

**STIMMBASIERTE AKUSTISCHE SCHLÄFRIGKEITSDETEKTION
(ACOUSTIC SLEEPINESS ANALYSIS)**

-

**EINE SIGNALVERARBEITUNGSBASIERTE UND
MUSTERERKENNUNGSALGORITHMENGESTÜTZTE
VERFAHRESENTWICKLUNG**

Inaugural-Dissertation

Zur Erlangung der Doktorwürde

des Fachbereichs Bildungswissenschaften

der Bergischen Universität Wuppertal

Vorgelegt von:

Jarek Krajewski

2007

Erstgutachter: Prof. Dr. Rainer Wieland

Zweitgutachter: Prof. Dr. Bernd Kröger

Die Dissertation kann wie folgt zitiert werden:

urn:nbn:de:hbz:468-20080379

[<http://nbn-resolving.de/urn/resolver.pl?urn=urn%3Anbn%3Ade%3A468-20080379>]

VORWORT

Eine solche Arbeit kann nur mit Unterstützung vieler anderer Menschen zustande kommen. Deswegen möchte ich an dieser Stelle die Gelegenheit ergreifen, allen Menschen zu danken, die durch tatkräftige Unterstützung zum Gelingen dieser Arbeit beigetragen haben. Allen voran ist meinem Doktorvater Herrn Prof. Rainer Wieland zu danken, der, obwohl mit dem üblichen Zeitmangel von Professoren geschlagen, stets ein offenes Ohr hatte und meine Arbeit bis zum Ende mit einer angenehmen Mischung aus Kreativität und konstruktiver Kritik begleitete.

Weiterhin möchte ich mich bei Prof. Bernd Kröger, Prof. Martin Golz, Prof. Jonathan Harrington und Prof. Torbjorn Akerstedt für die wichtigen Impulse bedanken, die manch einen meiner Gedanken zu ordnen halfen. Des Weiteren möchte ich mich bei Frau PD Dr. Barbara Wilhelm und Dr. Alexander Gundel für die konzeptuelle Unterstützung sowie das Vertrauen bedanken, das sie in mich gesetzt haben. Natürlich möchte ich meinen Hilfskräften und Diplomanden danken, die vor allem mit ihrem großen Engagement bei den vielen langen Schlafdeprivation-Studiennächten maßgeblich zum Gelingen dieser Arbeit beitrugen. Allen voran sind Herr Thomas Hörstermann, Herr Jonas Schmaling, Frau Sarah Groppe und Frau Julia Fink zu nennen.

Die vorliegende Arbeit und mein Weg dorthin wären nicht ohne die Unterstützung weiterer lieber Menschen möglich gewesen. Für den langjährigen emotionalen Halt und ihr Verständnis in Zeiten, in denen ich ihr nicht immer die verdiente Aufmerksamkeit zuteil kommen lassen konnte, möchte ich Wiebke Tofaute von ganzem Herzen danken. Außerdem danke ich meinen Freunden Andy, Wolfgang, Patty und Arnt, die sich nicht nur als Lektoren und Advocati Diabolorum meiner Arbeit verdient gemacht haben, sