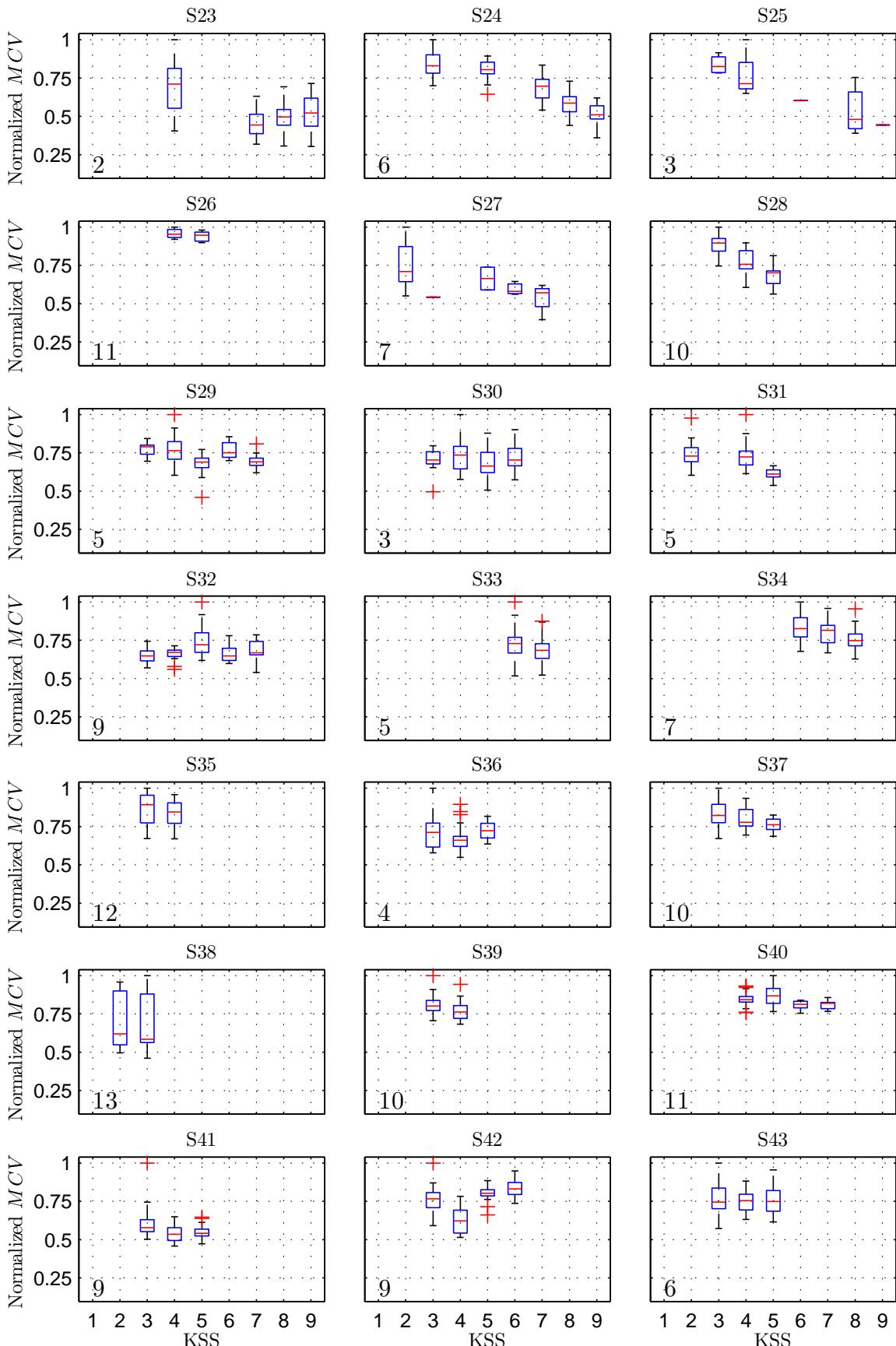


Figure F.7.: Boxplot of normalized feature MCV [mV/s] versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of MCV for each subject.

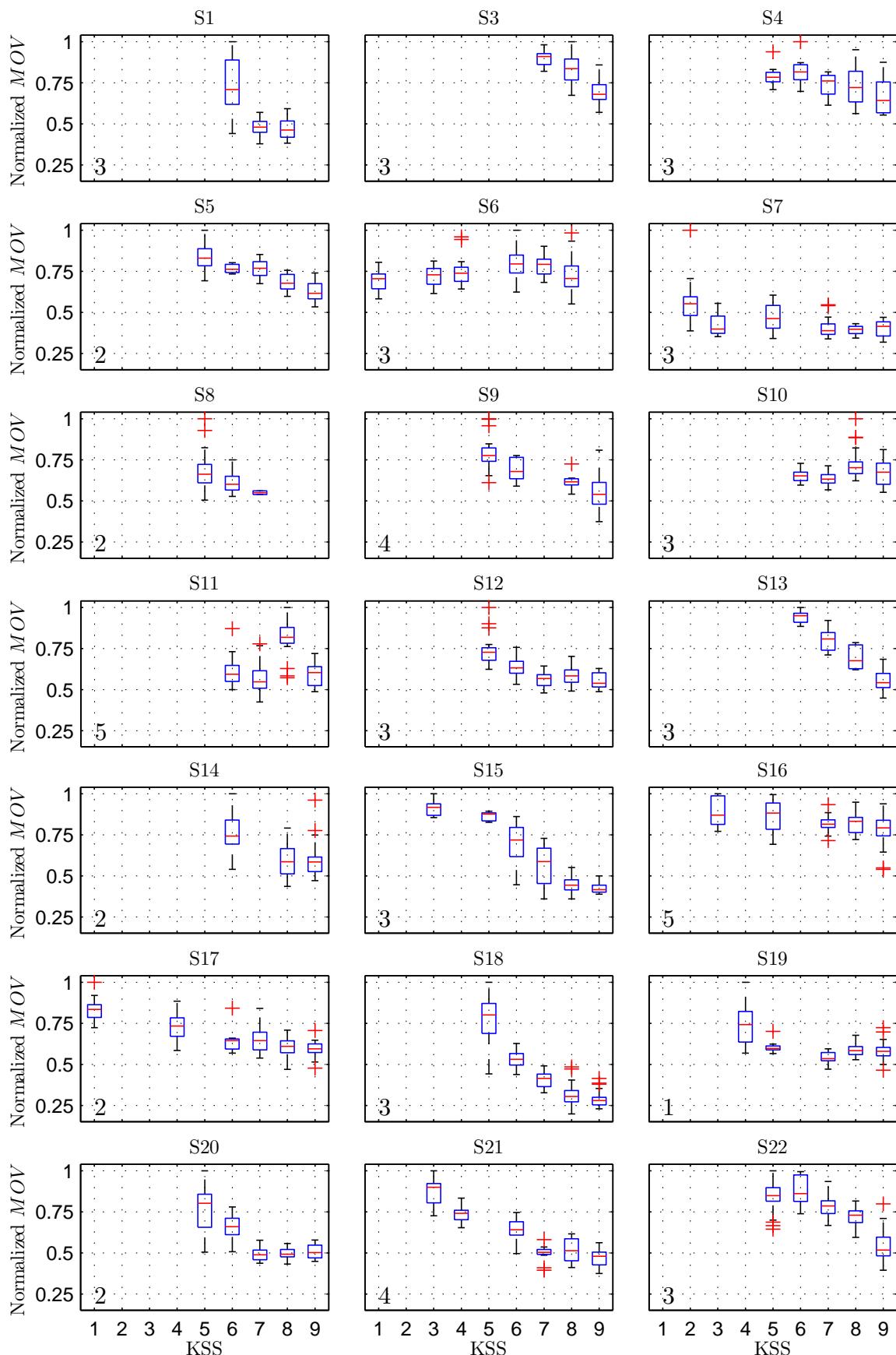


Figure F.8.: Boxplot of normalized feature MOV versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of MOV [mV/s] for each subject.

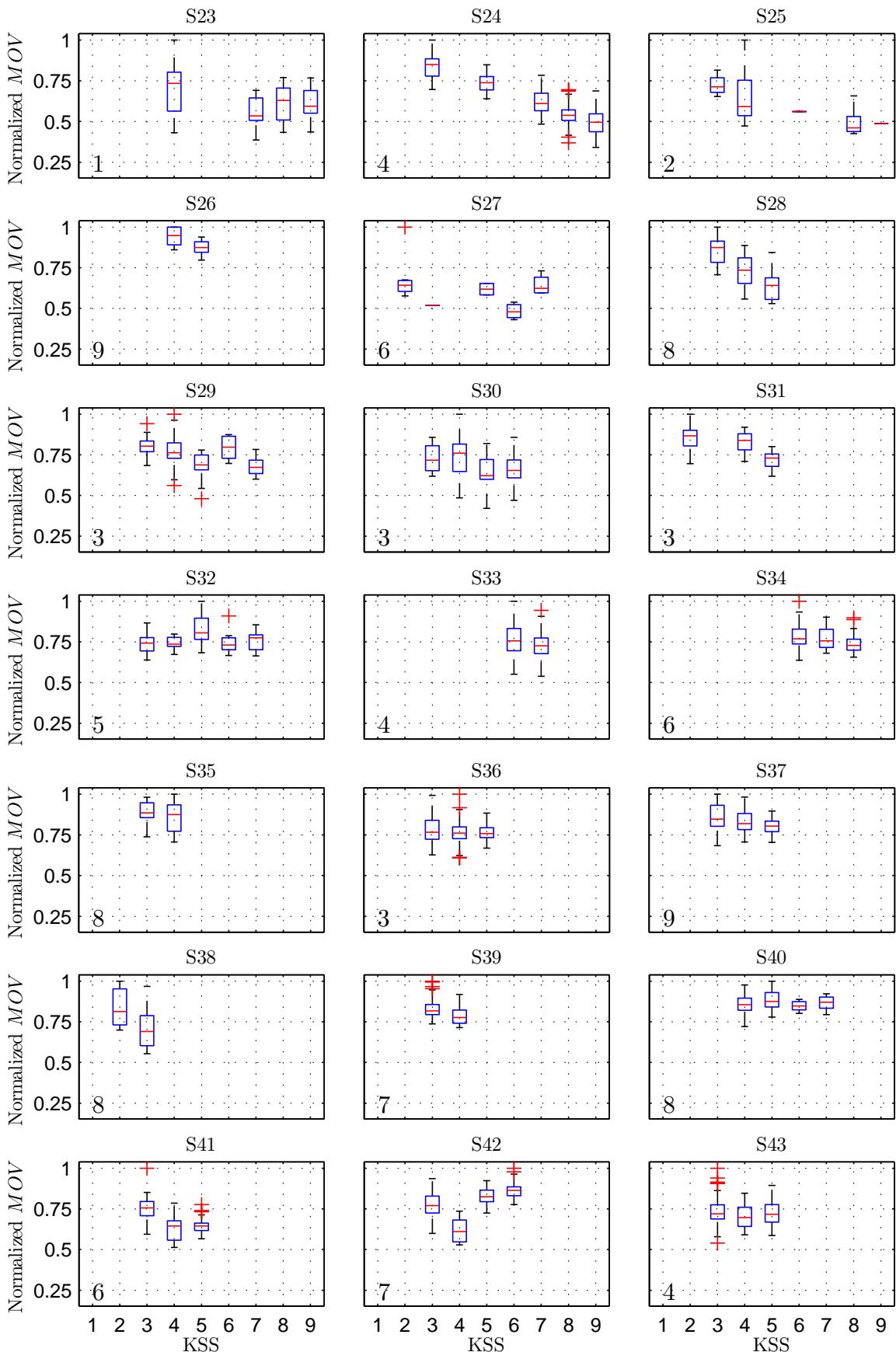


Figure F.9.: Boxplot of normalized feature *MOV* versus *KSS* for subjects S23 to S43. The values on the bottom left show the maximum of *MOV* [mV/s] for each subject.

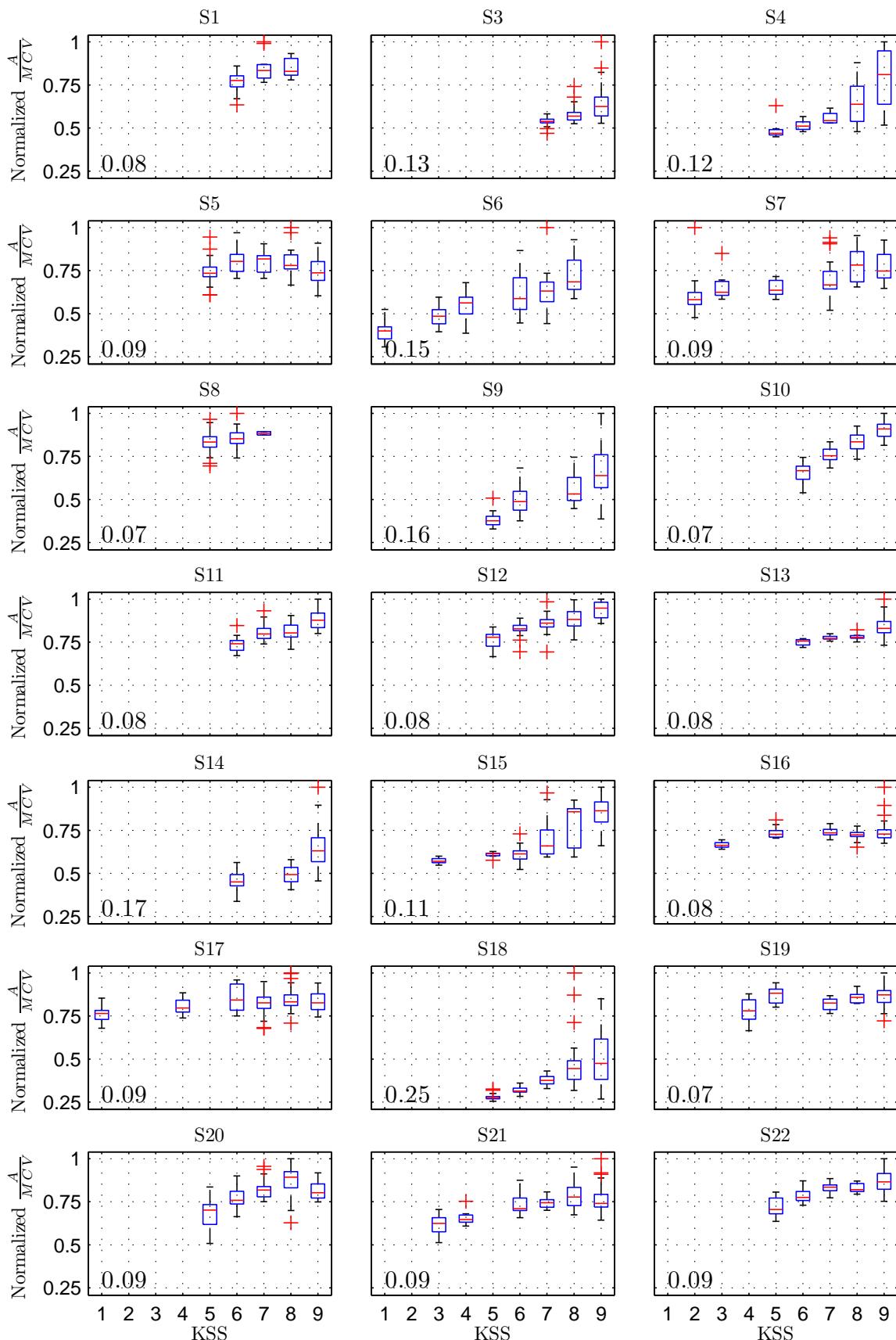


Figure F.10.: Boxplot of normalized feature A/MCV versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of A/MCV [s] for each subject.

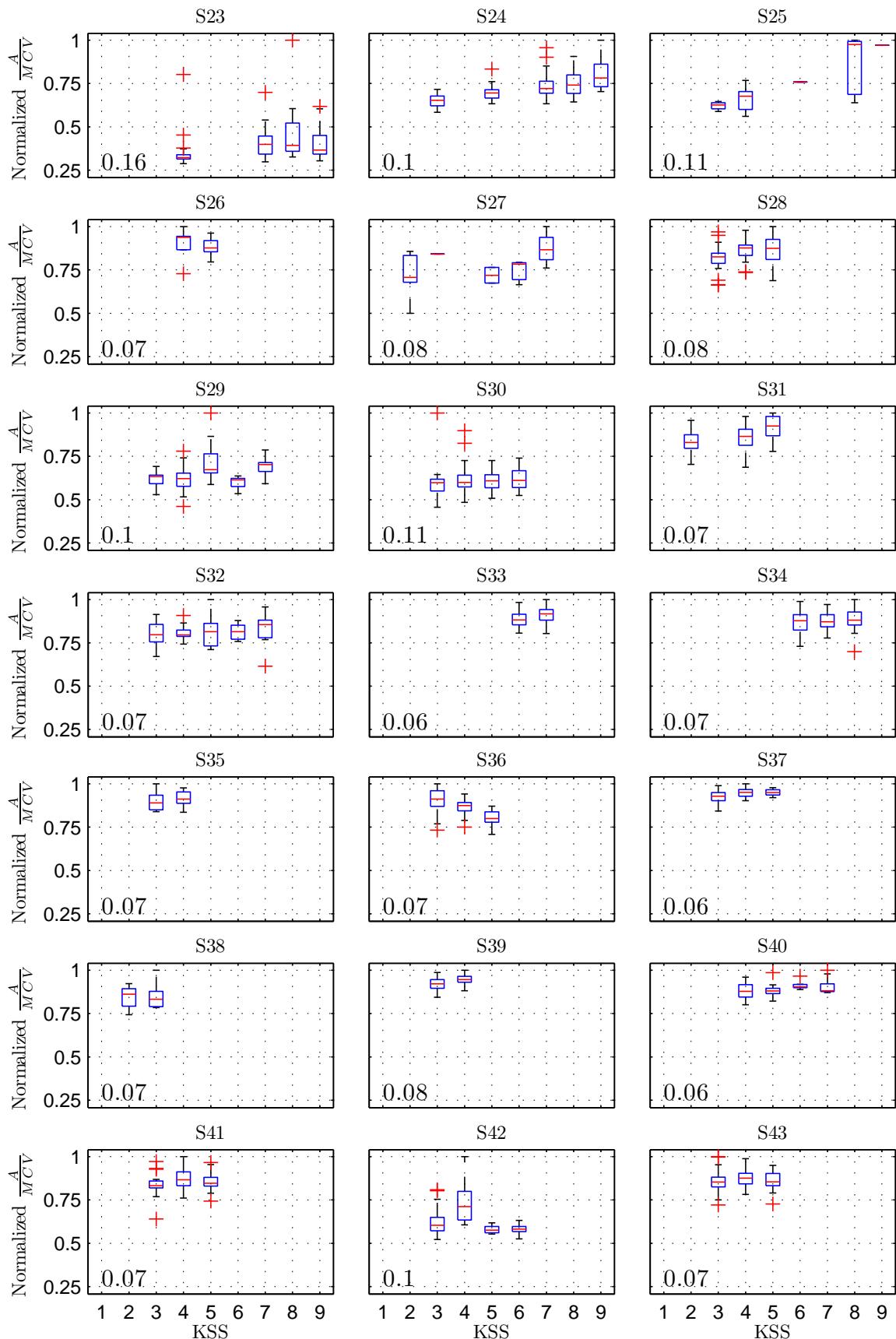


Figure F.11.: Boxplot of normalized feature A/MCV versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of A/MCV [s] for each subject.

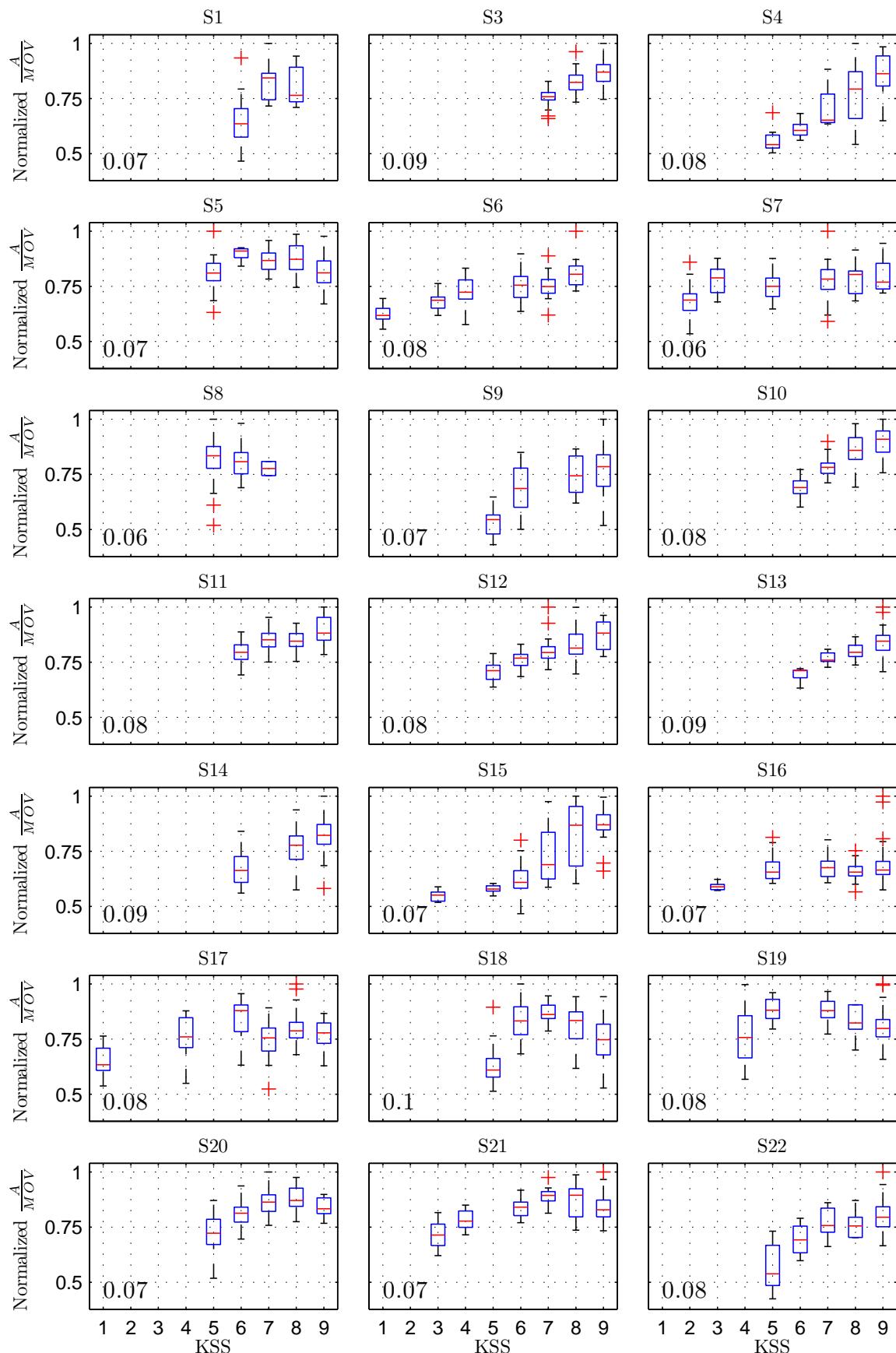


Figure F.12.: Boxplot of normalized feature A/MOV versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of A/MOV [s] for each subject.

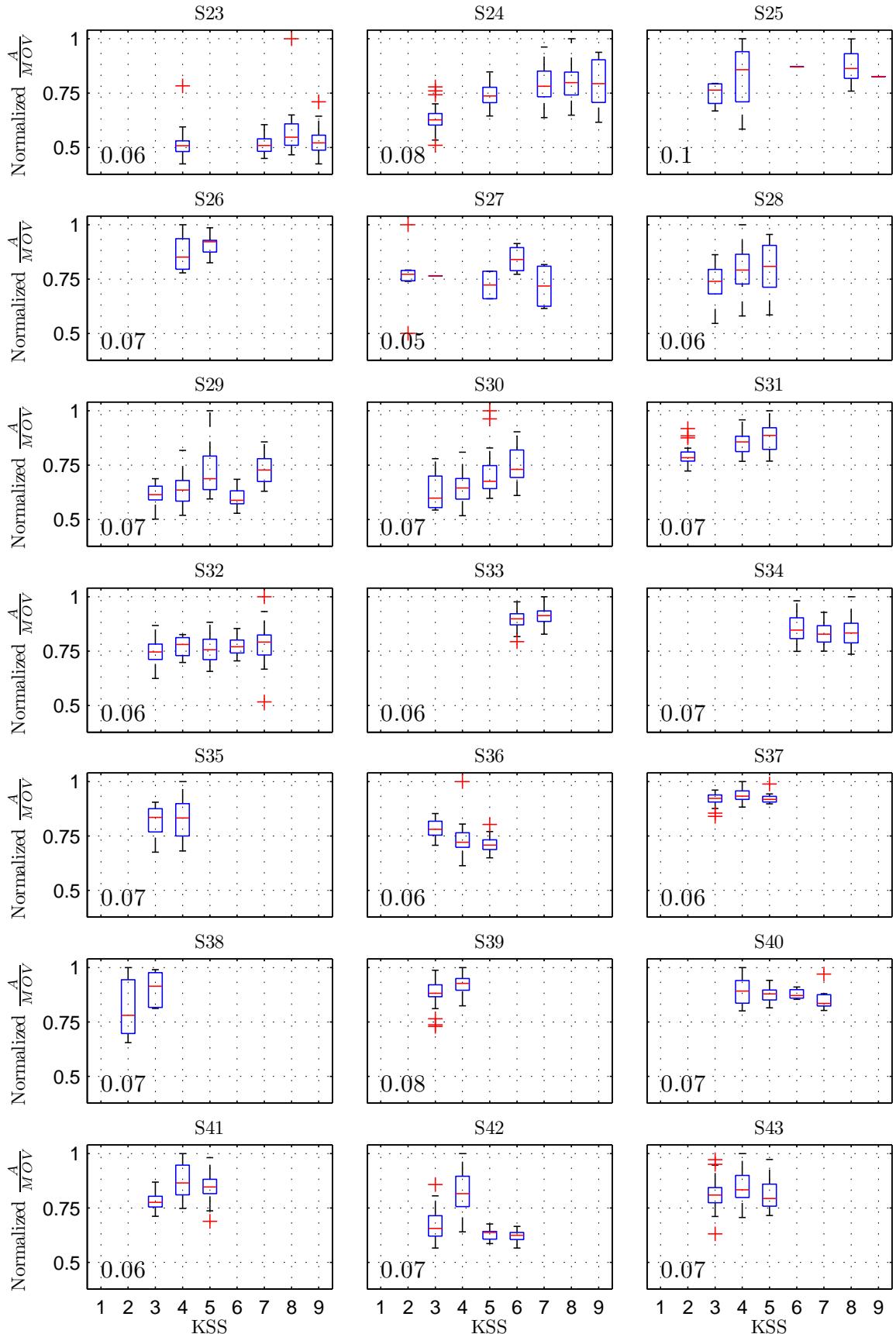


Figure F.13.: Boxplot of normalized feature A/MOV versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of A/MOV [s] for each subject.

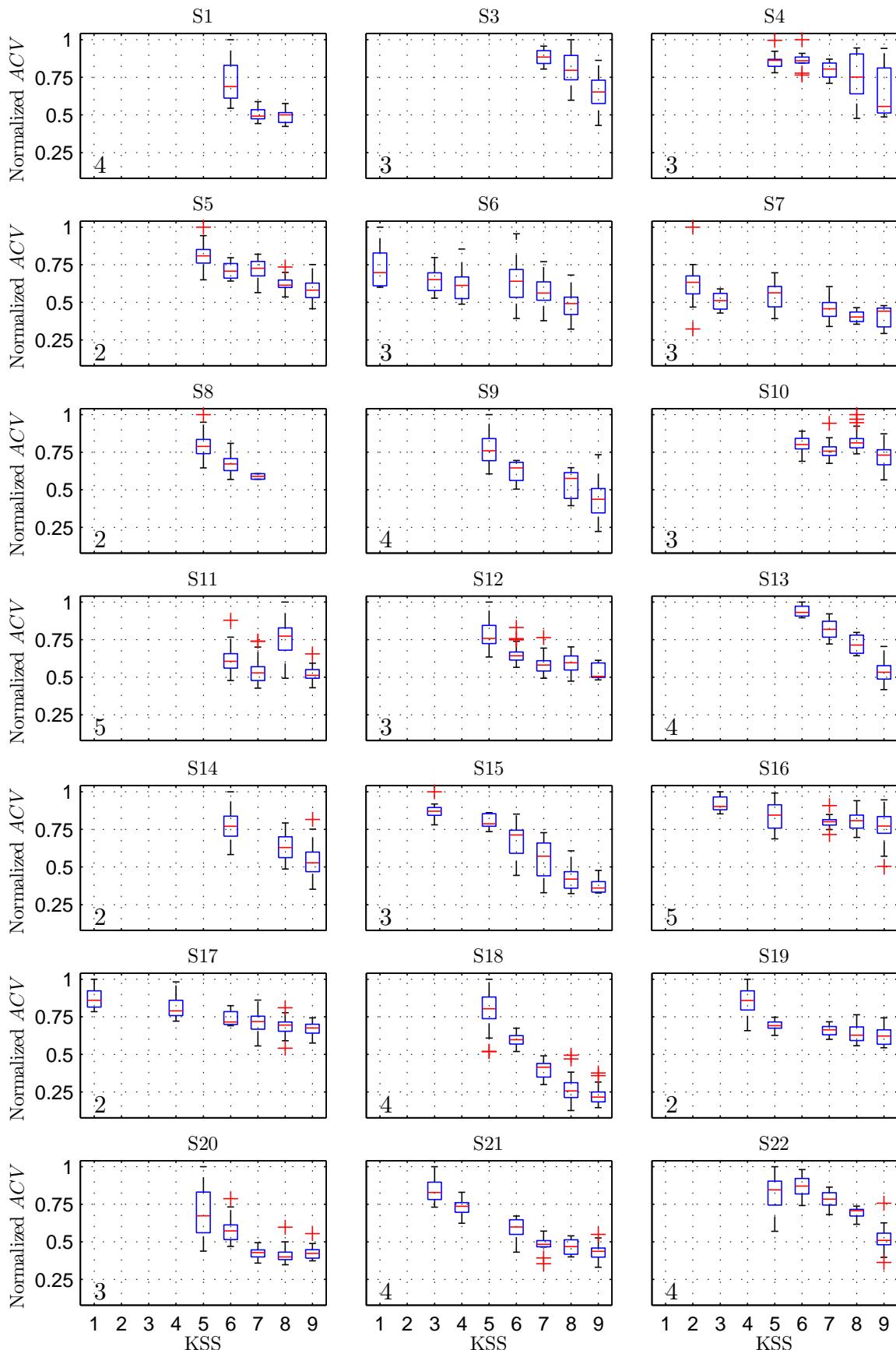


Figure F.14.: Boxplot of normalized feature ACV versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of ACV [mV/s] for each subject.

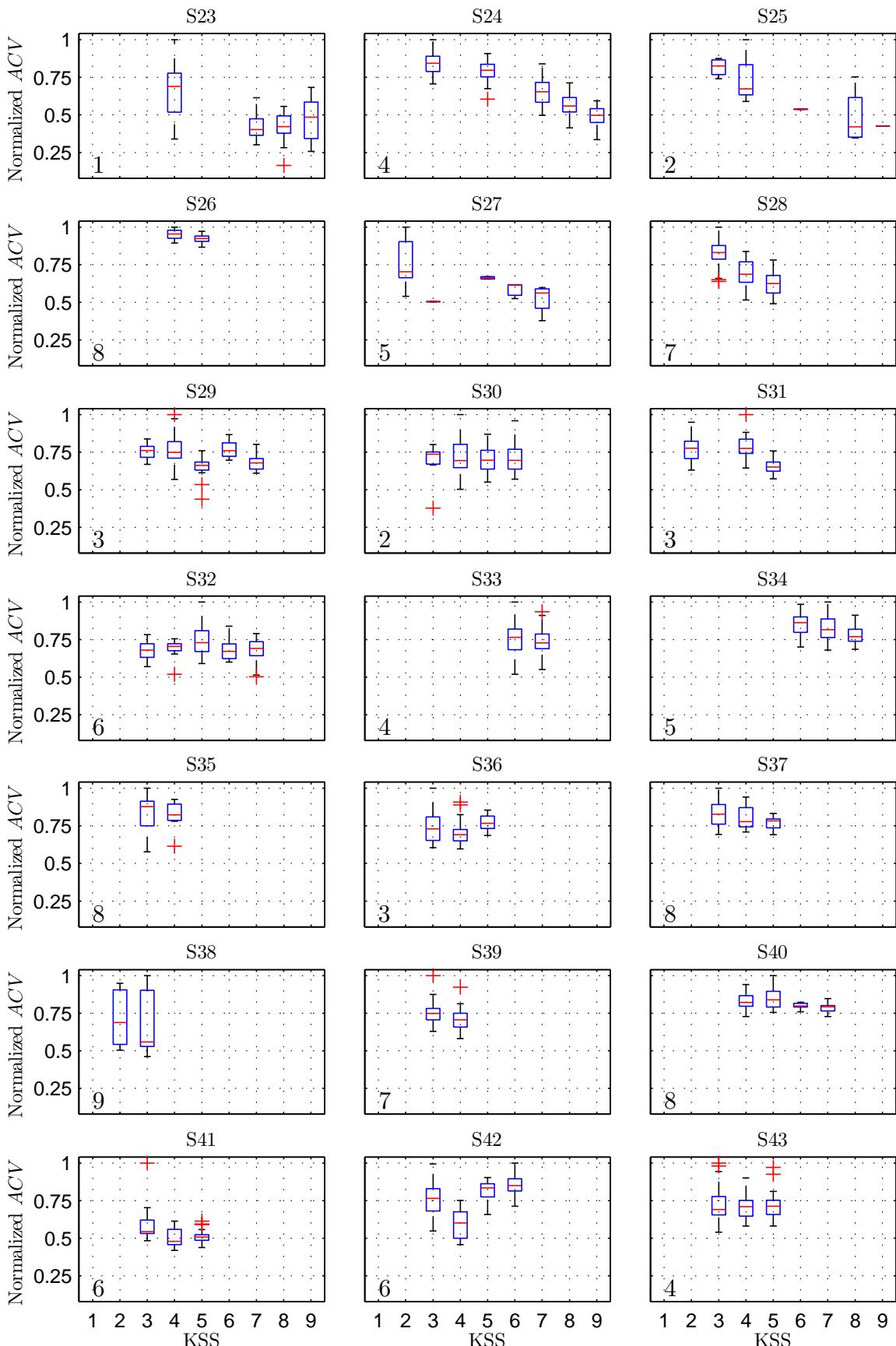


Figure F.15.: Boxplot of normalized feature ACV versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of ACV [mV/s] for each subject.

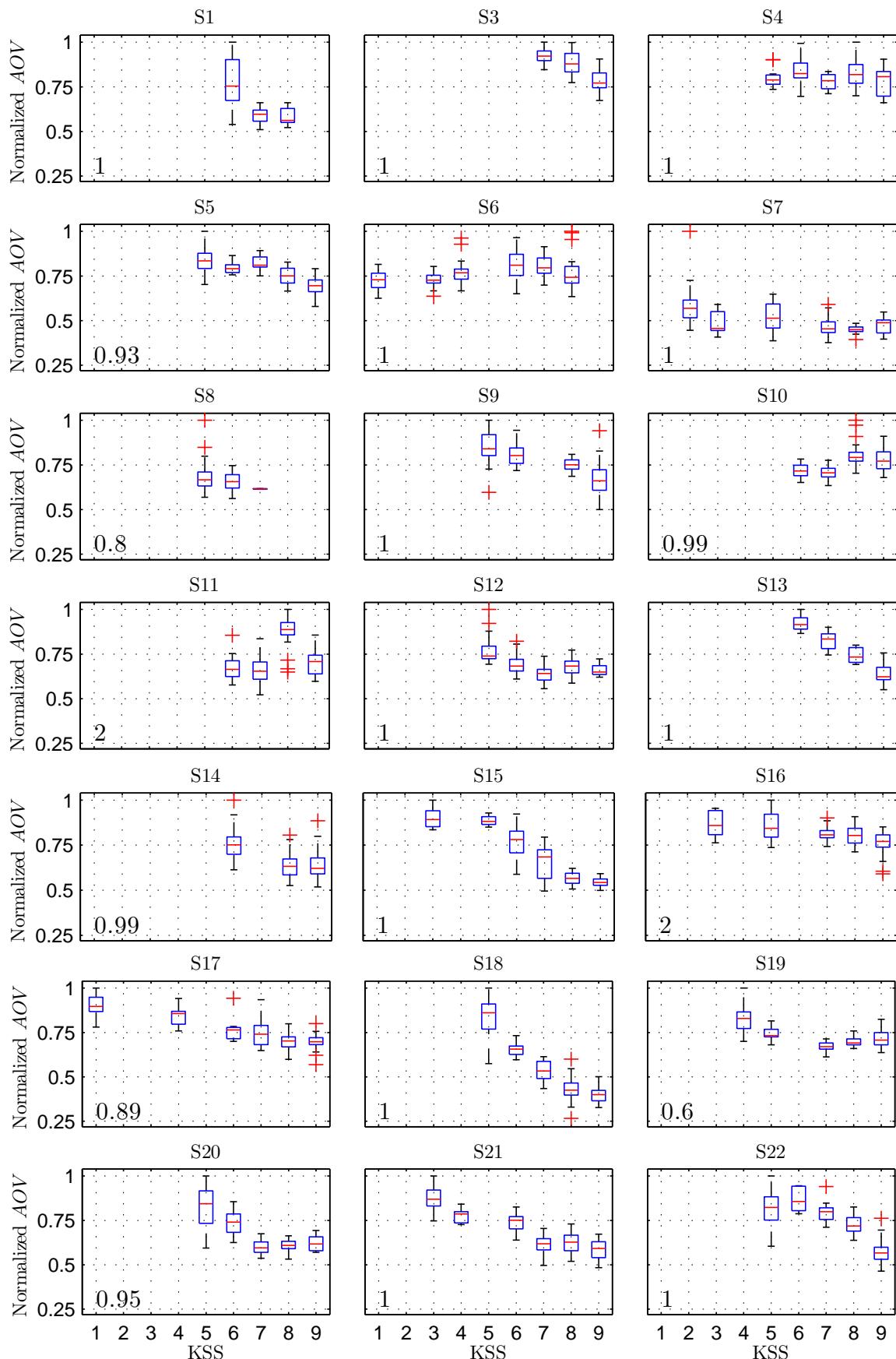


Figure F.16.: Boxplot of normalized feature AOV versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of AOV [mV/s] for each subject.

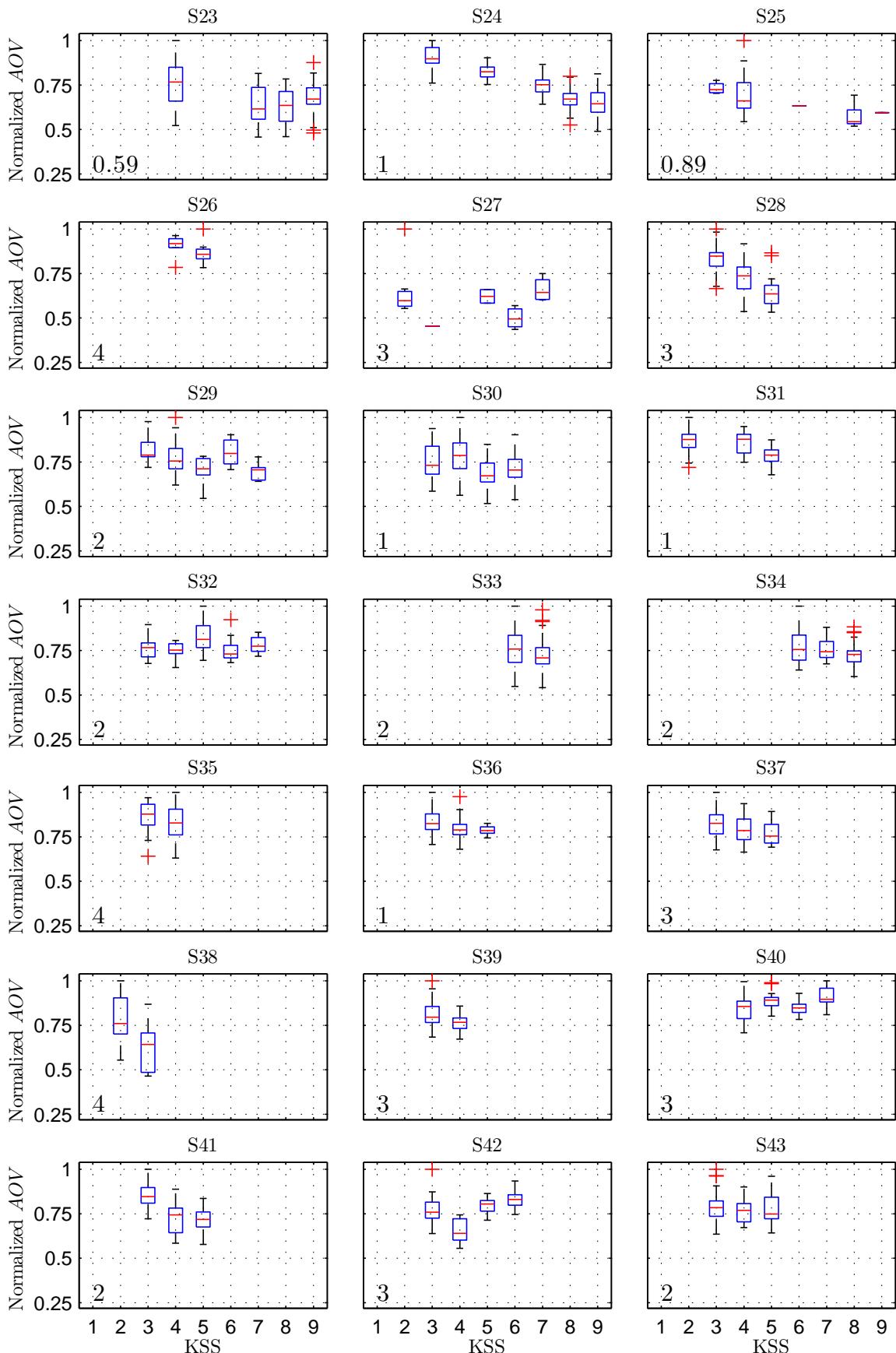


Figure F.17.: Boxplot of normalized feature AOV versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of AOV [mV/s] for each subject.

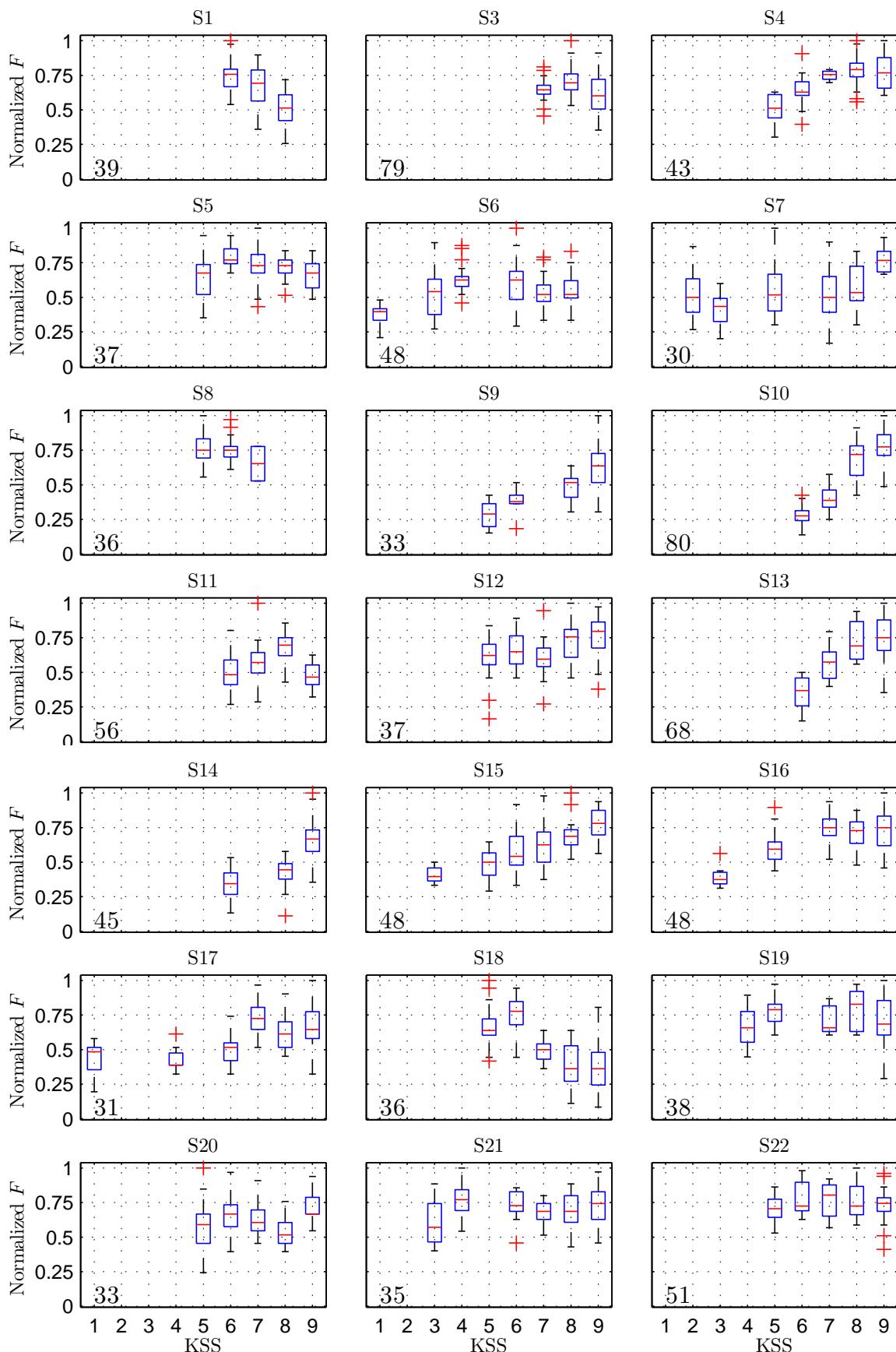


Figure F.18.: Boxplot of normalized feature F versus KSS for subjects S1 to S22 (except for subject S2).
The values on the bottom left show the maximum of F [1/min] for each subject.

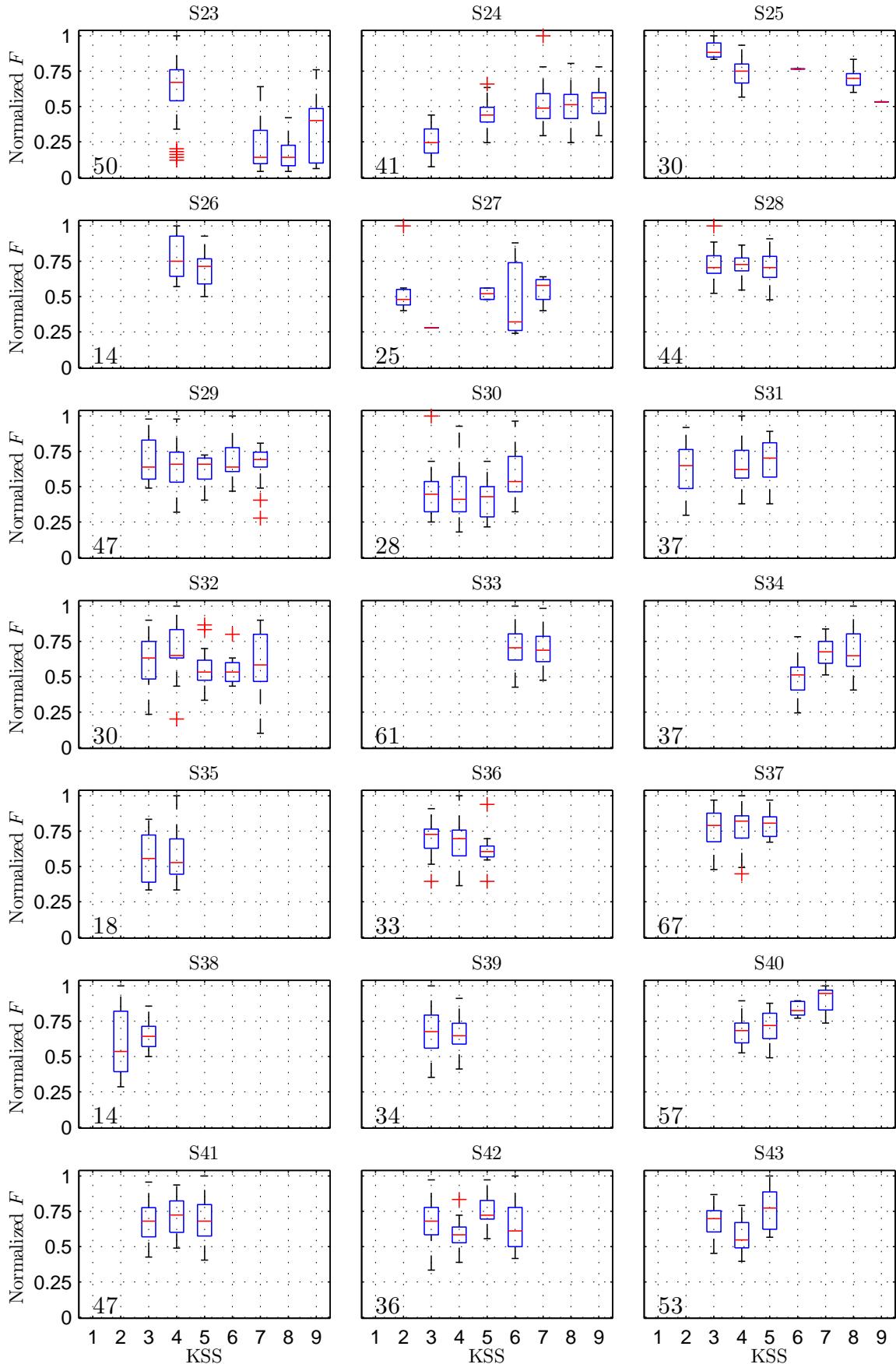


Figure F.19.: Boxplot of normalized feature F versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of F [1/min] for each subject.

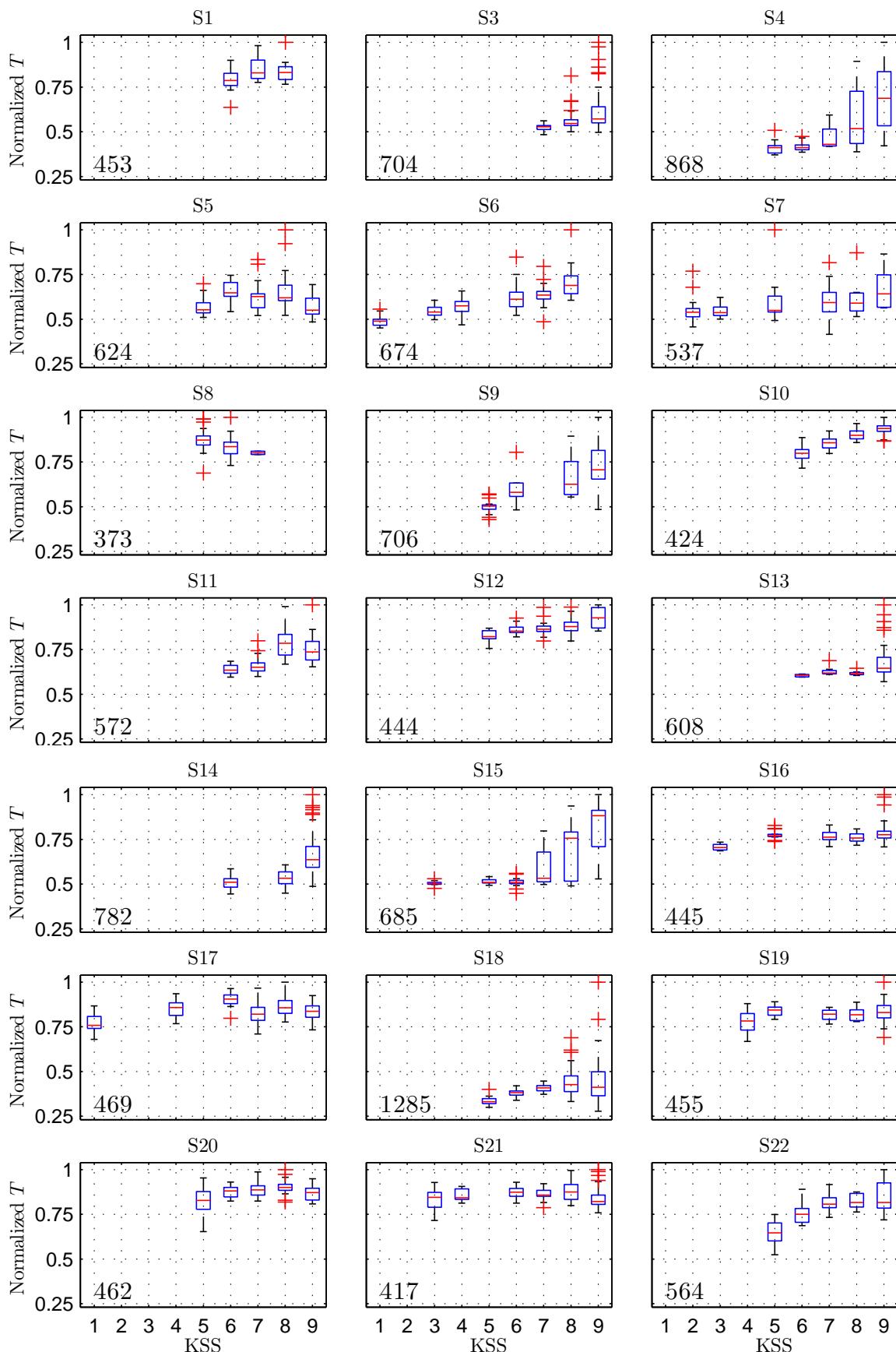


Figure F.20.: Boxplot of normalized feature T versus KSS for subjects S1 to S22 (except for subject S2).
The values on the bottom left show the maximum of T [s] for each subject.

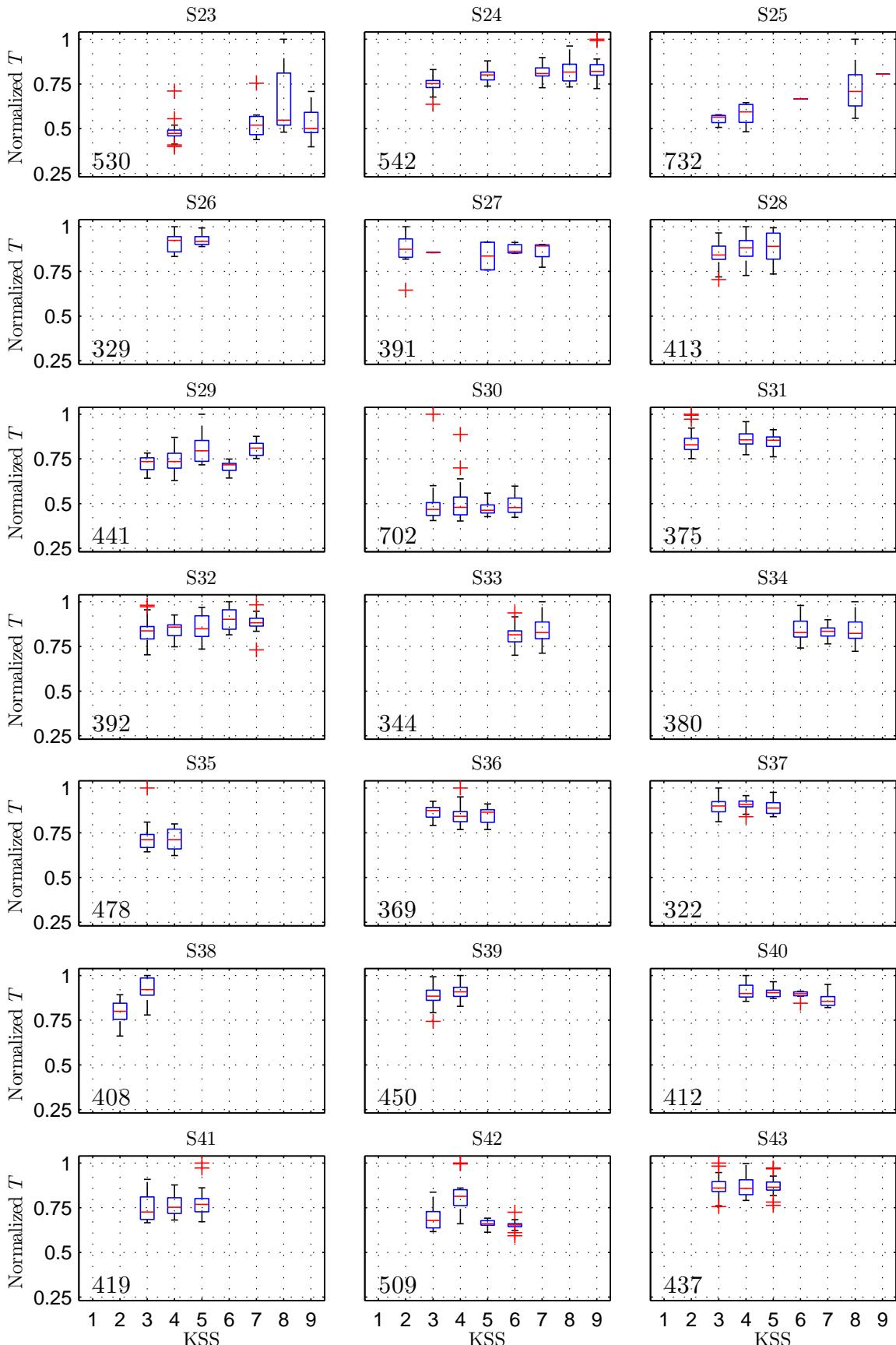


Figure F.21.: Boxplot of normalized feature T versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of T [s] for each subject.

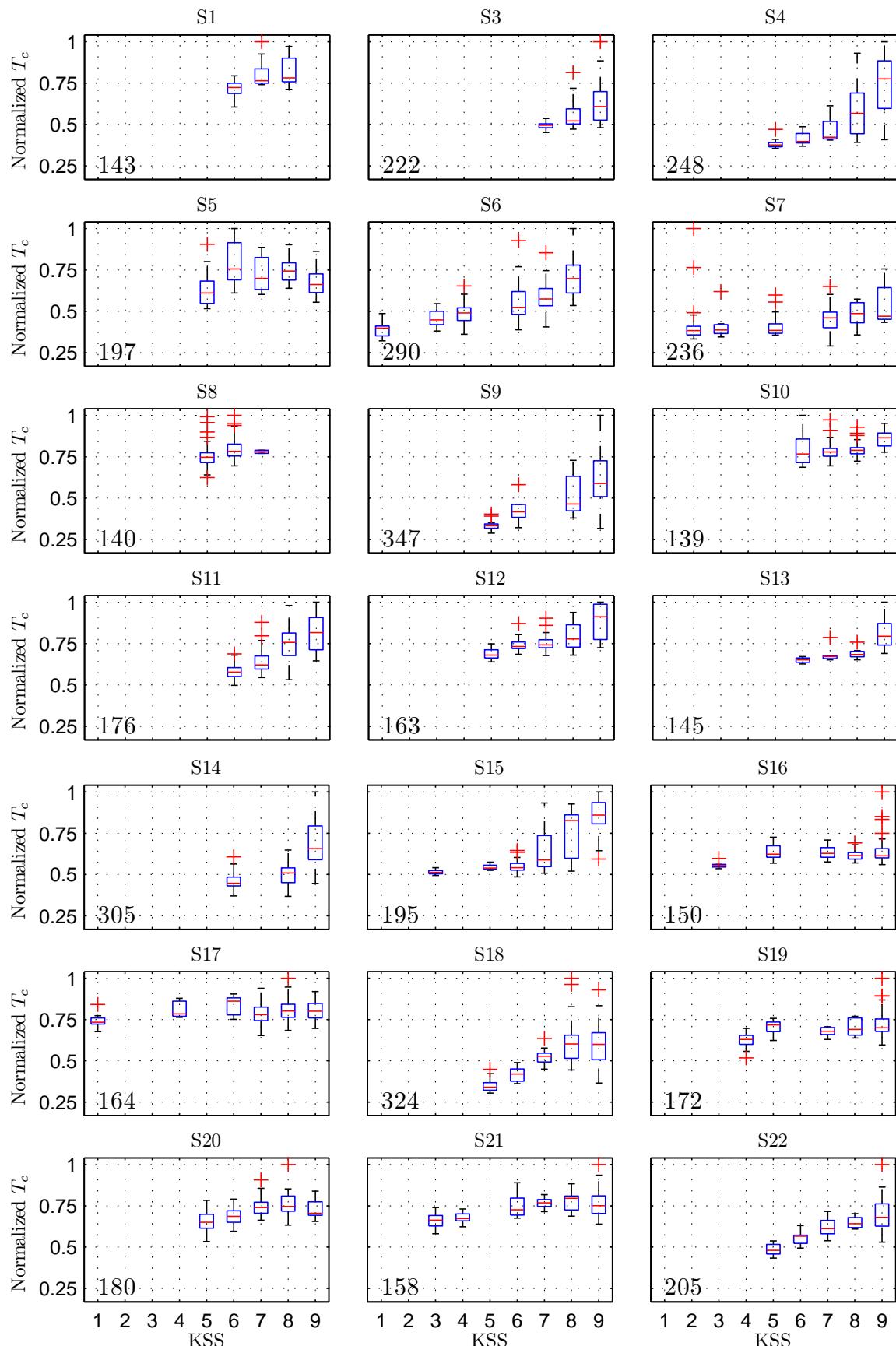


Figure F.22.: Boxplot of normalized feature T_c versus KSS for subjects S1 to S22 (except for subject S2).
The values on the bottom left show the maximum of T_c [ms] for each subject.

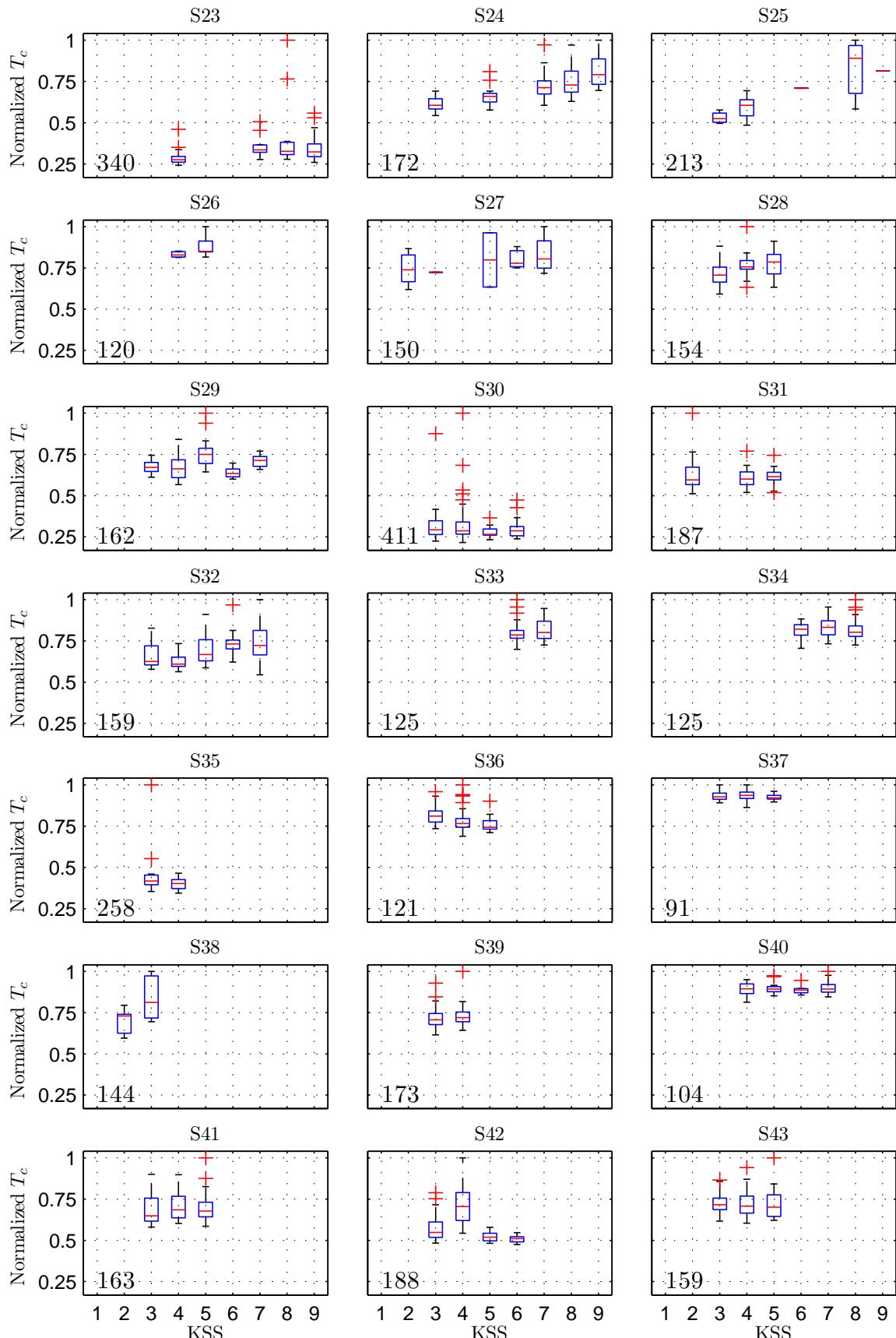


Figure F.23.: Boxplot of normalized feature T_c versus kss for subjects S23 to S43. The values on the bottom left show the maximum of T_c [ms] for each subject.

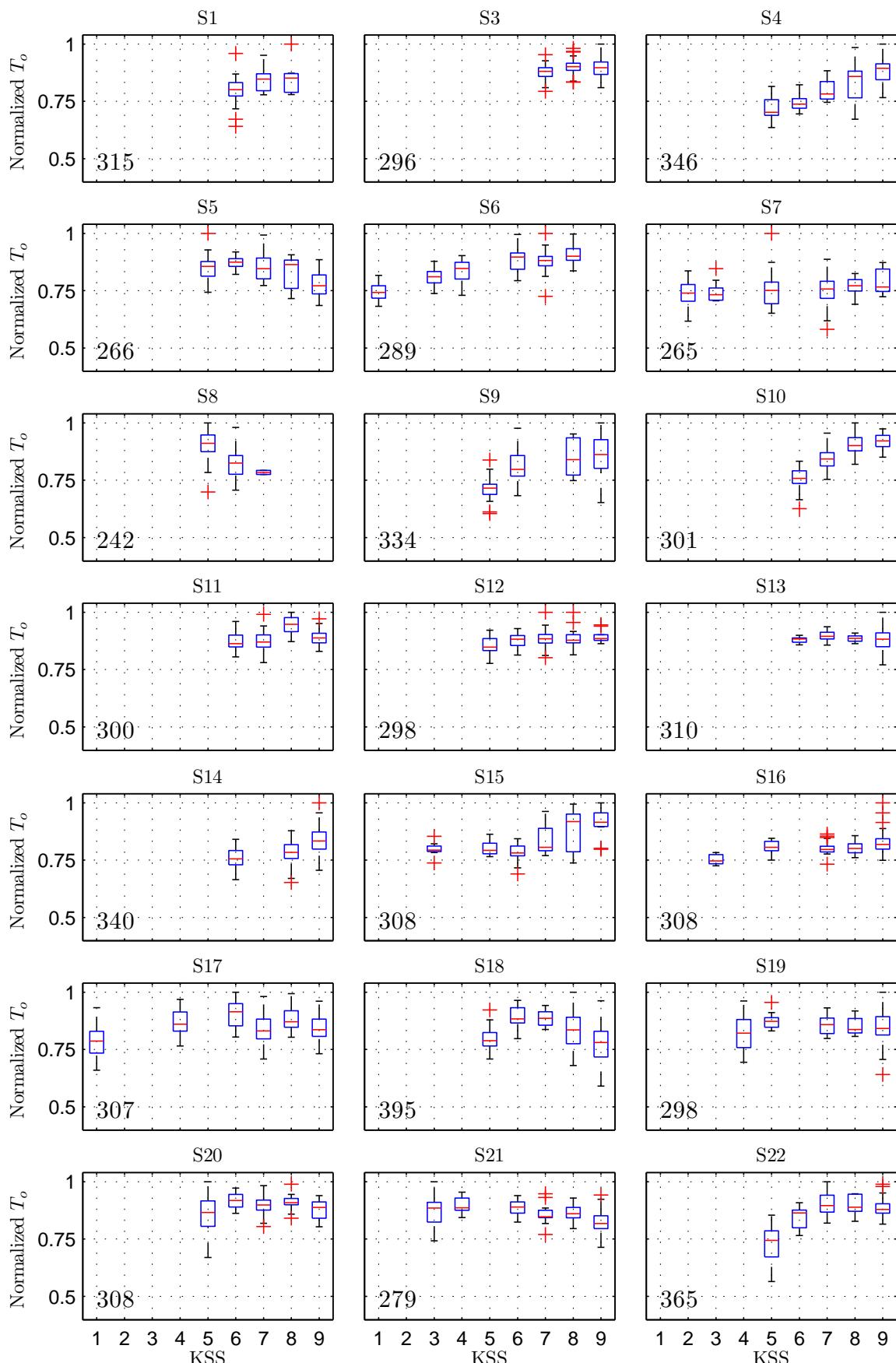


Figure F.24.: Boxplot of normalized feature T_o versus KSS for subjects S1 to S22 (except for subject S2).
The values on the bottom left show the maximum of T_o [ms] for each subject.

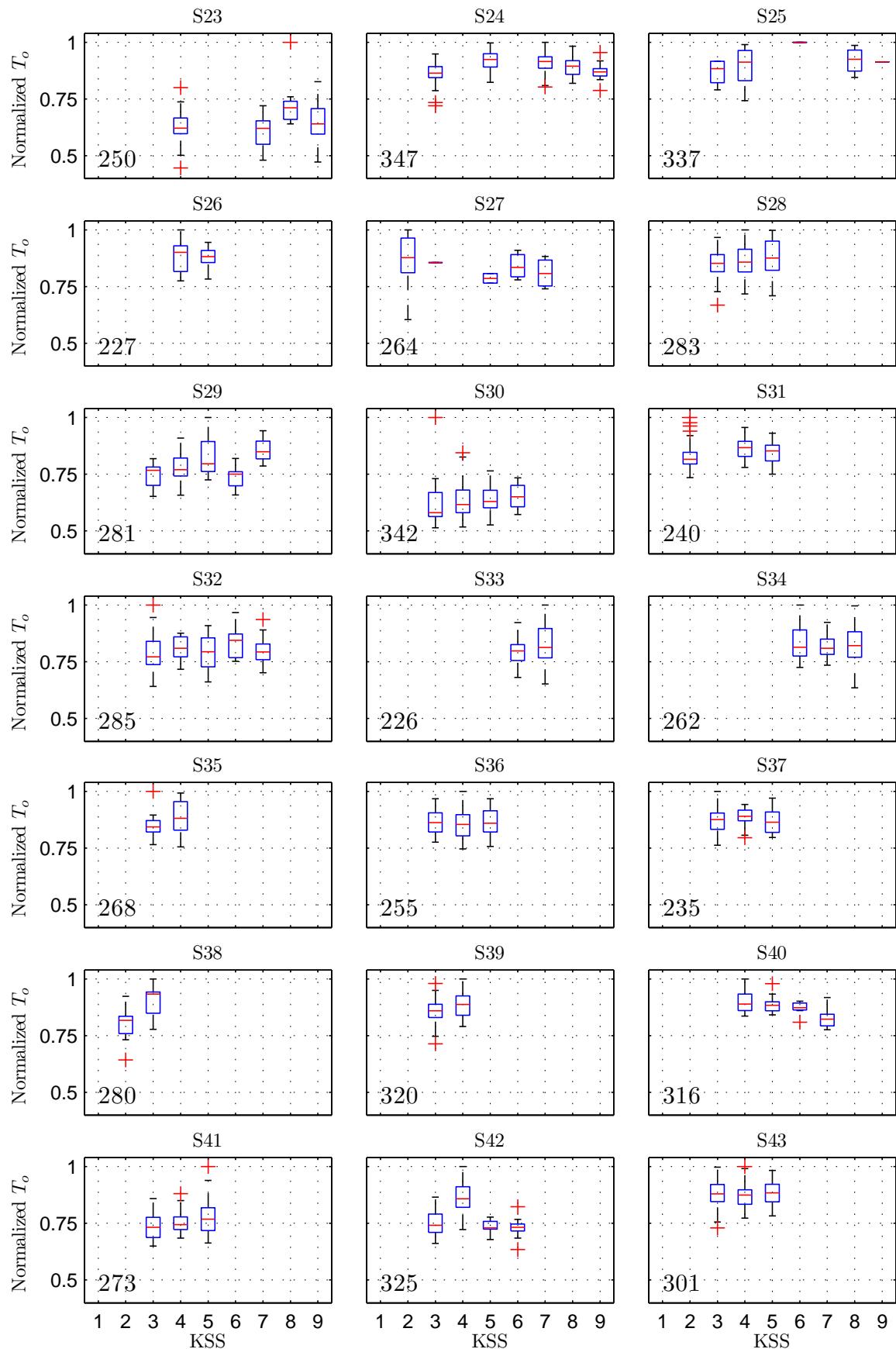


Figure F.25.: Boxplot of normalized feature T_o versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of T_o [ms] for each subject.

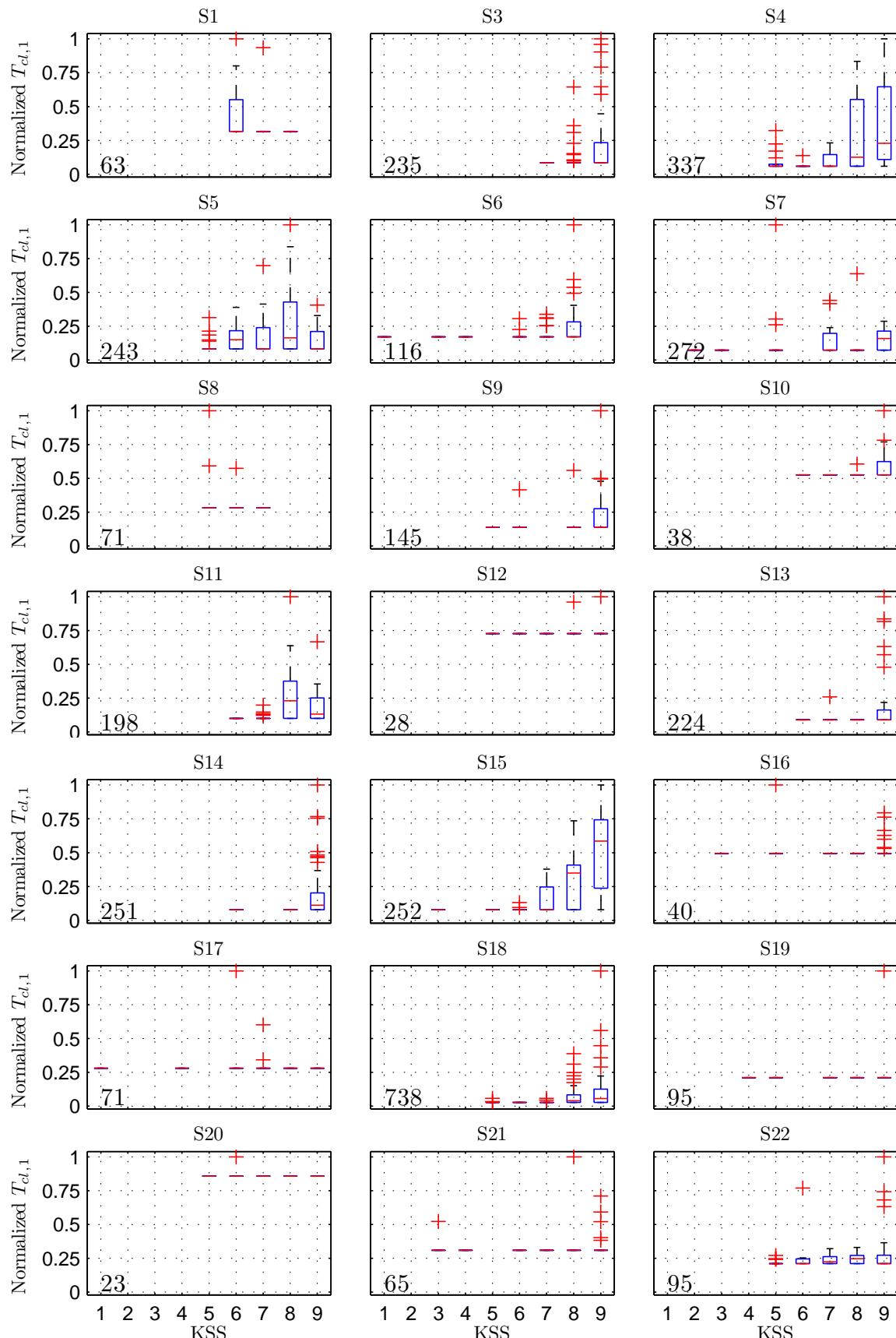


Figure F.26.: Boxplot of normalized feature $T_{cl,1}$ versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of $T_{cl,1}$ [ms] for each subject.

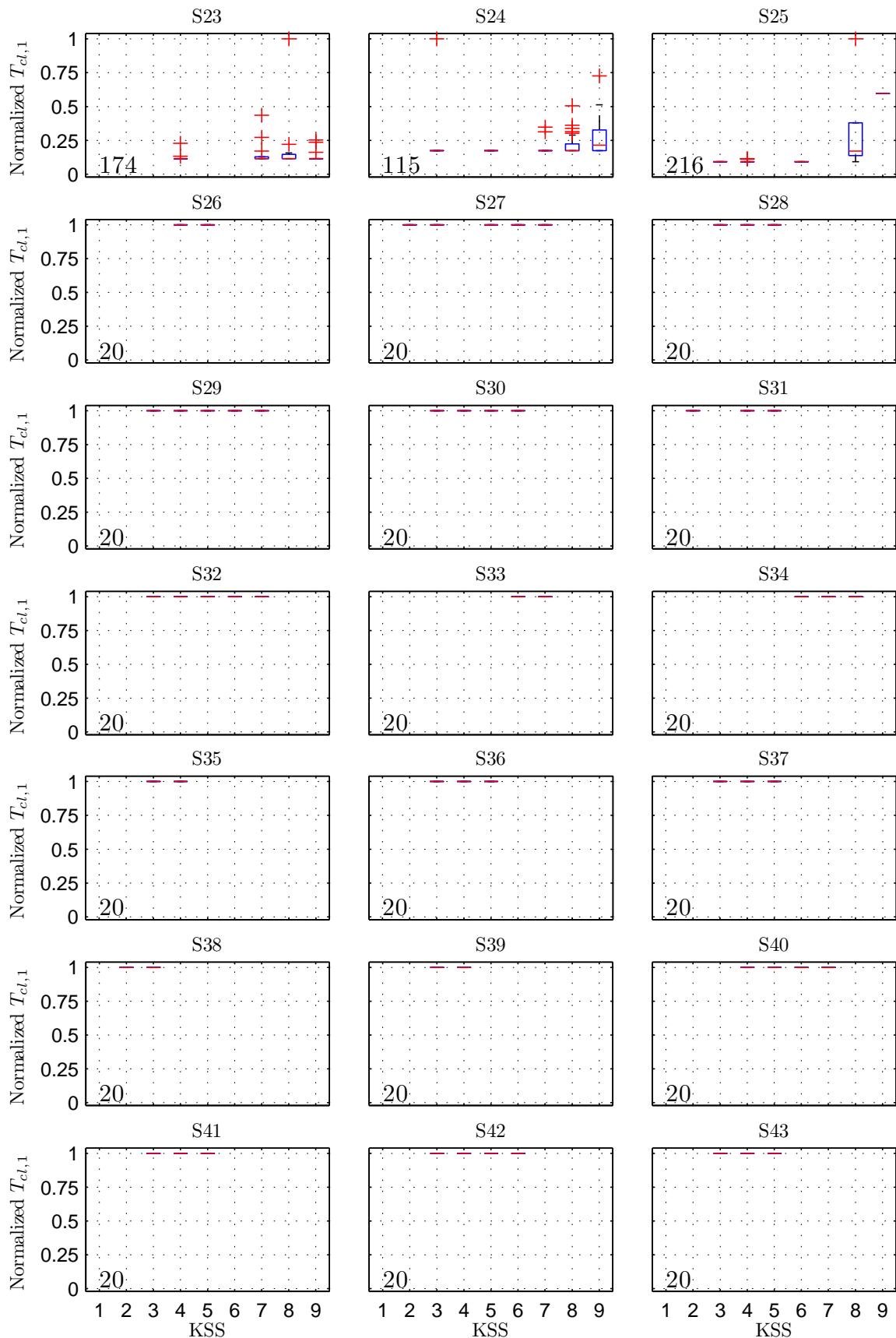


Figure F.27.: Boxplot of normalized feature $T_{cl,1}$ versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of $T_{cl,1}$ [ms] for each subject.

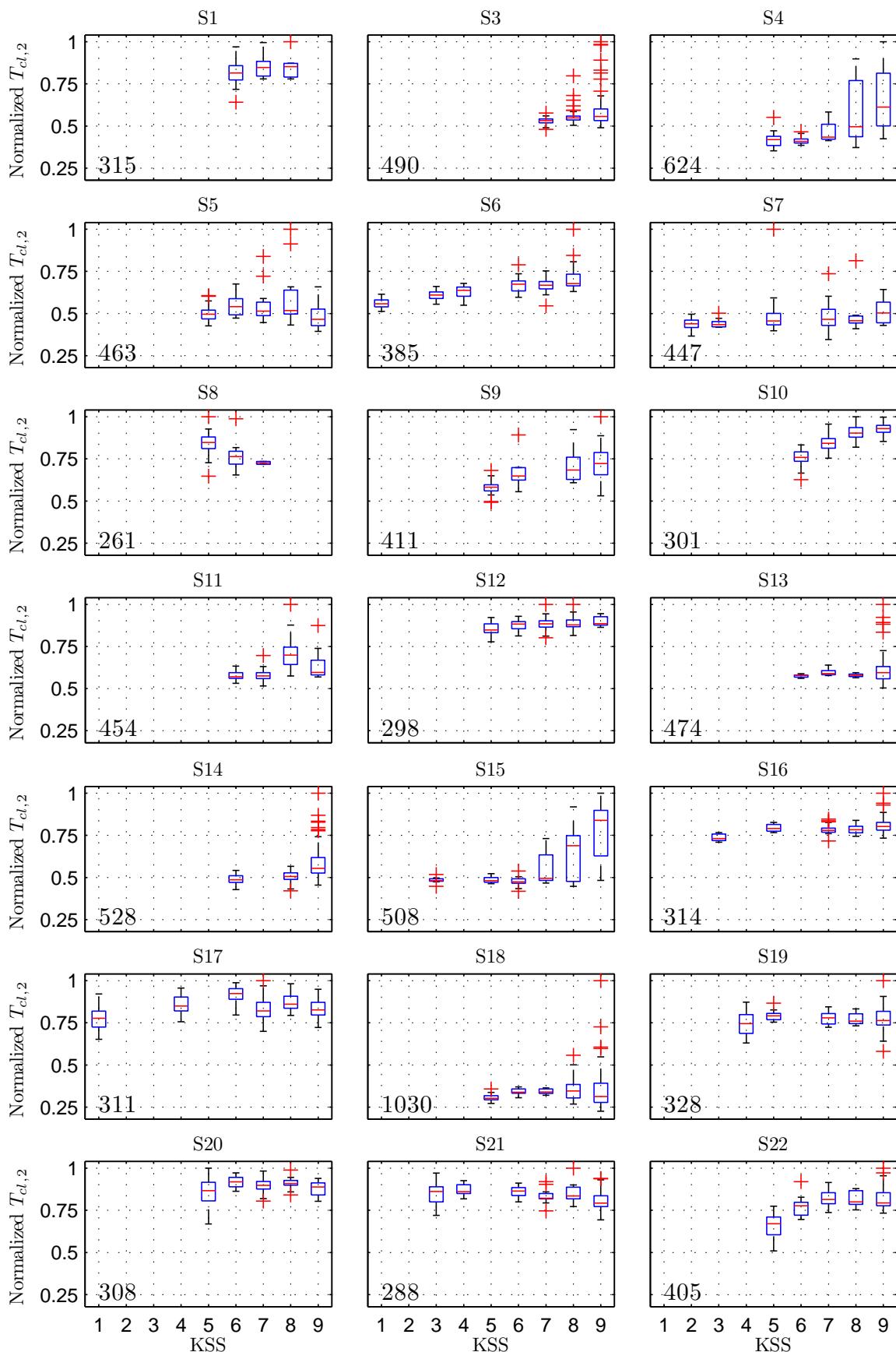


Figure F.28.: Boxplot of normalized feature $T_{cl,2}$ versus kss for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of $T_{cl,2}$ [ms] for each subject.

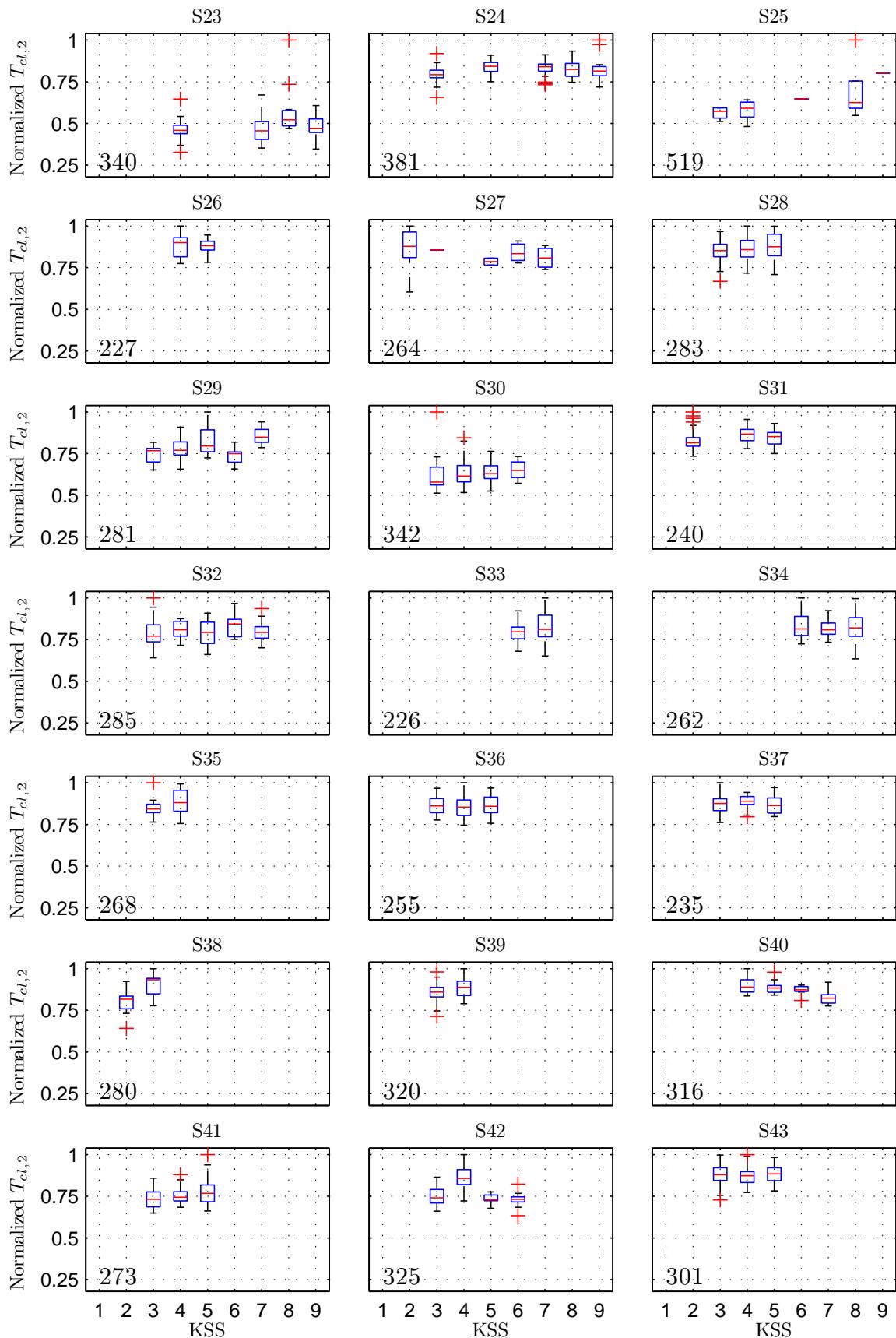


Figure F.29.: Boxplot of normalized feature $T_{cl,2}$ versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of $T_{cl,2}$ [ms] for each subject.

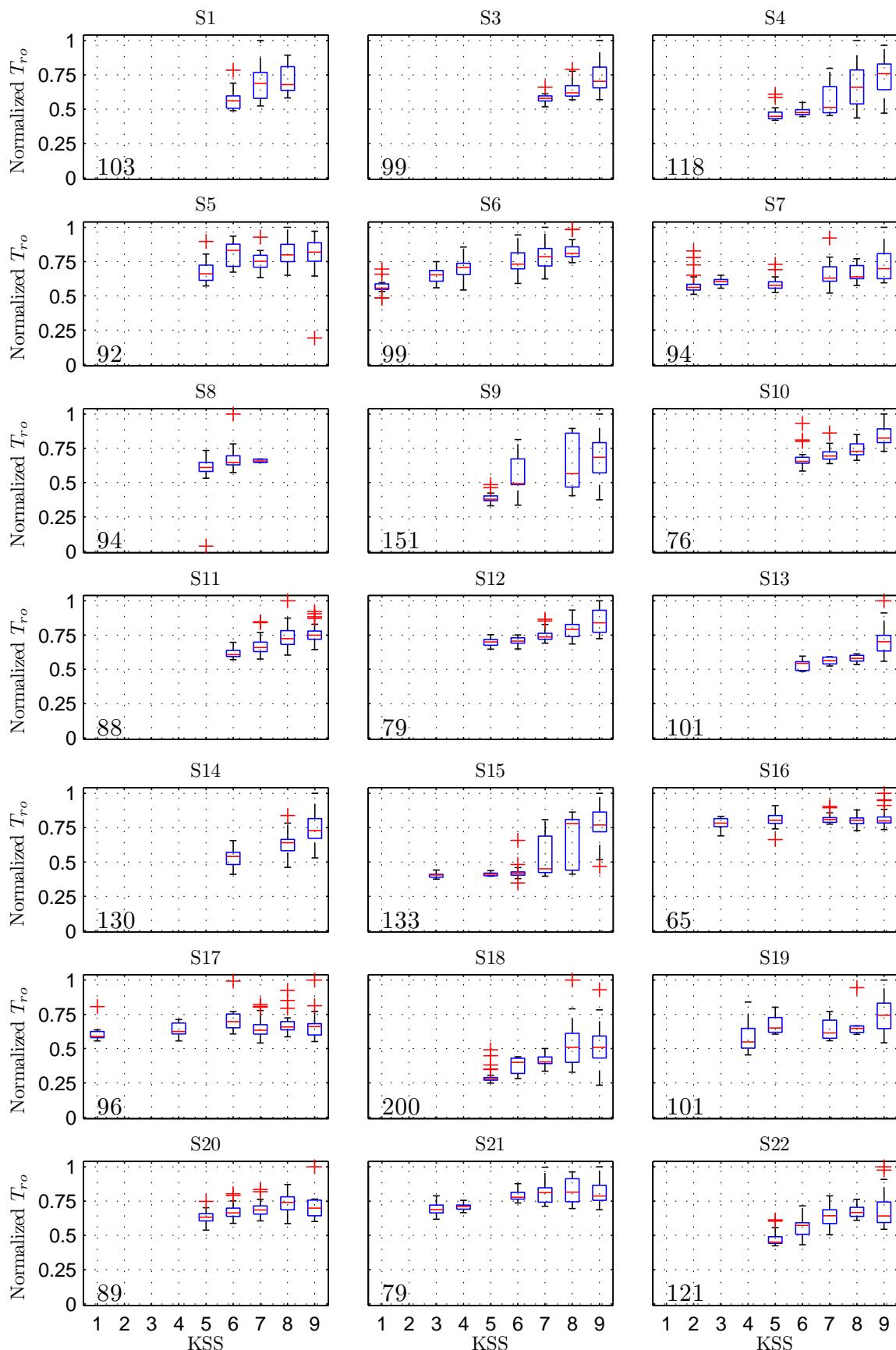


Figure F.30.: Boxplot of normalized feature T_{ro} versus KSS for subjects S1 to S22 (except for subject S2).
The values on the bottom left show the maximum of T_{ro} [ms] for each subject.

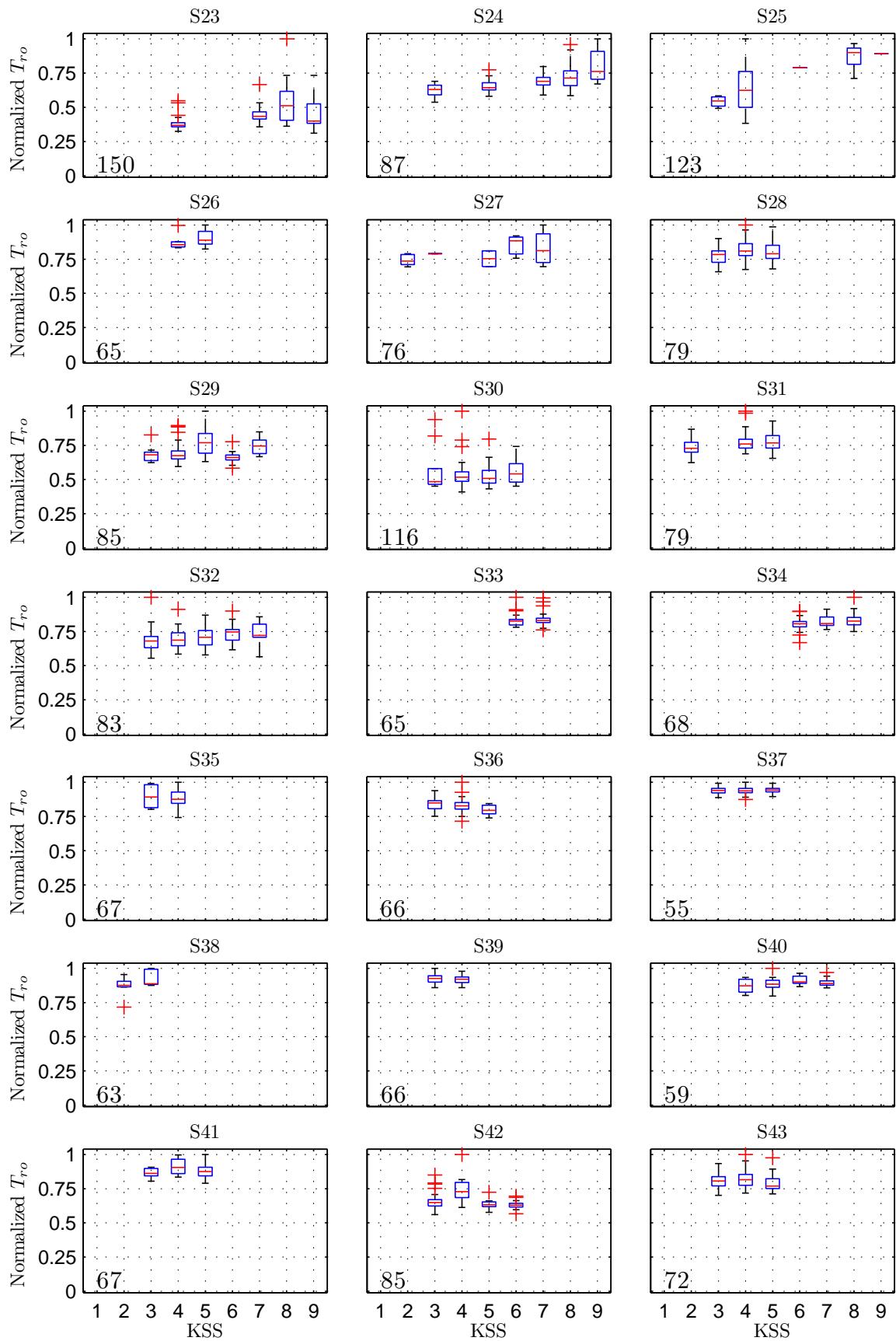


Figure F.31.: Boxplot of normalized feature T_{ro} versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of T_{ro} [ms] for each subject.

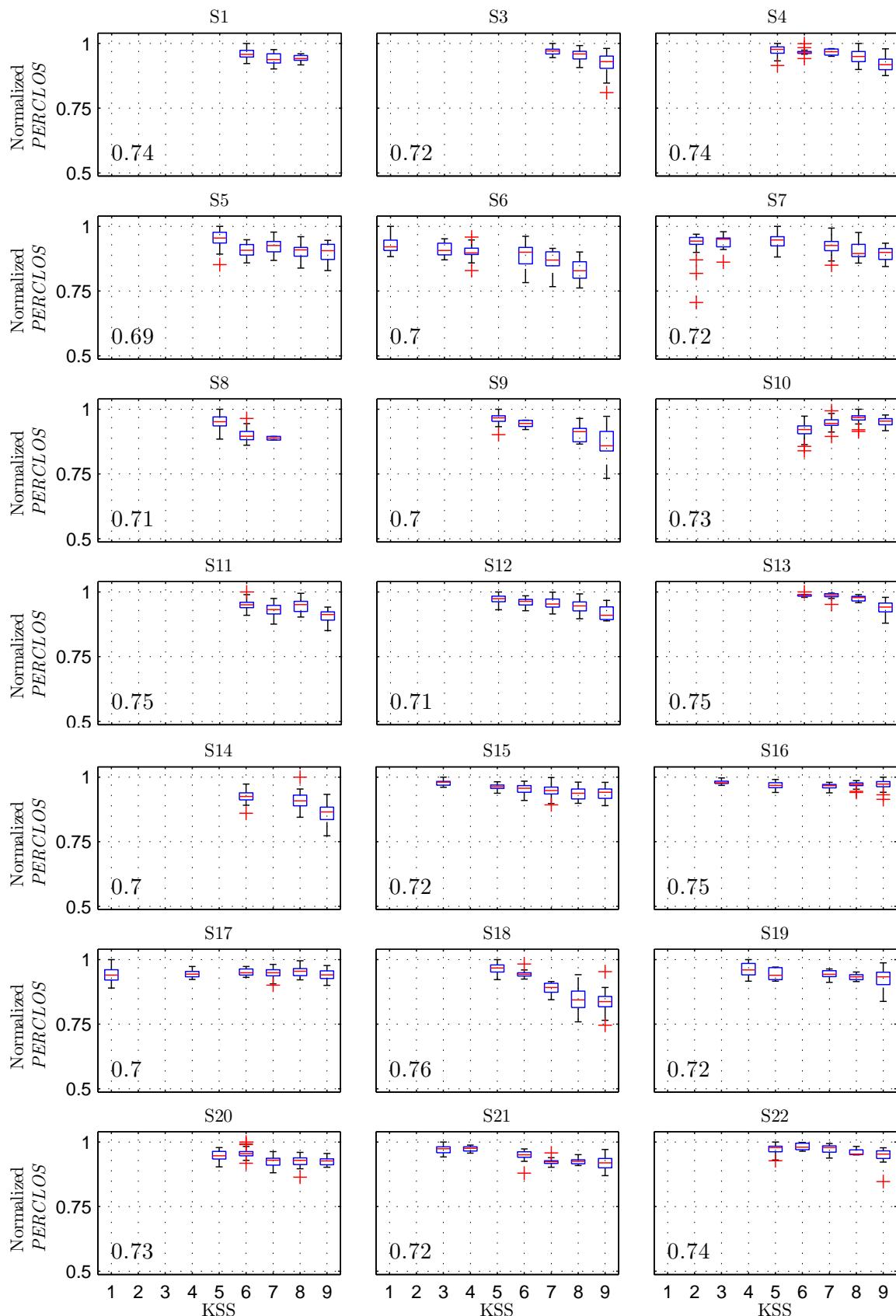


Figure F.32.: Boxplot of normalized feature *PERCLOS* versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of *PERCLOS* for each subject.

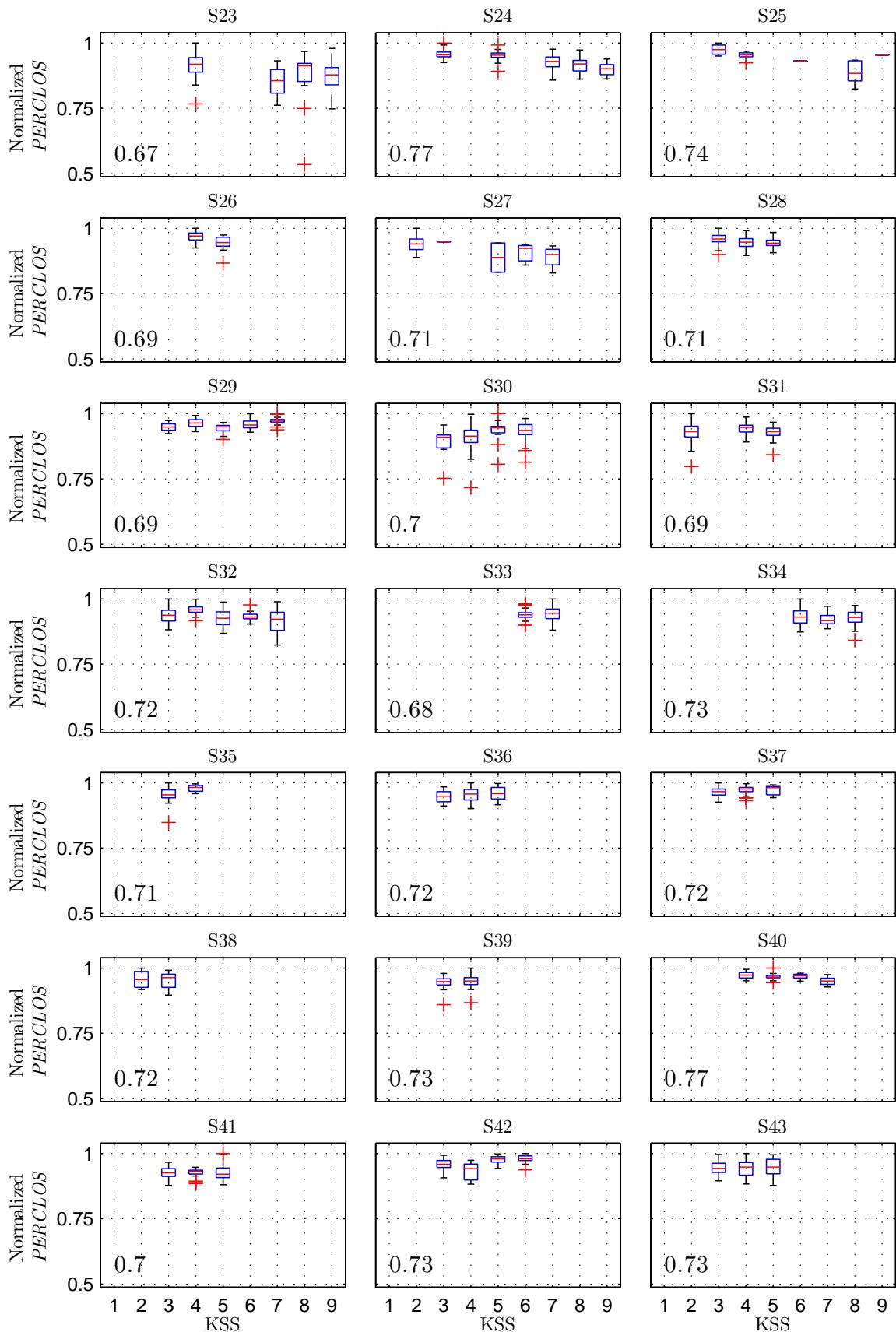


Figure F.33.: Boxplot of normalized feature *PERCLOS* versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of *PERCLOS* for each subject.

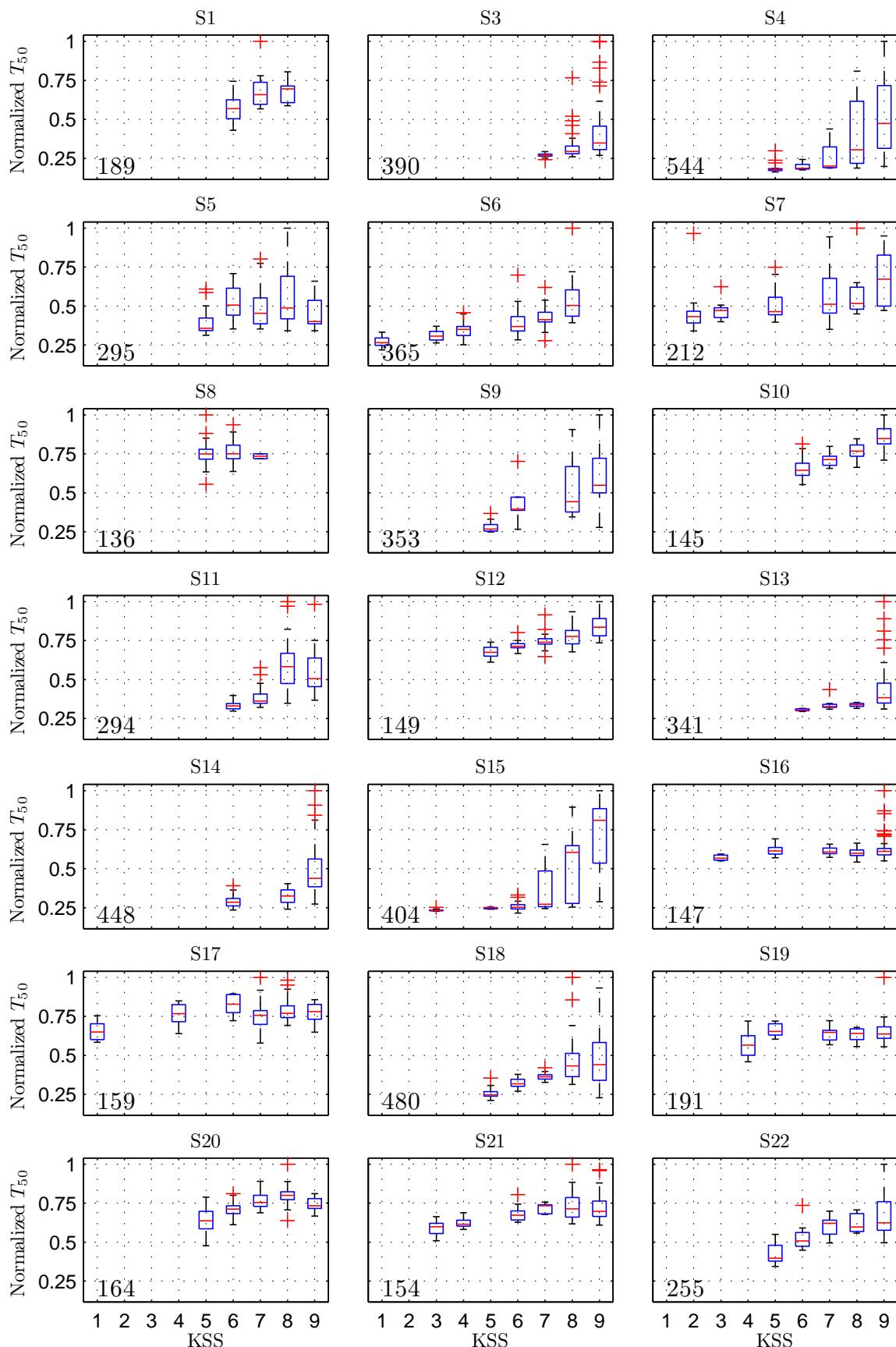


Figure F.34.: Boxplot of normalized feature T_{50} versus KSS for subjects S1 to S22 (except for subject S2).
The values on the bottom left show the maximum of T_{50} [ms] for each subject.

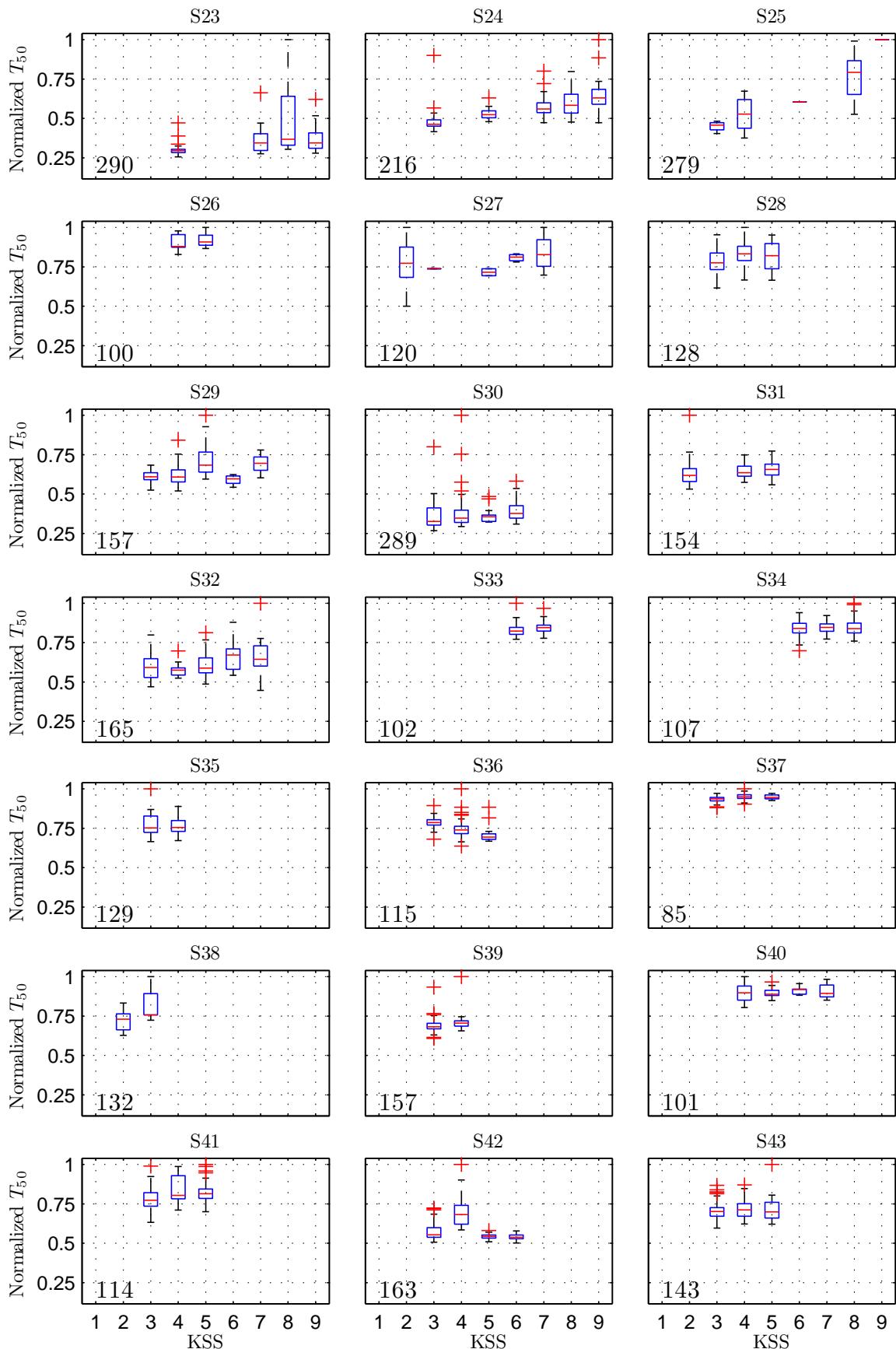


Figure F.35.: Boxplot of normalized feature T_{50} versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of T_{50} [ms] for each subject.

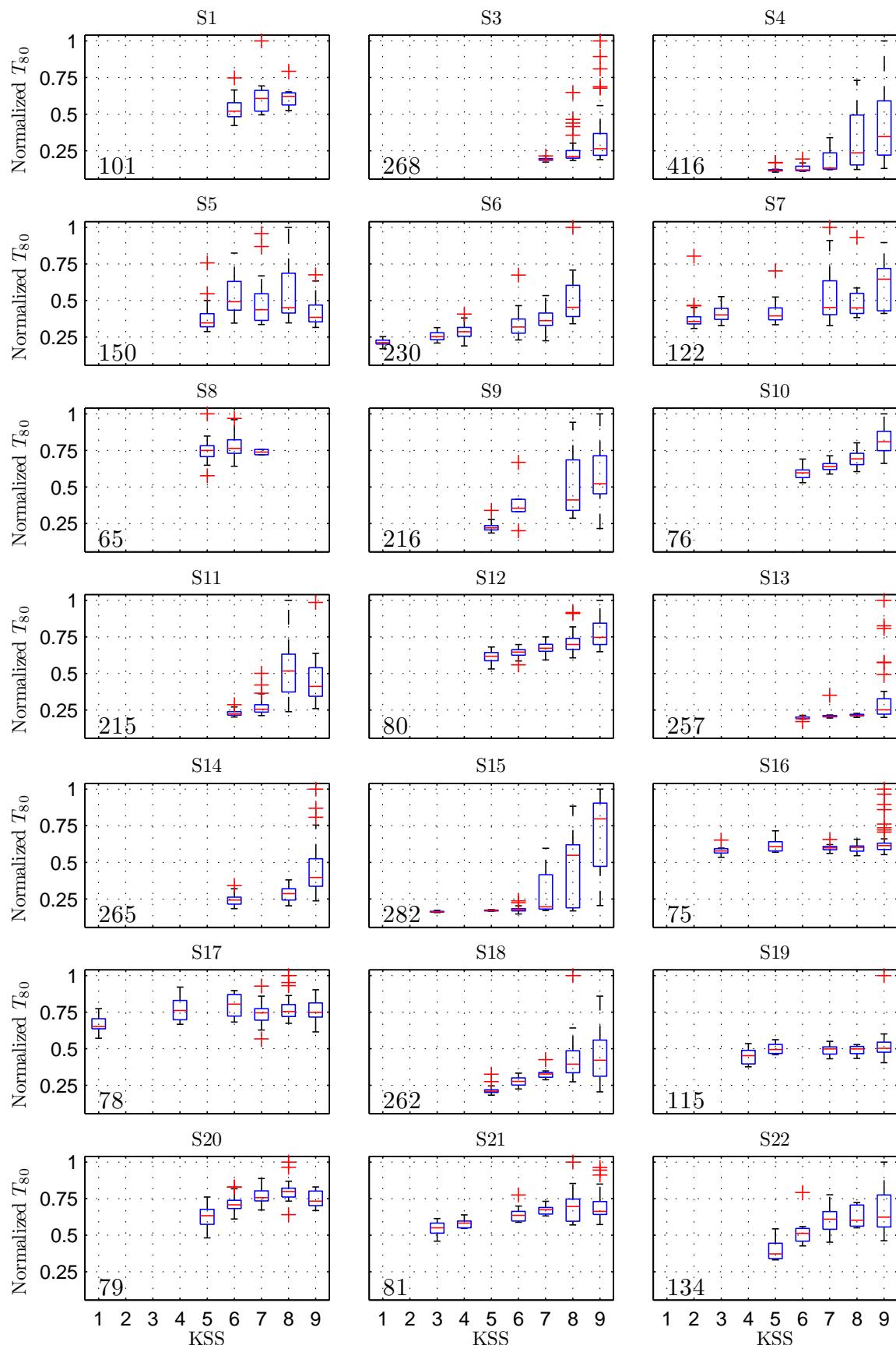


Figure F.36.: Boxplot of normalized feature T_{80} versus KSS for subjects S1 to S22 (except for subject S2).
The values on the bottom left show the maximum of T_{80} [ms] for each subject.

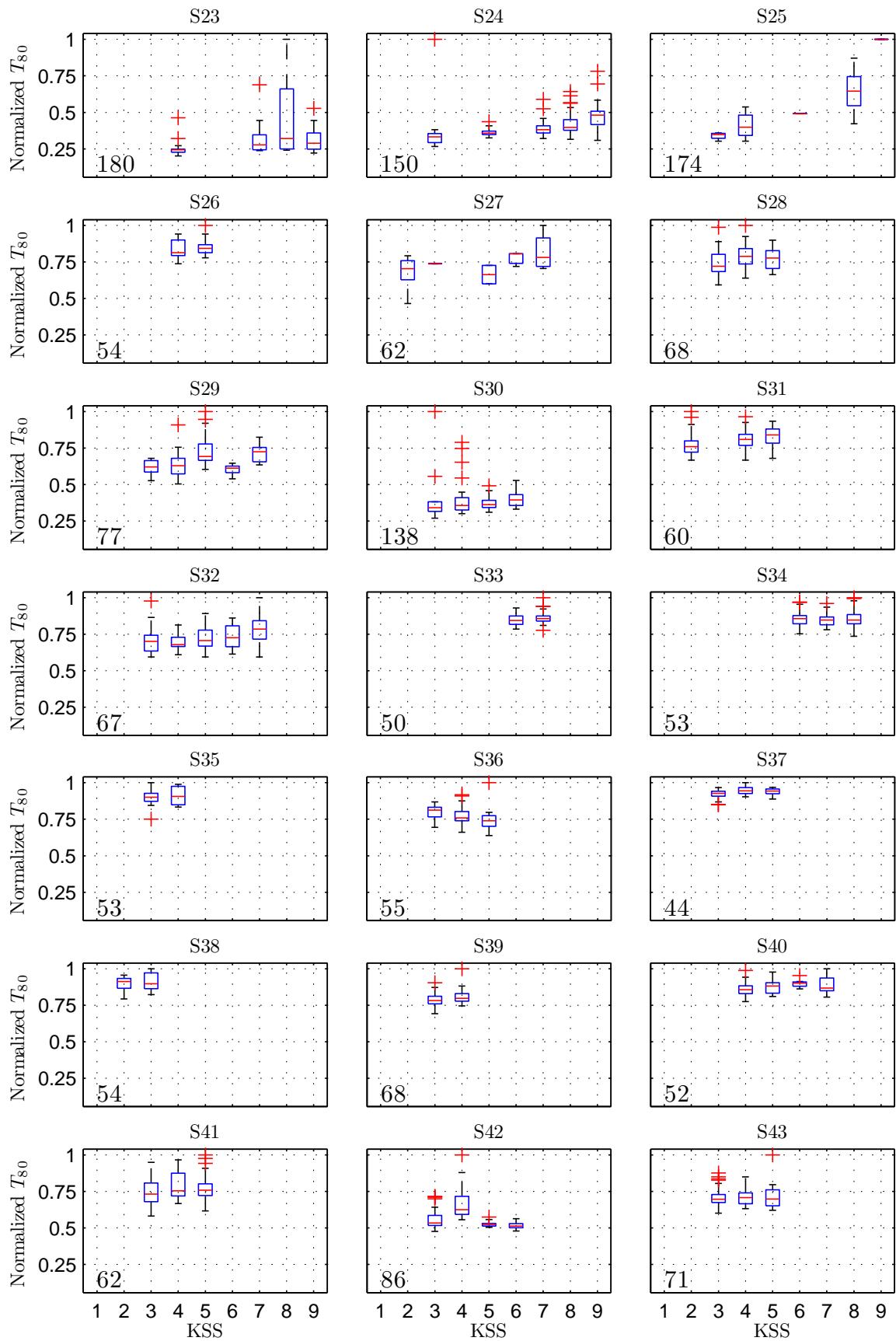


Figure F.37.: Boxplot of normalized feature T_{80} versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of T_{80} [ms] for each subject.

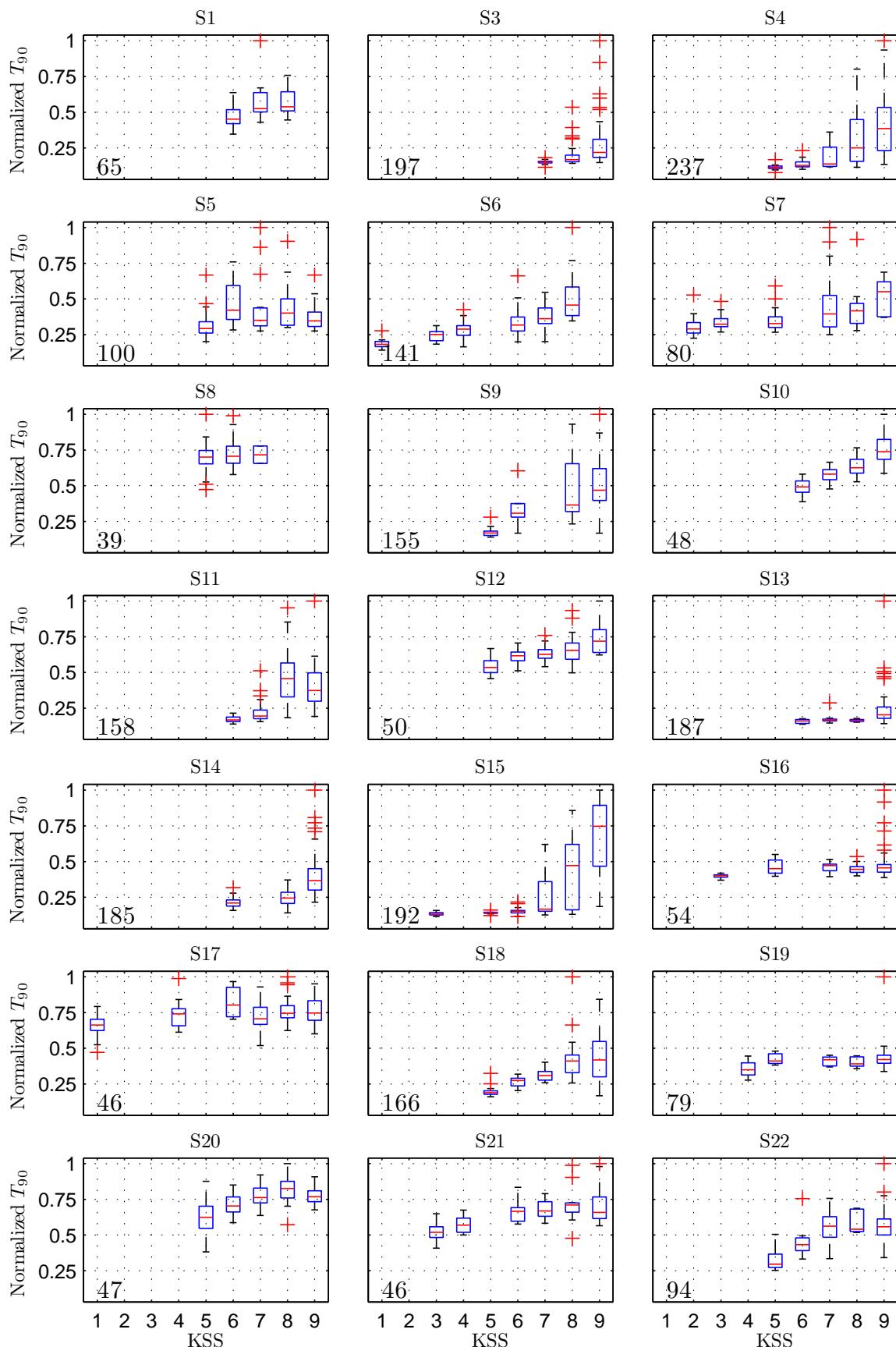


Figure F.38.: Boxplot of normalized feature T_{90} versus KSS for subjects S1 to S22 (except for subject S2).
The values on the bottom left show the maximum of T_{90} [ms] for each subject.

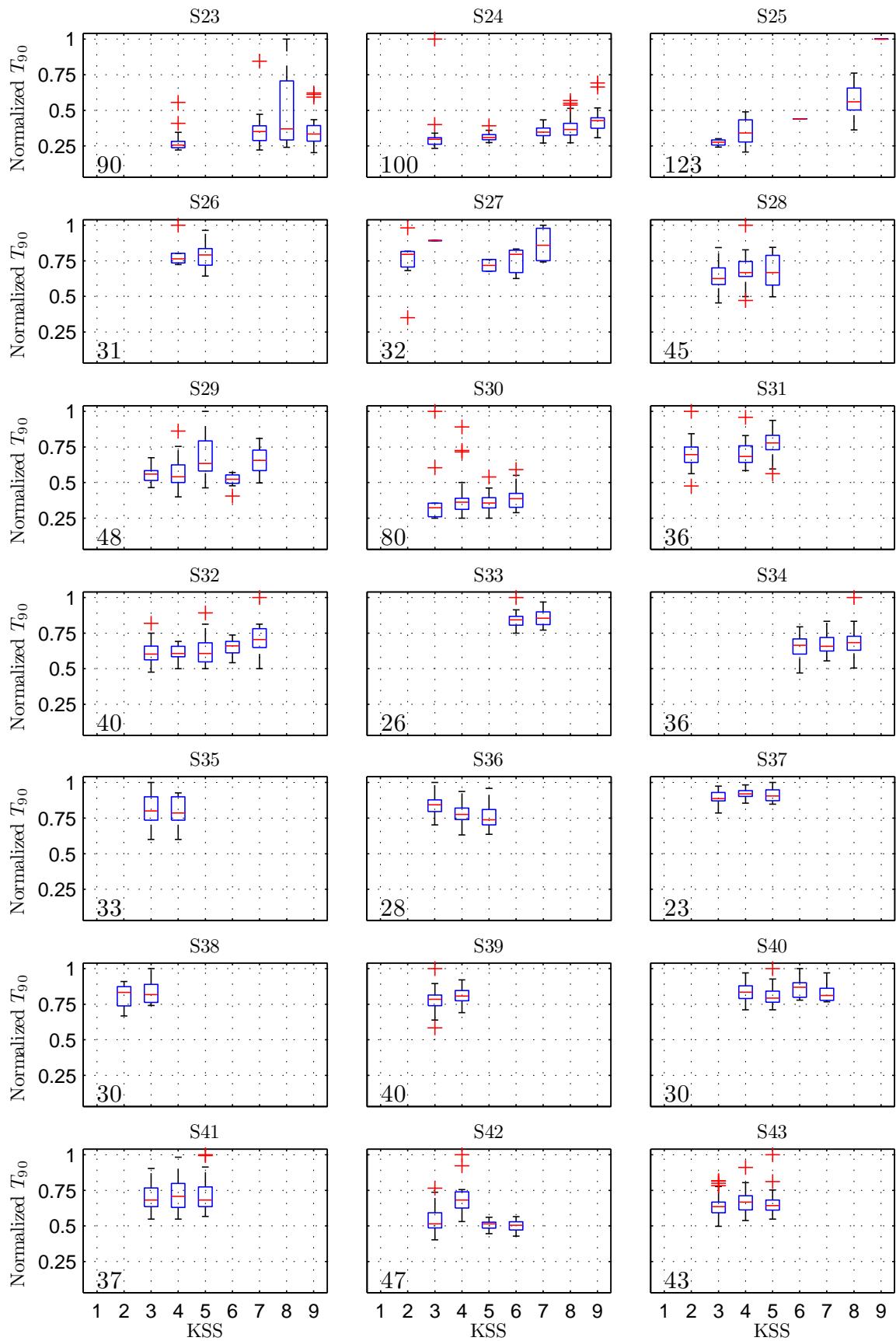


Figure F.39.: Boxplot of normalized feature T_{90} versus kss for subjects S23 to S43. The values on the bottom left show the maximum of T_{90} [ms] for each subject.

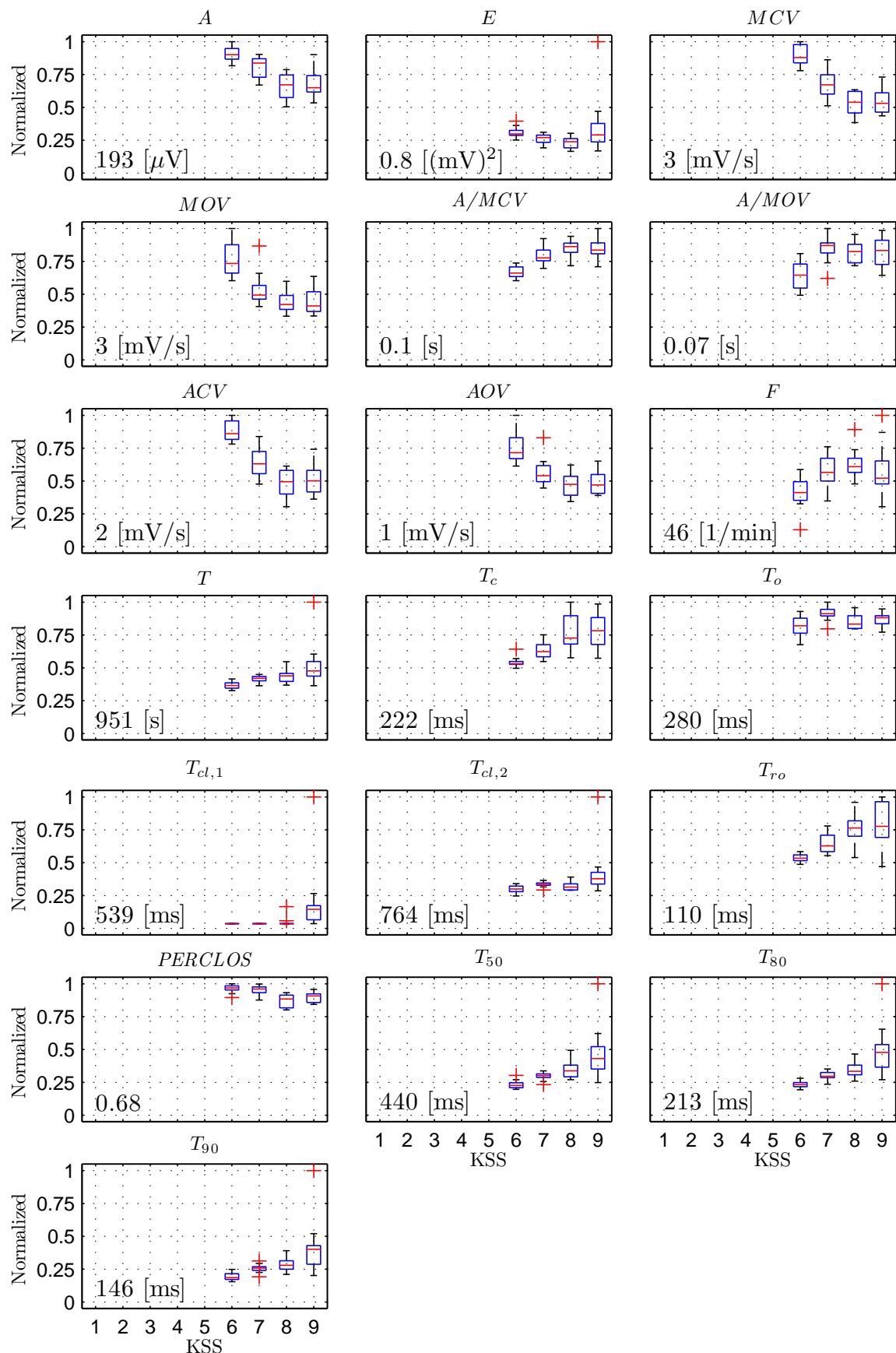


Figure F.40.: Boxplot of normalized features versus KSS values for subject S2. The maximum value of each feature is shown on the bottom left of plots.

F.3. Correlation between features using the Spearman's rank correlation coefficient

Figures F.41 and F.42 show the association between features for kss input-based features and drive time-based features with respect to the absolute value of the Spearman's rank correlation coefficient $|\rho_s|$. The p -value for feature pairs with the red \times were all larger than 0.05.

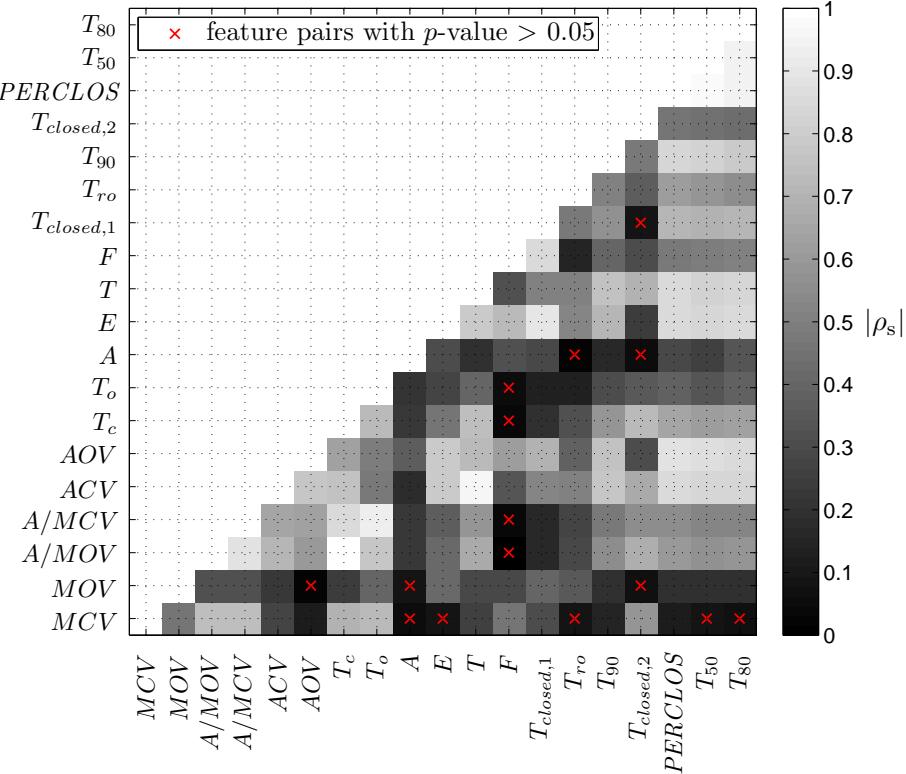


Figure F.41.: Absolute values of the Spearman's rank correlation coefficient $|\rho_s|$ calculated between kss input-based features

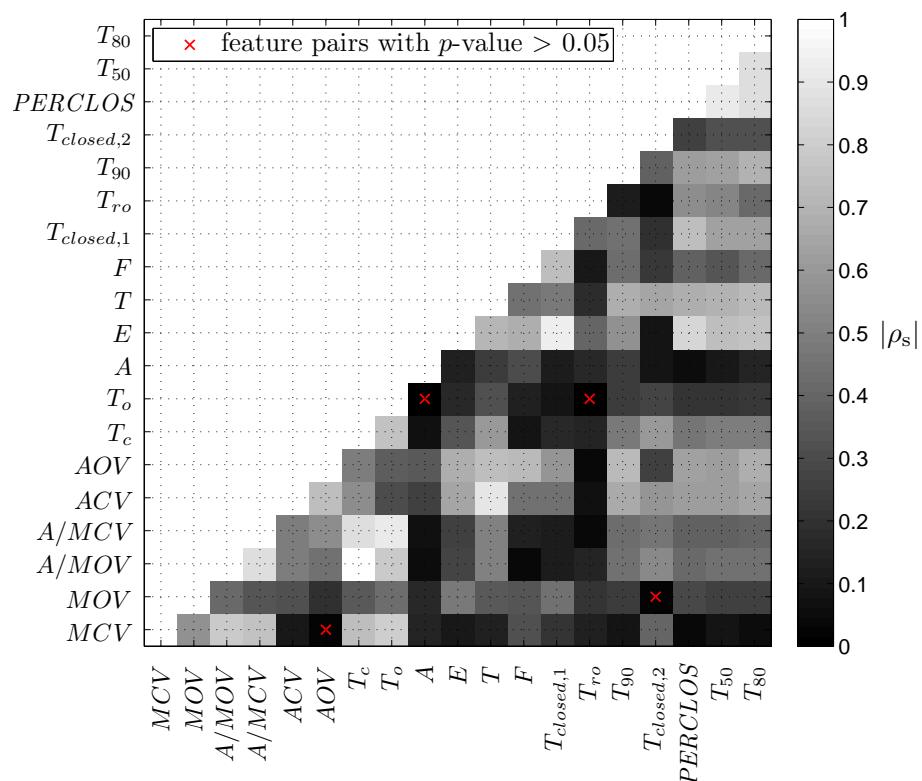


Figure F.42.: Absolute values of the Spearman's rank correlation coefficient $|\rho_s|$ calculated between drive time-based features

G. Gradient descent approach for training the ANN

In the following, it is explained how gradient descent is used for training the ANN. It is clear that the gradient descent approach in (8.19) should be calculated for the weights of each layer separately. Therefore, for the weights of hidden-to-output layer we have

$$\frac{\partial J}{\partial w_{jk}^{(2)}} = \frac{\partial J}{\partial net_k} \frac{\partial net_k}{\partial w_{jk}^{(2)}}. \quad (\text{G.1})$$

The last term can be easily calculated out of net_k definition in (8.15) which gives y_j . For the first term, the chain rule is applied as follows

$$\frac{\partial J}{\partial net_k} = \frac{\partial J}{\partial z_k} \frac{\partial z_k}{\partial net_k} = -(c_k - z_k) f'(net_k). \quad (\text{G.2})$$

$-\frac{\partial J}{\partial net_k}$ is also called *sensitivity* (ϑ_k) and denotes the changes of the training error J with respect to the net activation.

The final equation for updating weights of the hidden-to-output layer is

$$\Delta w_{jk}^{(2)} = \eta (c_k - z_k) f'(net_k) y_j = \eta \vartheta_k y_j. \quad (\text{G.3})$$

For the input-to-hidden layer, similarly, the chain rule is applied to (8.19) as follows

$$\frac{\partial J}{\partial w_{ij}^{(1)}} = \frac{\partial J}{\partial y_j} \frac{\partial y_j}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}^{(1)}}, \quad (\text{G.4})$$

where the first term is calculated as

$$\begin{aligned} \frac{\partial J}{\partial y_j} &= \frac{\partial J}{\partial z_k} \frac{\partial z_k}{\partial y_j} \\ &= \frac{\partial J}{\partial z_k} \frac{\partial z_k}{\partial net_k} \frac{\partial net_k}{\partial y_j} \\ &= - \sum_{k=1}^m \underbrace{(c_k - z_k) f'(net_k)}_{\vartheta_k} w_{jk}^{(2)}. \end{aligned} \quad (\text{G.5})$$

The other terms in (G.4) are calculated based on (8.14), namely $\frac{\partial y_j}{\partial net_j} = f'(net_j)$ and $\frac{\partial net_j}{\partial w_{ij}^{(1)}} = x_i$. Similar to the ϑ_k , ϑ_j is also defined as

$$\vartheta_j \equiv f'(net_j) \sum_{k=1}^m w_{jk}^{(2)} \vartheta_k, \quad (\text{G.6})$$

which links the sensitivity at a hidden layer to that of the output layer. Similar to (G.3), the learning rule of the input-to-hidden layer is

$$\Delta w_{ij}^{(1)} = \eta x_i f'(net_j) \sum_{k=1}^m w_{jk}^{(2)} \vartheta_k = \eta x_i \vartheta_j. \quad (\text{G.7})$$

Now it is clear, why this algorithm is called back-propagation. The reason is that it calculates the error and *propagates* it *back* as sensitivities ϑ_k from output to the hidden layer in order to learn the weights of the input to hidden layer. In other words, the error at a specific layer can be calculated, only if the error at the next layer is available.

H. On the understanding of the dual form of the optimization problem

H.1. Karush-Kuhn-Tucker theorem

For a better understanding of the Karush-Kuhn-Tucker theorem, following definitions should be reviewed first according to Chiang (2007) and Schmieder (2009):

- *Convex set*: a set \mathcal{C} is a convex set, if the line connecting any two points in \mathcal{C} also lies in \mathcal{C} (see Figure H.1), i.e.

$$(\theta x_1 + (1 - \theta) x_2) \in \mathcal{C} \quad \forall x_1, x_2 \in \mathcal{C}, \forall \theta \in [0, 1] \quad (\text{H.1})$$

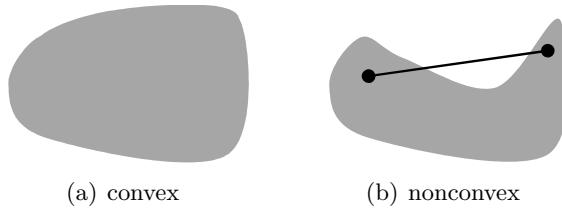


Figure H.1.: Examples of convex and non-convex sets

- *Convex function*: a function is referred to as a convex function, if the value of the function between any two points of it is smaller than the linear interpolation of these points (see Figure H.2). Mathematically, we have, $f: \mathbb{R}^n \mapsto \mathbb{R}$ is a convex function, if $\mathcal{C} = \mathbf{dom} f$ is a convex set and

$$f(\theta x_1 + (1 - \theta) x_2) \leq \theta f(x_1) + (1 - \theta) f(x_2) \quad \forall x_1, x_2 \in \mathcal{C}, \forall \theta \in [0, 1]. \quad (\text{H.2})$$

$\mathbf{dom} f$ denotes the domain of the function f .

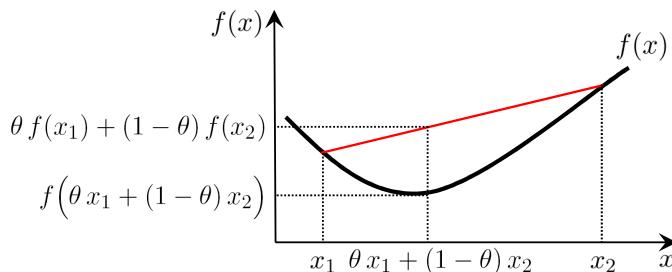


Figure H.2.: An example of a convex function

- *Affine function*: an affine function is both convex and concave. A function f is concave, if $-f$ is convex.

Theorem (Karush-Kuhn-Tucker Theorem (Cristianini and Shawe-Taylor, 2000))
Given an optimization problem with convex domain $\Omega \subseteq \mathbb{R}^n$,

$$\begin{aligned} & \text{minimise} && f(\mathbf{w}), \quad \mathbf{w} \in \Omega \\ & \text{subject to} && g_i(\mathbf{w}) \leq 0, \quad i = 1, \dots, k, \\ & && h_i(\mathbf{w}) = 0, \quad i = 1, \dots, m, \end{aligned}$$

with $f \in \mathcal{C}^1$ convex and g_i, h_i affine, the necessary and sufficient conditions for a normal point \mathbf{w}^* to be an optimum are the existence of $\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*$ such that

$$\begin{aligned} \frac{\partial \mathcal{L}(\mathbf{w}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)}{\partial \mathbf{w}} &= \mathbf{0}, \\ \frac{\partial \mathcal{L}(\mathbf{w}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}} &= 0, \\ \alpha_i^* g_i(\mathbf{w}^*) &= 0, \quad i = 1, \dots, k, \\ g_i(\mathbf{w}^*) &\leq 0, \quad i = 1, \dots, k, \\ \alpha_i^* &\geq 0, \quad i = 1, \dots, k. \end{aligned}$$

\mathcal{L} denotes the Lagrangian function defined in (8.27).

H.2. Extraction of the dual problem for the soft margin classifier

Similar to the linearly separable data set, the Lagrangian function for the L_1 -norm case of the soft margin SVM can be defined base on Lagrangian multipliers α_i and β_i

$$\mathcal{L}(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i (y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i) - \sum_{i=1}^N \beta_i \xi_i, \quad (\text{H.3})$$

where $\boldsymbol{\alpha} = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_N]^T$ and $\boldsymbol{\beta} = [\beta_1 \ \beta_2 \ \dots \ \beta_N]^T$. The conditions of the KKT, which should be fulfilled, are as follows

$$\frac{\partial \mathcal{L}(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \mathbf{w}} = \mathbf{0} \implies \mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \quad (\text{H.4})$$

$$\frac{\partial \mathcal{L}(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial b} = 0 \implies \sum_{i=1}^N \alpha_i y_i = 0. \quad (\text{H.5})$$

$$\frac{\partial \mathcal{L}(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \boldsymbol{\xi}} = \mathbf{0} \implies \alpha_i + \beta_i = C, \quad i = 1, \dots, N. \quad (\text{H.6})$$

$$\alpha_i (y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i) = 0, \quad i = 1, \dots, N \quad (\text{H.7})$$

$$\beta_i \xi_i = 0, \quad i = 1, \dots, N \quad (\text{H.8})$$

$$\alpha_i \geq 0, \quad \beta_i \geq 0, \quad \xi_i \geq 0, \quad i = 1, \dots, N. \quad (\text{H.9})$$

Based on the above equations three cases may happen

- $\alpha_i = 0 \implies \xi_i = 0$. As mentioned in Section 8.3.2, this corresponds to the correct classification of \mathbf{x}_i .

- $0 < \alpha_i < C \implies y_i (\mathbf{w}^T \mathbf{x}_i + b) + \xi_i = 1$ and $\xi_i = 0$, which leads to $y_i (\mathbf{w}^T \mathbf{x}_i + b) = 1$ and \mathbf{x}_i is consequently support vector.
- $\alpha_i = C \implies y_i (\mathbf{w}^T \mathbf{x}_i + b) + \xi_i = 0$ and $\xi_i \geq 0$, which means that \mathbf{x}_i is support vector. Depending on the value of ξ_i , \mathbf{x}_i might be misclassified (see Section 8.3.2).

List of figures

1.1	The evolution of number of killed and injured persons in traffic accidents of Germany	1
1.2	The evolution of number of vehicle accidents due to driver drowsiness and the number of injured persons involved in them	2
1.3	A typical steering event detected by the Attention Assist as a drowsiness-related steering wheel movement	8
1.4	Tool chain of this thesis	10
2.1	32-electrode arrangement of EEG (excluding 4 electrodes for eye movement data collection)	17
2.2	ActiCAP measurement system for EEG recording by Brain Products GmbH	17
2.3	EEG signals showing α -bursts with closed eyes versus open eyes	18
2.4	Frequency components of the α -bursts by applying the Fourier transform to the wave of the O2 electrode shown in Figure 2.3	18
2.5	Sensitivity of the ASR to auditory and visuomotor secondary tasks and the corresponding number of horizontal saccades	20
2.6	ASR with different control signals	21
2.7	ASR before and recalculated after applying the control signal	22
2.8	EOG electrodes attached around the eyes for collecting horizontal and vertical eye movement data	24
2.9	An example of the drift in the collected EOG data (vertical component)	25
3.1	Structure of the human eye while transmitting the ray of light	32
3.2	Eye muscles	32
3.3	Different categories of eye movements based on their velocity	32
3.4	Representative examples of blinks measured by the vertical ($V(n)$) and horizontal ($H(n)$) components of the EOG	34
3.5	$H(n)$ and $V(n)$ representing different types of saccades due to horizontal, vertical and diagonal eye movements	35
4.1	Selected tracks of the Applus+ IDIADA proving ground	38
4.2	$H(n)$ and $V(n)$ of subject S2 for baseline P1 of track 1	40
4.3	$H(n)$ and $V(n)$ of subject S2 for baseline of track 2	40
4.4	Spectrogram and power spectral density (PSD) of 20s of $V(n)$ of subject S4 for track 1 and track 2	41
4.5	Boxplot of the values listed in Table 4.1	42
4.6	Detected large amplitude bumps based on the EWVAR of wheel speed sensor data	42
4.7	$V(n)$ of subject S2 and S6 for small (top plot) and large (bottom plot) amplitude bumps of track 3	43
4.8	$H(n)$ of subject S8 for track 4, top/bottom: right/left curve	44
4.9	Calculated sawtooth occurrence time period Δt_c versus curve radius r for $\delta = 4^\circ, 6^\circ, 10^\circ$	45
4.10	Daytime experiment's route, about 130 km	47
4.11	One block of daytime real road experiment with secondary tasks	48
4.12	In-vehicle setup of the daytime experiment with secondary tasks	49
4.13	Nighttime experiment's route, about 450 km	50
5.1	Drift removal by applying a median filter to $V(n)$ to improve blink detection. Top: awake phase, bottom: drowsy phase	54
5.2	Information loss of slow blinks by median filter method	54

5.3	$V(n)$ and its derivative $V'(n)$ representing eye blinks during the awake phase	56
5.4	Normalized histogram of all detected potential blinks and their clustering thresholds by the k -means clustering method for 11 subjects	57
5.5	Simultaneous detection of saccades by eye blink detection algorithm	57
5.6	Normalized histogram of all detected potential saccades and blinks with long eye closure. Their clustering thresholds are also shown.	58
5.7	Possible combinations of two vertical saccades in $V(n)$	59
5.8	Flow chart of the derivate-based method for blink detection	60
5.9	The impact of window length L_{win} on the efficiency of the STFT	63
5.10	Examples of typical mother wavelets	64
5.11	Scaling and translation of the mother wavelet with varying a and b	64
5.12	Scalogram of the cwt for $x(t)$ signal shown in Figure 5.9, top plot: $1 \leq a \leq 256$, bottom plot: $1 \leq a \leq 20$	65
5.13	Scalograms of the cwt with different mother wavelets for $V(n)$ signal of the awake phase (left plots) and the drowsy phase (right plots) of the drive	66
5.14	cwt with different mother wavelets for $V(n)$ signal of the awake phase (left plots) and the drowsy phase (right plots) of the drive with $a = 5, 10$ and 15	68
5.15	Comparison of $\mathcal{X}_\psi(a, b)$ with the Haar wavelet at $a = 5, 10, 15, 30$ and 100 with the negative of the derivative of the EOG signal $-V'(n)$ for the awake and drowsy phases of the drive .	69
5.16	Comparison of CWT at $a = 10, 30$ and 100 for the detection of fast (the first 20 s) and slow (the last 20 s) blinks	70
5.17	Detected and accepted peaks at different scales of $\mathcal{X}_\psi(a, b)$ signals	70
5.18	Flow chart of the cwt-based method for blink detection	72
5.19	Schematic of spaces spanned by scaling and wavelet functions	74
5.20	J -stage decomposition tree	76
5.21	Three-stage decomposition of the EOG signal during the awake phase by db4 wavelet . .	77
5.22	Three-stage decomposition of the EOG signal during the drowsy phase by db4 wavelet . .	78
5.23	J -stage reconstruction tree	79
5.24	Example 1: denoising of the EOG signal by removing different coefficients during the reconstruction	80
5.25	Example 2: denoising of the EOG signal by removing different coefficients during the reconstruction	81
5.26	Scatter plot: $\varepsilon_1(n)$ versus $\varepsilon_2(n)$ shown in Figure 5.25	82
5.27	Two examples of drift removal with the wavelet decomposition and reconstruction for awake (top) and drowsy (bottom) phases of $V(n)$	83
5.28	RC and PC of vertical saccade and blink detections for the derivative-based algorithm and the median filter-based method during the awake and drowsy phases	84
5.29	Average duration (first row), amplitude (second row) and number of blinks (third row) versus self-estimated drowsiness level for subjects S15, S16 and S18 based on the derivative-based algorithm and the median filter-based method	85
5.30	RC and PC of blink detection for the derivative based algorithm and the wavelet transform method during the awake and drowsy phases	87
5.31	Setting the threshold for distinguishing between blinks and vertical saccades in an online implementation of the detection method	89
6.1	Saccade rate for the variable time-on-task (four blocks) for all subjects	92
6.2	Percentage of saccades time-locked to blinks for all subjects and all blocks during the visuomotor task	93
6.3	Percentage of saccades time-locked to blinks with respect to saccade direction averaged over all blocks during the visuomotor task	94
6.4	Scatter plot: number of saccades accompanied by blinks with respect to their direction during the visuomotor task for all subjects. Ellipses show two clusters.	94
6.5	Scatter plots of blink rate for visuomotor vs. driving and auditory vs. driving task. Pearson correlation coefficient (ρ_p) and the corresponding p -values are provided as well.	96

6.6	Percentage of blinks time-locked to saccades for all subjects averaged over all blocks during the visuomotor task	97
6.7	Scatter plot: blink rate versus saccade rate during the visuomotor task	97
6.8	EOG signals during the visuomotor and driving task for subject S8	98
6.9	EOG signals during the visuomotor and driving task for subject S1	98
6.10	Algorithm for determining the threshold of horizontal saccade detection	99
6.11	Histogram: absolute amplitude of saccades out of $H(n)$ signal for subject S1	100
6.12	The algorithm for balancing the number of small (N_s) and large-amplitude (N_l) saccades	100
6.13	Normalized histogram: amplitude of all horizontal saccades (dark bars) and those accompanied by blinks (light bars) for 12 subjects	101
6.14	Scatter plot: number of saccades in percent time-locked to the blinks with respect to their amplitude, i.e. small and large	101
7.1	KSS input-based feature aggregation method	104
7.2	Relative frequency of KSS values for two feature aggregation methods	105
7.3	Drive time-based feature aggregation method	106
7.4	MOV feature before and after baselining	107
7.5	$V(n)$ and its derivative $V'(n)$ representing eye blinks in awake and drowsy phases with the corresponding features	109
7.6	Boxplot of normalized drive time-based features combined for all subjects versus KSS values	110
7.7	Histogram of MCV and MOV with the estimated (est.) distribution (dist.) curves	111
7.8	Histogram of ACV and AOV with estimated (est.) distribution (dist.) curves. The outliers (values > 10 mV/s) are not shown.	112
7.9	Histogram of T_c and T_o with the estimated (est.) distribution (dist.) curves	114
7.10	Best linear fit to all baselined feature values of T_c and T_o	114
7.11	Comparison between T_c , $T_{cl,1}$ and T_o during the awake and drowsy phases of the drive for all subjects	115
7.12	Best linear fit to all baselined feature values of T_{ro}	116
7.13	Scatter plot of T versus T_x , $x \in \{50, 80, 90\}$. The red line shows the best linear fit with its equation on the top of each plot.	119
7.14	Examples of intended (takeover maneuver) and unintended (lane departure) lane change events visible in the vehicle's lateral distance signal	123
7.15	The mean of the EWVAR of lateral distance versus KSS values for 25 subjects who drove in the driving simulator. The standard deviations are also shown.	125
7.16	The mean of all baselined features over the first and the last 5 min before the first unintended lane departure event for 23 subjects who drove in the driving simulator	126
7.17	The mean of all baselined features over the first and the last 5 min of the drive for 18 subjects who drove under real conditions	128
7.18	The mean of all baselined features over the first and the last 5 min before an unintended microsleep for 11 subjects who drove in the driving simulator	130
7.19	Boxplot of baselined KSS input-based features for all subjects versus KSS values	133
7.20	Absolute values of Pearson correlation coefficient calculated between KSS input-based features	136
7.21	Absolute values of the Pearson correlation coefficient calculated between drive time-based features	136
7.22	Scatter plot: comparison of 50-Hz features with 40- and 30-Hz ones for the first 12 subjects - part 1	137
7.23	Scatter plot: comparison of 50-Hz features with 40- and 30-Hz ones for the first 12 subjects - part 2	138
7.24	Scatter plot: comparing AOV and MOV extracted based on 30, 40 and 50 Hz sampling rate. The lines indicate the best linear fits.	139
8.1	Examples of three classification rules	143
8.2	Distribution of classes for KSS input-based and drive time-based features	143

8.3	Applying the SMOTE to an imbalanced data set	148
8.4	Applying the SMOTE and the Tomek link cleaning technique to an imbalanced data	149
8.5	Architecture of a feed-forward neural network with 3 inputs, 3 neurons in one hidden layer and 2 outputs	150
8.6	Mathematical representation of the input-to-hidden layer of a network	150
8.7	Sigmoid activation function	151
8.8	Supervised classification by the ANN	152
8.9	ADR of the training and test sets of the binary subject-dependent ANN classifier for different numbers of neurons. Feature type: kss input-based features. Bars refer to the standard deviation of permutations.	154
8.10	ADR of the training and test sets of the binary and 3-class subject-dependent ANN classifier for different numbers of neurons. Feature type: drive time-based features. Bars refer to the standard deviation of permutations.	155
8.11	ADR of the training and test sets of the binary subject-dependent ANN classifier for different numbers of neurons. Feature type: imbalanced and balanced by SMOTE kss input-based features of driving simulator experiment. Bars refer to the standard deviation of permutations.	156
8.12	ADR of the kss input-based features for the real road experiment applied to the network trained based on the SMOTE. Bars refer to the standard deviation of permutations.	157
8.13	Comparing confusion matrix of the binary subject-dependent ANN classifier with that of the subject-independent	159
8.14	Different separating hyperplanes	160
8.15	Example of a linearly inseparable data	161
8.16	An example of feature mapping for a linearly inseparable data set	163
8.17	An example of the grid search for finding (C_0, γ_0) and $(C_{\text{opt}}, \gamma_{\text{opt}})$	165
8.18	j -fold cross validation method	165
8.19	An example of a 3-class classification with shaded areas as the unclassifiable. The arrows show the positive sides of the hyperplanes. Decision functions for the One-Against-All approach: \mathcal{H} : $\mathbf{w}_J^T \Phi(\mathbf{x}_i) + b_J = 0, J = 1, 2, 3$ and for the One-Against-One approach: \mathcal{H} : $\mathbf{w}_{I,J}^T \Phi(\mathbf{x}) + b_{I,J} = 0, I = 1, 2, 3, J = 1, 2, 3$ and $I \neq J$	166
8.20	Boxplot of C , γ , training and test accuracies for the balanced and imbalanced 2-class subject-dependent classification with the SVM for all 100 permutations. Feature type: kss input-based features	169
8.21	Boxplot of C , γ , training and test accuracies for the 2-class and 3-class subject-dependent classification with the SVM for all 100 permutations. Feature type: drive time-based features	170
8.22	Comparing confusion matrix of the binary subject-dependent SVM classifier with that of the subject-independent	172
8.23	ADR of the test sets of the binary subject-dependent k -NN classifier for different numbers of neighbors. Feature type: kss input-based features. Bars refer to the standard deviation of permutations.	173
8.24	ADR of the 2-class and 3-class subject-dependent k -NN classifier for different numbers of neighbors. Feature type: drive time-based features. Bars refer to the standard deviation of permutations.	174
8.25	ADR of the test sets of the binary subject-dependent k -NN classifier for different numbers of neighbors. Feature type: imbalanced kss input-based features of driving simulator experiment. Bars refer to the standard deviation of permutations.	175
8.26	Comparing confusion matrix of the binary subject-dependent k -NN classifier with that of the subject-independent	176
8.27	Comparing confusion matrices of the binary ANN, SVM and k -NN classifiers for the subject-dependent and subject-independent classifications. Feature type: drive time-based features	176
8.28	Comparing confusion matrices of the 3-class ANN, SVM and k -NN classifiers for the subject-dependent classification. Feature type: drive time-based features	177
8.29	Comparing confusion matrices of the binary subject-dependent ANN, SVM and k -NN classifiers for different kss input-based features	178

8.30	ADR of the training and test sets of the binary subject-dependent ANN classifier for different numbers of neurons based on GSRD case. Feature type: drive time-based features. Bars refer to the standard deviation of permutations.	180
8.31	Comparing confusion matrices of the binary ANN ($N_h = 10$), SVM and k -NN ($k = 7$) classifiers for the subject-dependent and subject-independent classifications of the GSRD case. Feature type: drive time-based features.	180
8.32	Boxplot of C , γ , training and test accuracies for the 2-class subject-dependent classification of the GSRD case with SVM for all 100 permutations. Feature type: drive time-based features	181
8.33	ADR of the 2-class subject-dependent k -NN classifier of the GSRD case for different numbers of neighbors. Feature type: KSS input-based features. Bars refer to the standard deviation of permutations.	182
8.34	Comparing confusion matrices of the binary subject-independent ANN ($N_h = 10$), SVM ($C = 90.5$, $\gamma = 1.4$) and k -NN ($k = 7$) classifiers for unseen real road drives of drop-outs. Feature type: drive time-based features.	182
8.35	The ANN classification accuracy of the best selected features by the SFFS algorithm from 1 to 10-feature combination. Feature type: drive time-based features.	185
8.36	M_R values of the best 10-feature combinations calculated based on the Pearson and Spearman's rank correlation coefficients	189
A.1	Geometrical representation of tracking two successive TPs during a curve negotiation	197
B.1	Description of a boxplot representation	199
E.1	Scaling and wavelet functions of two mother wavelets	211
F.1	Comparison of Spearman's rank correlation coefficient between statistical metrics of features and KSS values. Feature type: drive time-based features	213
F.2	Boxplot of normalized feature A versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of A [μ V] for each subject.	215
F.3	Boxplot of normalized feature A versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of A [μ V] for each subject.	216
F.4	Boxplot of normalized feature E versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of E [(mV) ²] for each subject.	217
F.5	Boxplot of normalized feature E versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of E [(mV) ²] for each subject.	218
F.6	Boxplot of normalized feature MCV versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of MCV [mV/s] for each subject.	219
F.7	Boxplot of normalized feature MCV [mV/s] versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of MCV for each subject.	220
F.8	Boxplot of normalized feature MOV versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of MOV [mV/s] for each subject.	221
F.9	Boxplot of normalized feature MOV versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of MOV [mV/s] for each subject.	222
F.10	Boxplot of normalized feature A/MCV versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of A/MCV [s] for each subject.	223
F.11	Boxplot of normalized feature A/MCV versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of A/MCV [s] for each subject.	224
F.12	Boxplot of normalized feature A/MOV versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of A/MOV [s] for each subject.	225
F.13	Boxplot of normalized feature A/MOV versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of A/MOV [s] for each subject.	226
F.14	Boxplot of normalized feature ACV versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of ACV [mV/s] for each subject.	227

F.15	Boxplot of normalized feature ACV versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of ACV [mV/s] for each subject.	228
F.16	Boxplot of normalized feature AOV versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of AOV [mV/s] for each subject.	229
F.17	Boxplot of normalized feature AOV versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of AOV [mV/s] for each subject.	230
F.18	Boxplot of normalized feature F versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of F [1/min] for each subject.	231
F.19	Boxplot of normalized feature F versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of F [1/min] for each subject.	232
F.20	Boxplot of normalized feature T versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of T [s] for each subject.	233
F.21	Boxplot of normalized feature T versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of T [s] for each subject.	234
F.22	Boxplot of normalized feature T_c versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of T_c [ms] for each subject.	235
F.23	Boxplot of normalized feature T_c versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of T_c [ms] for each subject.	236
F.24	Boxplot of normalized feature T_o versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of T_o [ms] for each subject.	237
F.25	Boxplot of normalized feature T_o versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of T_o [ms] for each subject.	238
F.26	Boxplot of normalized feature $T_{cl,1}$ versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of $T_{cl,1}$ [ms] for each subject.	239
F.27	Boxplot of normalized feature $T_{cl,1}$ versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of $T_{cl,1}$ [ms] for each subject.	240
F.28	Boxplot of normalized feature $T_{cl,2}$ versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of $T_{cl,2}$ [ms] for each subject.	241
F.29	Boxplot of normalized feature $T_{cl,2}$ versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of $T_{cl,2}$ [ms] for each subject.	242
F.30	Boxplot of normalized feature T_{ro} versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of T_{ro} [ms] for each subject.	243
F.31	Boxplot of normalized feature T_{ro} versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of T_{ro} [ms] for each subject.	244
F.32	Boxplot of normalized feature $PERCLOS$ versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of $PERCLOS$ for each subject.	245
F.33	Boxplot of normalized feature $PERCLOS$ versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of $PERCLOS$ for each subject.	246
F.34	Boxplot of normalized feature T_{50} versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of T_{50} [ms] for each subject.	247
F.35	Boxplot of normalized feature T_{50} versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of T_{50} [ms] for each subject.	248
F.36	Boxplot of normalized feature T_{80} versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of T_{80} [ms] for each subject.	249
F.37	Boxplot of normalized feature T_{80} versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of T_{80} [ms] for each subject.	250
F.38	Boxplot of normalized feature T_{90} versus KSS for subjects S1 to S22 (except for subject S2). The values on the bottom left show the maximum of T_{90} [ms] for each subject.	251
F.39	Boxplot of normalized feature T_{90} versus KSS for subjects S23 to S43. The values on the bottom left show the maximum of T_{90} [ms] for each subject.	252
F.40	Boxplot of normalized features versus KSS values for subject S2. The maximum value of each feature is shown on the bottom left of plots.	253
F.41	Absolute values of the Spearman's rank correlation coefficient $ \rho_s $ calculated between KSS input-based features	254

F.42 Absolute values of the Spearman's rank correlation coefficient $ \rho_s $ calculated between drive time-based features	255
H.1 Examples of convex and non-convex sets	259
H.2 An example of a convex function	259

List of tables

2.1	Karolinska Sleepiness Scale (kss)	27
2.2	Literature review of the length of time intervals between successive KSS inputs	28
2.3	Stanford Sleepiness Scale (sss)	29
4.1	Means of moving standard deviations of $H(n)$ for all rounds (R) and parts (P) of track 1 and track 2, for all subjects (excluding subject S3)	42
4.2	Summary of experiments studied in this work	51
5.1	confusion matrix: events of video labeling versus those of the proposed detection methods	83
6.1	Values of ANOVA to assess the significant difference between means of blink rates for all tasks	95
6.2	Contingency table: saccade amplitude versus occurrence of gaze shift-induced blinks, for subject S1, first selection procedure	102
7.1	Literature review of feature aggregation and the calculated statistic measure	104
7.2	Extracted blink features	108
7.3	Literature review of the experiment setups. n. s.: not specified	120
7.4	Literature review of the features introduced in this work. Trends versus drowsiness are either pos.: positive or neg.: negative. n. s.: the feature was studied without its trend being specified. * reduced vigilance, ** before a driving error, *** based on another end point for blinks	121
7.5	Left table: number of occurrences of KSS values at the time of first unintended lane departure and number of occurrences for the maximum value of KSS, if no lane departure was detected. Right table: confusion matrix.	124
7.6	Results of paired-sample t -test (t_0) and Wilcoxon signed-rank test (z_0) for all features shown in Figure 7.16. Red color indicates non-significant features.	127
7.7	Results of paired-sample t -test (t_0) and Wilcoxon signed-rank test (z_0) shown in Figure 7.17. Red color indicates non-significant features.	129
7.8	Left table: number of occurrences of KSS values at the time of first microsleep and the number of occurrences for the maximum value of KSS, if no microsleep was detected. Right table: confusion matrix.	129
7.9	Results of paired-sample t -test (t_0) and Wilcoxon signed-rank test (z_0) for all features shown in Figure 7.18. Red color indicates non-significant features.	131
7.10	Sorted Spearman's rank correlation coefficient ρ_s and Pearson correlation coefficient ρ_p between all KSS input-based features and KSS values ($N = 391$). All p -values were smaller than 0.05 except for red features.	134
7.11	Sorted Spearman's rank correlation coefficient ρ_s and Pearson correlation coefficient ρ_p between all drive time-based features and KSS values ($N = 4021$). All p -values were smaller than 0.05.	135
8.1	Confusion matrix of a binary classifier	144
8.2	Confusion matrix of the binary subject-dependent ANN classifier ($N_h = 5$). Feature type: KSS input-based features	154
8.3	Confusion matrices of the subject-dependent ANN classifiers for 2-class ($N_h = 10$) and 3-class ($N_h = 20$) cases. Feature type: drive time-based features	155

8.4	Confusion matrices of the binary subject-dependent ANN classifier for KSS input-based features of the driving simulator experiment. Left: imbalanced features ($N_h = 2$). Right: balanced features by SMOTE ($N_h = 2$)	156
8.5	Confusion matrices of the binary subject-dependent ANN classifier for KSS input-based features of the real road experiment applied to the network trained based on the SMOTE. Left: $N_h = 3$. Right: $N_h = 10$	158
8.6	Confusion matrix of the binary subject-independent ANN classifier for drive time-based features ($N_h = 2$)	158
8.7	Confusion matrix of the binary subject-dependent SVM classifier. Feature type: KSS input-based features	169
8.8	Confusion matrices of the subject-dependent SVM classifiers for the 2-class and 3-class cases. Feature type: drive time-based features	170
8.9	Confusion matrices of the binary subject-dependent SVM classifiers for KSS input-based features of driving simulator experiment. Left: imbalanced features, right: balanced features by considering different misclassification costs	170
8.10	Confusion matrix of the binary subject-dependent SVM classifier for KSS input-based features of the real road experiment applied to the model trained by considering different misclassification costs	171
8.11	Confusion matrix of the binary subject-independent SVM classifier for drive time-based features	172
8.12	Confusion matrix of the binary subject-dependent k -NN classifier for $k = 5$. Feature type: KSS input-based features	174
8.13	Confusion matrices of the subject-dependent k -NN classifier ($k = 7$) for the 2-class and 3-class cases. Feature type: drive time-based features.	174
8.14	Confusion matrix of the binary subject-dependent k -NN classifier ($k = 7$). Feature type: imbalanced KSS input-based features of driving simulator experiment	175
8.15	Confusion matrix of the binary subject-independent k -NN classifier for drive time-based features ($k = 9$)	175
8.16	Best selected feature combination set by the SFFS and ANN classifier from 1 to 10 features. Feature type: drive time-based features	185
8.17	Confusion matrices of the binary subject-dependent ANN classifier ($N_h = 10$) for drive time-based features. Left: classification with 19 features. Right: classification with 4 features	186
8.18	Confusion matrices of the binary subject-dependent ANN classifier ($N_h = 10$) for drive time-based features of drop-outs. Left: classification with 19 features. Right: classification with 4 features	186
8.19	Values of $\Delta\gamma$ calculated based on the MIA method for drive time-based features ($\check{D} = 4$) .	187
8.20	Best selected drive time-based features based on the CFS method and Pearson correlation coefficient (Red features were also selected by the Spearman's rank correlation coefficient in Table 8.21.)	188
8.21	Best selected drive time-based features based on the CFS method and Spearman's rank correlation coefficient (Red features were selected by the Pearson correlation coefficient in Table 8.20.)	189
8.22	Confusion matrices of the binary subject-dependent ANN classifier ($N_h = 10$) for drive time-based features and $k_{CFS} = 4$. Left: Pearson correlation coefficient. Right: Spearman's rank correlation.	190
8.23	Best selected drive time-based features based on the CFS method using the Pearson correlation coefficient regardless of the number of features	190
8.24	Best selected drive time-based features based on the CFS method using the Spearman's rank correlation coefficient regardless of the number of features	190
D.1	Typical data set of one-way ANOVA	206
D.2	Contingency table with two categories	209

Bibliography

- Abdellaoui, A. (2013). Driver state classification based on eye movements by support vector machine. Master's thesis, University of Stuttgart.
- Abe, S. (2010). *Support Vector Machines for Pattern Classification*. Advances in Computer Vision and Pattern Recognition. Springer.
- Addison, P. S. (2010). *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*. Taylor & Francis.
- Akbani, R., Kwek, S., and Japkowicz, N. (2004). Applying support vector machines to imbalanced datasets. In Boulicaut, J.-F., Esposito, F., Giannotti, F., and Pedreschi, D., editors, *Machine Learning: ECML 2004*, volume 3201 of *Lecture Notes in Computer Science*, pages 39–50. Springer Berlin Heidelberg.
- Åkerstedt, T. and Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *The International journal of neuroscience*, 52(1-2):29–37.
- Åkerstedt, T., Peters, B., Anund, A., and Kecklund, G. (2005). Impaired alertness and performance driving home from the night shift: a driving simulator study. *Journal of sleep research*, 14(1):17–20.
- Anund, A. (2009). *Sleepiness at the wheel*. PhD thesis, Karolinska Institutet, Stockholm.
- Anund, A., Kecklund, G., Vadeby, A., Hjälmdahl, M., and Åkerstedt, T. (2008). The alerting effect of hitting a rumble strip - a simulator study with sleepy drivers. *Accident Analysis & Prevention*, 40(6):1970 – 1976.
- Applus+ IDIADA (2014). [Online; accessed 15-October-2014] <http://www.applusidiada.com/en/>.
- Arnedt, J. T., Geddes, M. A. C., and MacLean, A. W. (2005). Comparative sensitivity of a simulated driving task to self-report, physiological, and other performance measures during prolonged wakefulness. *Journal of Psychosomatic Research*, 58(1):61 – 71.
- Artusi, R., Verderio, P., and Marubini, E. (2002). Bravais-pearson and spearman correlation coefficients: meaning, test of hypothesis and confidence interval. *International Journal of Biological Markers*, 17(2):148–151.
- Asa, B. and Weston, J. (2010). A user's guide to support vector machines. In Carugo, O. and Eisenhaber, F., editors, *Data Mining Techniques for the Life Sciences*, volume 609 of *Methods in Molecular Biology*, pages 223–239. Humana Press.
- Authié, C. N. and Mestre, D. R. (2011). Optokinetic nystagmus is elicited by curvilinear optic flow during high speed curve driving. *Vision Research*, 51(16):1791 – 1800.
- Baranski, J. V. (2007). Fatigue, sleep loss, and confidence in judgment. *Journal of Experimental Psychology: Applied*, 13(4):182–196.
- Barea, R., Boquete, L., Ortega, S., López, E., and Rodríguez-Ascariz, J. M. (2012). EOG-based eye movements codification for human computer interaction. *Expert Systems with Applications*, 39(3):2677 – 2683.
- Barr, L., Howarth, H., Popkin, S., and Carroll, R. J. (2005). A review and evaluation of emerging driver fatigue detection measures and technologies. In *Proceedings of the International Conference on Fatigue Management in Transportation Operations*, Seattle, USA.
- Bartula, M., Tigges, T., and Muehlsteff, J. (2013). Camera-based system for contactless monitoring of respiration. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 2672–2675.

- Beideman, L. R. and Stern, J. A. (1977). Aspects of the eye blink during simulated driving as a function of alcohol. *Human Factors*, 19:73–77.
- Belz, S. M., Robinson, G. S., and Casali, J. G. (2004). Temporal separation and self-rating of alertness as indicators of driver fatigue in commercial motor vehicle operators. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1):154–169.
- Bergasa, L., Nuevo, J., Sotelo, M., Barea, R., and Lopez, M. E. (2006). Real-time system for monitoring driver vigilance. *Intelligent Transportation Systems, IEEE Transactions on*, 7(1):63–77.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer.
- Bouchner, P., Piekník, R., Novotny, S., Pekny, J., Hajny, M., and Borzová, C. (2006). Fatigue of car drivers-detection and classification based on the experiments on car simulators. In *6th WEEAS International Conference on Simulation, Modeling and Optimization*, pages 727–732, Lisbon, Portugal.
- Brain Products GmbH (2009). Selecting a suitable EEG recording cap - tutorial. [Online; accessed 14-August-2014] <http://www.brainproducts.com/downloads.php?kid=8>.
- Brown, T., Lee, J., Schwarz, C., Fiorentino, D., and McDonald, A. (2014). Assessing the feasibility of vehicle-based sensors to detect drowsy driving. Technical report, National Advanced Driving Simulator, The University of Iowa.
- Bulling, A., Ward, J., Gellersen, H., and Troster, G. (2011). Eye movement analysis for activity recognition using electrooculography. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(4):741–753.
- Burrus, C. S., Gopinath, R. A., and Guo, H. (1998). *Introduction to wavelets and wavelet transforms: a primer*. Prentice Hall.
- Caffier, P. P., Erdmann, U., and Ullsperger, P. (2003). Experimental evaluation of eye-blink parameters as a drowsiness measure. *European Journal of Applied Physiology*, 89(3-4):319–325.
- Chang, C. C. and Lin, C. J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357.
- Chiang, C. (2007). *Optimization of Communication Systems, Lecture 1B: Convex Sets and Convex Functions*. Electrical Engineering Department, Princeton University. [Online; accessed 07-July-2014] <https://www.princeton.edu/~chiangm/ele53911b.pdf>.
- Chua, E. C., Tan, W., Yeo, S., Lau, P., Lee, I., Mien, I. H., Puvanendran, K., and Gooley, J. (2012). Heart rate variability can be used to estimate sleepiness-related decrements in psychomotor vigilance during total sleep deprivation. *Sleep*, 35(3):325–334.
- Čolić, A., Marques, O., and Furht, B. (2014). *Driver Drowsiness Detection*. Springer.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.
- Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27.
- Crawshaw, J. and Chambers, J. (2001). *A Concise Course in Advanced Level Statistics: With Worked Examples*. Nelson Thornes.
- Cristianini, N. and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press.
- Daimler AG (2008). Hightech report 02. Technical report.
- Daimler AG (2014a). Attention Assist. [Online; accessed 24-August-2014] <http://www.daimler.com/dccom/0-5-1210218-1-1210332-1-0-0-1210228-0-0-135-0-0-0-0-0-0-0.html>.

- Daimler AG (2014b). DISTRONIC PLUS with Steering Assist. [Online; accessed 22-January-2015] [http://www.daimler.com/dccom/0-5-1210218-1-1210321-1-0-0-1210228-0-0-135-0-0-0-0-0-0.html](http://www.daimler.com/dccom/0-5-1210218-1-1210321-1-0-0-1210228-0-0-135-0-0-0-0-0-0-0.html).
- Damousis, I. G. and Tzovaras, D. (2008). Fuzzy fusion of eyelid activity indicators for hypovigilance-related accident prediction. *Intelligent Transportation Systems, IEEE Transactions on*, 9(3):491–500.
- DESTATIS (2013a). Unfallentwicklung auf Deutschen Straßen. [Online; accessed 20-November-2014] https://www.destatis.de/DE/Publikationen/Thematisch/TransportVerkehr/Verkehrsunfaelle/PK_Unfallentwicklung_PDF.pdf;jsessionid=6C80F60C49D0724AACAFAD373BC163C7.cae2?__blob=publicationFile.
- DESTATIS (2013b). Verkehrsunfälle-Zeitreihen 2012. [Online; accessed 23-March-2014] <https://www.destatis.de/DE/Publikationen/Thematisch/TransportVerkehr/Verkehrsunfaelle/VerkehrsunfaelleZeitreihen.html>.
- Dinges, D. F. and Grace, R. (1998). PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance. Technical Report Tech. Rep. FHWA-MCRT-98-006, Federal Highway Administration. Office of motor carriers.
- Dong, Y., Hu, Z., Uchimura, K., and Murayama, N. (2011). Driver inattention monitoring system for intelligent vehicles: A review. *Intelligent Transportation Systems, IEEE Transactions on*, 12(2):596–614.
- Duchowski, A. (2007). *Eye Tracking Methodology: Theory and Practice*. Springer.
- Duda, R. O., Hart, P. E., and Stork, D. G. (2012). *Pattern Classification*. Wiley.
- Dukas, R. (1998). *Cognitive Ecology: The Evolutionary Ecology of Information Processing and Decision Making*. Cognitive Ecology: The Evolutionary Ecology of Information Processing and Decision Making. University of Chicago Press.
- Dureman, E. I. and Bodéén, C. (1972). Fatigue in simulated car driving. *Ergonomics*, 15(3):299–308.
- Ebrahim, P. (2011). Drowsiness detection using lane data - event-based and driver model approaches. Master's thesis, University of Stuttgart.
- Ebrahim, P., Stolzmann, W., and Yang, B. (2013a). Eye movement detection for assessing driver drowsiness by electrooculography. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, pages 4142–4148.
- Ebrahim, P., Stolzmann, W., and Yang, B. (2013b). Road dependent driver eye movements under real driving conditions by electrooculography. In *Communications, Signal Processing, and their Applications (ICCSA), 2013 1st International Conference on*, pages 1–6.
- Ebrahim, P., Stolzmann, W., and Yang, B. (2013c). Spontaneous vs. gaze shift-induced blinks for assessing driver drowsiness/inattention by electrooculography. In *Driver Distraction and Inattention, 2013 3rd International Conference on*.
- Eoh, H. J., Chung, M. K., and Kim, S. (2005). Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *International Journal of Industrial Ergonomics*, 35(4):307 – 320.
- Ergoneers GmbH (2014). Eye-tracking glasses - Dikablis essential. [Online; accessed 30-September-2014] <http://www.ergoneers.com/wp-content/uploads/2014/09/Dikablis-Essential-Eye-Tracking-Glasses.pdf>.
- Eskandarian, A., Sayed, R., Delaigue, P., Blum, J., and Mortazavi, A. (2007). Advanced driver fatigue research. Technical report, Center for Intelligent Systems Research (CISR) , School of Engineering and Applied Science, The George Washington University.
- Evinger, C., Manning, K., Pellegrini, J., Basso, M., Powers, A., and Sibony, P. (1994). Not looking while leaping: the linkage of blinking and saccadic gaze shifts. *Experimental Brain Research*, 100(2):337–344.

- Fairclough, S. H. and Gildeade, K. (2014). *Advances in Physiological Computing*. Human–Computer Interaction Series. Springer London, Limited.
- Field, A. (2007). *Discovering Statistics Using SPSS*. Introducing Statistical Methods Series. SAGE Publications.
- Ford Motor Company (2010). Ford technology news brief. [Online; accessed 25-August-2014] <http://technology.fordmedia.eu/documents/newsletter/FordTechnologyNewsletter082010.pdf>.
- Friedrichs, F., Hermannstädter, P., and Yang, B. (2011). Consideration of influences on driver state classification. In *2nd International Conference on Driver Distraction and Inattention*.
- Friedrichs, F., Miksch, M., and Yang, B. (2010). Estimation of lane data-based features by odometric vehicle data for driver state monitoring. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, pages 611–616.
- Friedrichs, F. and Yang, B. (2010a). Camera-based drowsiness reference for driver state classification under real driving conditions. In *Intelligent Vehicles Symposium (IV), 2010 IEEE*, pages 101 –106.
- Friedrichs, F. and Yang, B. (2010b). Drowsiness monitoring by steering and lane data based features under real driving conditions. In *18th European Signal Processing Conference (EUPISCO)*.
- Fürsich, A. (2009). Klassifikation des fahrerzustandes mit hidden markov modellen und bayes-netzen. Master's thesis, University of Stuttgart.
- Gault, T. R. and Farag, A. A. (2013). A fully automatic method to extract the heart rate from thermal video. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, pages 336–341.
- Gershon, P., Shinar, D., Oron-Gilad, T., Parmet, Y., and Ronen, A. (2011). Usage and perceived effectiveness of fatigue countermeasures for professional and nonprofessional drivers. *Accident Analysis & Prevention*, 43(3):797 – 803.
- Gosling, J. (1995). *Introductory Statistics*. Quicksmart University Guides Series. Pascal Press.
- Hall, M. A. (1999). *Correlation-based Feature Selection for Machine Learning*. PhD thesis, University of Waikato, Department of computer science.
- Hall, M. A. (2000). Correlation-based feature selection for discrete and numeric class machine learning. In *ICML*, pages 359–366.
- Hallvig, D., Anund, A., Fors, C., Kecklund, G., Karlsson, J. G., Wahde, M., and Åkerstedt, T. (2013). Sleepy driving on the real road and in the simulator-a comparison. *Accident Analysis & Prevention*, 50:44 – 50.
- Hammoud, R. I. (2008). *Passive Eye Monitoring: Algorithms, Applications and Experiments*, chapter 14, page 315. Signals and Communication Technology. Springer London, Limited.
- Hargutt, V. (2003). *Das Lidschlussverhalten als Indikator für Aufmerksamkeits- und Müdigkeitsprozesse bei Arbeitshandlungen*. Fortschritt-Berichte VDI.: Biotechnik, Medizintechnik. VDI-Verlag.
- He, H. and Garcia, E. A. (2009). Learning from imbalanced data. *Knowledge and Data Engineering, IEEE Transactions on*, 21(9):1263–1284.
- He, H., Pang, C., and Li, Q. (2010). Driver fatigue monitoring method based on multi-information fusion. In *Image and Signal Processing (CISP), 2010 3rd International Congress on*, volume 1, pages 110–113.
- Hermannstädter, P. and Yang, B. (2013). Driver distraction assessment using driver modeling. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, pages 3693–3698.
- Hirshkowitz, M. (2013). Fatigue, sleepiness, and safety: Definitions, assessment, methodology. *Sleep Medicine Clinics*, 8(2):183 – 189.
- Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., and Dement, W. C. (1973). Quantification of sleepiness: A new approach. *Psychophysiology*, 10(4):431–436.

- Hoel, J., Jaffard, M., and Van Elslande, P. (2010). Attentional competition between tasks and its implications. In *European Conference on Human Centred Design for Intelligent Transport Systems*.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and van de Weijer, J. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. OUP Oxford.
- Horak, K. (2011). Fatigue features based on eye tracking for driver inattention system. In *Telecommunications and Signal Processing (TSP), 2011 34th International Conference on*, pages 593–597.
- Horne, J. A. and Reyner, L. A. (1996). Counteracting driver sleepiness: effects of napping, caffeine, and placebo. *Psychophysiology*, 33(3):306–309.
- Hsu, C. W., Chang, C. C., and Lin, C. J. (2003). A practical guide to support vector classification. [Online; accessed 10-April-2015] <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
- Hu, S. and Zheng, G. (2009). Driver drowsiness detection with eyelid related parameters by support vector machine. *Expert Systems with Applications*, 36(4):7651 – 7658.
- Huang, R., Chang, S., Hsiao, Y., Shih, T., Lee, S., Ting, H., and Lai, C. (2012). Strong correlation of sleep onset between EOG and EEG sleep stage 1 and 2. In *Computer, Consumer and Control (IS3C), 2012 International Symposium on*, pages 614–617.
- Ingre, M., Åkerstedt, T., Peters, B., Anund, A., and Kecklund, G. (2006). Subjective sleepiness, simulated driving performance and blink duration: examining individual differences. *Journal of Sleep Research*, 15(1):47–53.
- ISO 15007 (2013). Road vehicles - Measurement of driver visual behavior with respect to transport information and control systems - Part 1: Definitions and parameters. ISO 15007-1:2013.
- Jain, A. K., Mao, J., and Mohiuddin, K. M. (1996). Artificial neural networks: a tutorial. *Computer*, 29(3):31–44.
- James, W. (1981). *The Principles of Psychology (Vol. I)*. Cambridge, MA: Harvard University Press. (see: James, W., The Principles of Psychology, H. Holt and Co., New York, 1890.).
- Jammes, B., Sharaby, H., and Esteve, D. (2008). Automatic EOG analysis: A first step toward automatic drowsiness scoring during wake-sleep transitions. *Somnologie - Schlaforschung und Schlafmedizin*, 12(3):227–232.
- Johns, M., Tucker, A., Chapman, R., Crowley, K., and 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, M. W. (1991). A new method for measuring daytime sleepiness: the epworth sleepiness scale. *Sleep*, 14(6):540–545.
- Johns, M. W. (2003). The amplitude-velocity ratio of blinks: a new method for monitoring drowsiness. *Sleep*, 26:A51.
- Johns, M. W. and Tucker, A. J. (2005). The amplitude-velocity ratios of eyelid movements during blinks: changes with drowsiness. *Sleep*, 28:A122.
- Jürgensohn, T., Neculau, M., and Willumeit, H. P. (1991). Visual scanning pattern in curve negotiation. *Vision in Vehicles III*, pages 171–178.
- Kadambe, S., Murray, R., and Boudreax-Bartels, G. F. (1999). Wavelet transform-based QRS complex detector. *Biomedical Engineering, IEEE Transactions on*, 46(7):838–848.
- Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., and Fukasawa, K. (2006). Validation of the karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology*, 117(7):1574 – 1581.
- Kandil, F. I., Rotter, A., and Lappe, M. (2010). Car drivers attend to different gaze targets when negotiating closed vs. open bends. *Journal of Vision*, 10(4):24.1–11.
- Kecklund, G. and Åkerstedt, T. (1993). Sleepiness in long distance truck driving: an ambulatory EEG study of night driving. *Ergonomics*, 36(9):1007–1017.

- Keller, W. (2004). *Wavelets in Geodesy and Geodynamics*. Walter de Gruyter.
- Kincses, W. E., Hahn, S., Schrauf, M., and Schmidt, E. A. (2008). Measuring driver's mental workload using EEG. *ATZ worldwide*, 110(3):12–17.
- Kira, K. and Rendell, L. A. (1992a). The feature selection problem: Traditional methods and a new algorithm. In *Proceedings of AAAI*.
- Kira, K. and Rendell, L. A. (1992b). A practical approach to feature selection. In *Proceedings of the ninth international workshop on Machine learning*, pages 249–256. Morgan Kaufmann Publishers Inc.
- Kircher, A., Uddman, M., and Sandin, J. (2002). Vehicle control and drowsiness. Technical report, Swedish National Road and Transport Research Institute.
- Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., and Ramsey, D. J. (2006). The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. Technical Report DOT HS 810 594, Virginia Tech Transportation Institute, 3500 Transportation Research Plaza (0536) Blacksburg, Virginia 24061.
- Knipling, R. R. and Wierwille, W. W. (1994). Vehicle-based drowsy driver detection: Current status and future prospects. In *Moving Toward Deployment. Proceedings of the IVHS America Annual Meeting. 2 Volumes*, number Volume 1.
- Kolo, B. (2011). *Binary and Multiclass Classification*. Weatherford Press.
- Kranjec, J., Beguš, S., Geršak, G., and Drnovšek, J. (2014). Non-contact heart rate and heart rate variability measurements: A review. *Biomedical Signal Processing and Control*, 13(0):102 – 112.
- Krupiński, R. and Mazurek, P. (2010). Convergence improving in evolution-based technique for estimation and separation of electrooculography and blinking signals. In Piętka, E. and Kawa, J., editors, *Information Technologies in Biomedicine*, volume 69 of *Advances in Intelligent and Soft Computing*, pages 293–302. Springer Berlin Heidelberg.
- Kumar, D. and Poole, E. (2002). Classification of EOG for human computer interface. In *Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint*, volume 1, pages 64 – 67.
- Lal., S. K. L. (2001). *The psychophysiology of driver fatigue/drowsiness : electroencephalography, electro-oculogram, electrocardiogram and psychological effects*. PhD thesis, University of Technology, Sydney. Faculty of Science.
- Land, M. F. and Lee, D. N. (1994). Where we look when we steer. *Nature*, 369:742 – 744.
- Leigh, R. J. and Zee, D. S. (1999). *The Neurology of Eye Movements*, chapter 4. Contemporary Neurology Series. Oxford University Press.
- Lemieux, C. (2009). *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer Series in Statistics. Springer.
- Li, H. D., Liang, Y. Z., Xu, Q. S., Cao, D. S., B., T. B., Deng, B. C., and Lin, C. C. (2011a). Recipe for uncovering predictive genes using support vector machines based on model population analysis. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, 8(6):1633–1641.
- Li, L., Xie, M., and Dong, H. (2011b). A method of driving fatigue detection based on eye location. In *Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on*, pages 480–484.
- Liang, Y. and Lee, J. D. (2010). Combining cognitive and visual distraction: Less than the sum of its parts. *Accident Analysis & Prevention*, 42(3):881 – 890.
- Liang, Y., Reyes, M. L., and Lee, J. D. (2007). Real-time detection of driver cognitive distraction using support vector machines. *Intelligent Transportation Systems, IEEE Transactions on*, 8(2):340–350.
- Liu, C. C., Hosking, S. G., and Lenné, M. G. (2009). Predicting driver drowsiness using vehicle measures: Recent insights and future challenges. *Journal of Safety Research*, 40(4):239–245.

- Lugger, M. (2011). *Mehrstufige Klassifikation paralinguistischer Eigenschaften aus Sprachsignalen mit Hilfe neuartiger Merkmale*. PhD thesis, University of Stuttgart.
- Magosso, E., Provini, F., Montagna, P., and Ursino, M. (2006). A wavelet based method for automatic detection of slow eye movements: A pilot study. *Medical Engineering & Physics*, 28(9):860 – 875.
- Mallat, S. (2009). *A wavelet tour of signal processing: the sparse way*. Academic Press Elsevier, Amsterdam; Heidelberg [u.a.], 3. ed. edition.
- Martínez, M., Soria, E., Magdalena, R., Serrano, A., Martín, J., and Vila, J. (2008). Comparative study of several fir median hybrid filters for blink noise removal in electrooculograms. *WSEAS Trans. Sig. Proc.*, 4(3):53–59.
- May, J. F. (2011). Chapter 21 - driver fatigue. In Porter, B. E., editor, *Handbook of Traffic Psychology*, pages 287 – 297. Academic Press, San Diego.
- McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.
- Moller, H. J., Kayumov, L., Bulmash, E. L., Nhan, J., and Shapiro, C. M. (2006). Simulator performance, microsleep episodes, and subjective sleepiness: normative data using convergent methodologies to assess driver drowsiness. *Journal of Psychosomatic Research*, 61(3):335 – 342.
- Moller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4):525 – 533.
- Monk, T. H. (1989). A visual analogue scale technique to measure global vigor and affect. *Psychiatry Research*, 27(1):89–99.
- Montgomery, D. C. and Runger, G. C. (2006). *Applied statistics and probability for engineers*. Wiley.
- Morris, T. L. and Miller, J. C. (1996). Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology*, 42(3):343 – 360.
- NHTSA (1998). Drowsy driving and automobile crashes. Technical report, National Highway Traffic Safety Administration.
- Niedermeyer, E. and da Silva, F. L. D. (2005). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. LWW Doody's all reviewed collection. Lippincott Williams & Wilkins.
- Niemann, H. (2003). *Klassifikation von Mustern*.
- Nunez, P. L. and Srinivasan, R. (2006). *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press.
- O'Hanlon, J. F. (1972). Heart rate variability: a new index of driver alertness/fatigue. Technical report, SAE Technical Paper.
- O'Hanlon, J. F. and Kelley, G. R. (1977). Comparison of performance and physiological changes between drivers who perform well and poorly during prolonged vehicular operation. In Mackie, R., editor, *Vigilance*, volume 3 of *NATO Conference Series*, pages 87–109. Springer US.
- Olson, D. L. and Delen, D. (2008). *Advanced Data Mining Techniques*. Springer.
- Otmani, S., Pebayle, T., Roge, J., and Muzet, A. (2005). Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers. *Physiology & Behavior*, 84(5):715 – 724.
- Oxford (2014). Oxford dictionaries. [Online; accessed 11-August-2014] <http://www.oxforddictionaries.com/definition/english/distraction>.
- Papadelis, C., Chen, Z., Kourtidou-Papadeli, C., Bamidis, P. D., Chouvarda, I., Bekiaris, E., and Maglaveras, N. (2007). Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents. *Clinical Neurophysiology*, 118(9):1906 – 1922.
- Pettitt, M., Burnett, G., and Stevens, A. (2005). Defining driver distraction. In *Proceedings of the 12th ITS World Congress*, San Francisco, USA, ITS America.

- Philip, P., Sagaspe, P., Taillard, J., Valtat, C., Moore, N., Åkerstedt, T., Charles, A., and Bloulac, B. (2005). Fatigue, sleepiness, and performance in simulated versus real driving conditions. *Sleep*, 28(12):1511–1516.
- Picot, A., Caplier, A., and Charbonnier, S. (2009). Comparison between EOG and high frame rate camera for drowsiness detection. In *Applications of Computer Vision (WACV), 2009 Workshop on*, pages 1–6.
- Picot, A., Charbonnier, S., and Caplier, A. (2010). Drowsiness detection based on visual signs: blinking analysis based on high frame rate video. In *Instrumentation and Measurement Technology Conference (I2MTC), 2010 IEEE*, pages 801–804.
- Pilutti, T. and Ulsoy, A. G. (1999). Identification of driver state for lane-keeping tasks. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 29(5):486–502.
- Pimenta, P. A. D. M. (2011). Driver drowsiness classification based on lane and steering behavior. Master’s thesis, University of Stuttgart.
- Platho, C., Pietrek, A., and Kolrep, H. (2013). Erfassung der Fahrermüdigkeit. *Berichte der Bundesanstalt für Straßenwesen. Unterreihe Fahrzeugtechnik*, F 89.
- Pouliarikas, A. D. (2009). *Transforms and Applications Handbook, Third Edition*. Electrical Engineering Handbook. Taylor & Francis.
- Priddy, K. L. and Keller, P. E. (2005). *Artificial Neural Networks: An Introduction*. Tutorial Text Series. SPIE Press.
- Pudil, P., Ferri, F. J., Novovičoá, J., and Kittler, J. (1994). Floating search methods for feature selection with nonmonotonic criterion functions. In *Pattern Recognition, 1994. Vol. 2-Conference B: Computer Vision & Image Processing., Proceedings of the 12th IAPR International. Conference on*, volume 2, pages 279–283. IEEE.
- Rantanen, E. M. and Goldberg, J. H. (1999). The effect of mental workload on the visual field size and shape. *Ergonomics*, 42(6):816–834. PMID: 10340026.
- Records, R. E. (1979). *Physiology of the human eye and visual system*. Harper & Row.
- Reddy, M. S., Narasimha, B., Suresh, E., and Rao, K. (2010). Analysis of EOG signals using wavelet transform for detecting eye blinks. In *Wireless Communications and Signal Processing (WCSP), 2010 International Conference on*, pages 1–4.
- Regan, M. A., Hallett, C., and Gordon, C. P. (2011). Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident Analysis & Prevention*, 43(5):1771 – 1781.
- Riemersma, J. B. J., Sanders, A. F., Wildervanck, C., and Gaillard, A. W. (1977). Performance decrement during prolonged night driving. In Mackie, R., editor, *Vigilance*, volume 3 of *NATO Conference Series*, pages 41–58. Springer US.
- Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14(5):465 – 471.
- Rosario, H. D., Solaz, J., Rodriguez, N., and Bergasa, L. M. (2010). Controlled inducement and measurement of drowsiness in a driving simulator. *IET Intelligent Transport Systems*, 4 (4):280–288.
- Rowland, L. M., Thomas, M. L., Thorne, D. R., Sing, H. C., Krichmar, J. L., Davis, H. Q., Balwinski, S. M., Peters, R. D., Kloepfel-Wagner, E., Redmond, D. P., Alicandri, E., and Belenky, G. (2005). Oculomotor responses during partial and total sleep deprivation. *Aviation, space, and environmental medicine*, 76(7):C104–C113.
- Royal, D. (2003). Volume I: Findings. national survey of distracted and drowsy driving. attitudes and behavior: 2002 (vol. 1. findings). Technical Report DOT HS 809 566, U.S. Department of Transportation National Highway Traffic Safety Administration (NHTSA), Washington.
- Santillán-Guzmán, A. (2014). *Digital Enhancement of EEG/MEG signals*. PhD thesis, University of Kiel.
- Saroj, K. L. L. and Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biological Psychology*, 55(3):173 – 194.

- Sağlam, M., Lehnert, N., and Glasauer, S. (2011). Optimal control of natural eye-head movements minimizes the impact of noise. *The Journal of Neuroscience*, 31(45):16185–16193.
- Savitzky, A. and Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedure. *Anal. Chem.* (1964), pages 1627–1639.
- Schleicher, R., Galley, N., Briest, S., and Galley, L. (2008). Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics*, 51(7):982–1010.
- Schmidt, E. A., Schrauf, M., Simon, M., Buchner, A., and Kincses, W. E. (2011). The short-term effect of verbally assessing drivers' state on vigilance indices during monotonous daytime driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 14(3):251 – 260.
- Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., and Kincses, W. E. (2009). Drivers' misjudgment of vigilance state during prolonged monotonous daytime driving. *Accident Analysis & Prevention*, 41(5):1087 – 1093.
- Schmieder, F. (2009). Support vector machine in emotion in recognition. Master's thesis, University of Stuttgart.
- Seekircher, J., Woltermann, B., Gern, A., Janssen, R., Mehren, D., and Lallinger, M. (2009). Das Auto lernt sehen - kamerabasierte Assistenzsysteme. *ATZextra*, 14(1):64–71.
- Shahid, A., Wilkinson, K., Marcu, S., and Shapiro, C. M. (2012a). Epworth sleepiness scale (ESS). In Shahid, A., Wilkinson, K., Marcu, S., and Shapiro, C. M., editors, *STOP, THAT and One Hundred Other Sleep Scales*, pages 149–151. Springer New York.
- Shahid, A., Wilkinson, K., Marcu, S., and Shapiro, C. M. (2012b). Karolinska sleepiness scale (KSS). In Shahid, A., Wilkinson, K., Marcu, S., and Shapiro, C. M., editors, *STOP, THAT and One Hundred Other Sleep Scales*, pages 209–210. Springer New York.
- Shahid, A., Wilkinson, K., Marcu, S., and Shapiro, C. M. (2012c). Stanford sleepiness scale (SSS). In Shahid, A., Wilkinson, K., Marcu, S., and Shapiro, C. M., editors, *STOP, THAT and One Hundred Other Sleep Scales*, pages 369–370. Springer New York.
- Sigari, M. H. (2009). Driver hypo-vigilance detection based on eyelid behavior. In *Advances in Pattern Recognition, 2009. ICAPR '09. Seventh International Conference on*, pages 426–429.
- Simon, M. (2013). *Neurophysiologische Analyse des kognitiven Fahrerzustandes*. PhD thesis, Eberhard Karls University of Tübingen.
- Simon, M., Schmidt, E. A., Kincses, W. E., Fritzsche, M., Bruns, A., Aufmuth, C., Bogdan, M., R., W., and Schrauf, M. (2011). EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions. *Clinical Neurophysiology*, 122(6):1168 – 1178.
- Sirevaag, E. J. and Stern, J. A. (2000). Ocular measures of fatigue and cognitive factors. *Engineering Psychophysiology: Issues and Applications.*, R. Backs & W. Boucsein (eds.), L. Erlbaum Associates Press, New Jersey.
- Skipper, J. H. and Wierwille, W. W. (1986). Drowsy driver detection using discriminant analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 28(5):527–540.
- Soman, K. P., Ramachandran, K. I., and Resmi, N. G. (2010). *Insight Into Wavelets : from Theory to Practice*. PHI Learning.
- Sommer, D. and Golz, M. (2010). Evaluation of perclos based current fatigue monitoring technologies. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 4456–4459.
- Sonneleitner, A., Simon, M., Kincses, W. E., Buchner, A., and Schrauf, M. (2012). Alpha spindles as neurophysiological correlates indicating attentional shift in a simulated driving task. *International Journal of Psychophysiology*, 83(1):110 – 118.
- Sonneleitner, A., Simon, M., Kincses, W. E., and Schrauf, M. (2011). Assessing driver state - neurophysiological correlates of attentional shift during real road. In *2nd International Conference on Driver Distraction and Inattention*.

- Stern, J. A., Boyer, D., and Schroeder, D. (1994). Blink rate: a possible measure of fatigue. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 36:285–297.
- Stern, J. A., Walrath, L. C., and Goldstein, R. (1984). The endogenous eyeblink. *Psychophysiology*, 21(1):22–33.
- Stern, R. M., Ray, W. J., and Quigley, K. S. (2001). *Psychophysiological Recording*. Oxford University Press.
- Straube, A. and Büttner, U. (2007). *Neuro-ophthalmology: Neuronal Control of Eye Movements*. Developments in ophthalmology. Karger.
- Summala, H., Häkkänen, H., Mikkola, T., and Sinkkonen, J. (1999). Task effects on fatigue symptoms in overnight driving. *Ergonomics*, 42(6):798–806. PMID: 10340025.
- Suzuki, M., Yamamoto, N., Yamamoto, O., Nakano, T., and Yamamoto, S. (2006). Measurement of driver's consciousness by image processing -a method for presuming driver's drowsiness by eye-blanks coping with individual differences -. In *Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on*, volume 4, pages 2891–2896.
- Svensson, U. (2004). Blink behavior based drowsiness detection method development and validation. Master's thesis, Linköping University.
- Tefft, B. C. (2014). Prevalence of motor vehicle crashes involving drowsy drivers, United States, 2009 – 2013 (November 2014). Technical report, AAA Foundation for Traffic Safety.
- Thiffault, P. and Bergeron, J. (2003). Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis & Prevention*, 35(3):381–391.
- Thorslund, B. (2003). Electrooculogram analysis and development of a system for defining stages of drowsiness. Master's thesis, Linköping University.
- Tinati, M. A. and Mozaffary, B. (2006). A wavelet packets approach to electrocardiograph baseline drift cancellation. *International Journal of Biomedical Imaging*, pages 1–9.
- Tomek, I. (1976). Two modifications of CNN. *Systems, Man and Cybernetics, IEEE Transactions on*, SMC-6(11):769–772.
- Tran, Y., Wijesuriya, N., Tarvainen, M., Karjalainen, P., and Craig, A. (2009). The relationship between spectral changes in heart rate variability and fatigue. *Journal of Psychophysiology*, 23(3):143–151.
- Uhlich, S. (2006). Emotion recognition of speech signals. Master's thesis, University of Stuttgart.
- Veropoulos, K., Campbell, C., and Cristianini, N. (1999). Controlling the sensitivity of support vector machines. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 55–60.
- Verwey, W. B. and Zaidel, D. M. (2000). Predicting drowsiness accidents from personal attributes, eye blinks and ongoing driving behaviour. *Personality and Individual Differences*, 28(1):123–142.
- ViaMichelin (2014). Michelin. [Online; accessed 26-August-2014] <http://www.viamichelin.de/web/Karten-Stadtplan>.
- Volkswagen AG (2014). Fatigue Detection. [Online; accessed 20-November-2014] http://www.volkswagen.com.au/en/technology_and_service/technical-glossary/fatigue-detection.html.
- Volvo Group (2014). Driver Alert control. [Online; accessed 24-August-2014] <http://www.volvcars.com/de/sales-services/service/specialsales/Pages/techniklexikon-d.aspx>.
- von Helmholtz, H. (1925). *Treatise on Physiological Optics. Volume III: The Perceptions of Vision*. The Optical Society of America.
- Wallén Warner, H., Ljung Aust, M., Sandin, J., Johansson, E., and Björklund, G. (2008). Manual for DREAM 3.0, driving reliability and error analysis method. Technical report, Deliverable D5.6 of the EU FP6 project SafetyNet, TREN-04-FP6TRSI2.395465/506723.

- Wei, Z. and Lu, B. (2012). Online vigilance analysis based on electrooculography. In *Neural Networks (IJCNN), The 2012 International Joint Conference on*, pages 1 –7.
- Wigh, F. (2007). Detection of driver unawareness based on long- and short-term analysis of driver lane keeping. Master's thesis, Linköpings University.
- Williamson, A., Friswell, R., Olivier, J., and Grzebieta, R. (2014). Are drivers aware of sleepiness and increasing crash risk while driving? *Accident Analysis & Prevention*, 70(0):225 – 234.
- Yang, B. (2014). Detection and pattern recognition. Slides to the lecture.
- Young, L. and Sheena, D. (1975). Survey of eye movement recording methods. *Behavior Research Methods*, 7:397–429.
- Young, R. K. (1993). *Wavelet Theory and Its Applications*. Kluwer international series in engineering and computer science: VLSI, computer architecture, and digital signal processing. Springer US.
- Zamora, M. E. (2001). *The study of the sleep and vigilance Electroencephalogram using neural network methods*. PhD thesis, University of Oxford.
- Zeeb, E. (2010). Daimler's new full-scale, high-dynamic driving simulator – a technical overview. In *Conference Proc. Driving Simulator Conference Europe, Paris*.
- Zulley, J. and Popp, R. (2012). Müdigkeit im Straßenverkehr. [Online; accessed 04-April-2013] http://www.adac.de/_mmm/pdf/vm_muedigkeit_im_strassenverkehr_flyer_48789.pdf.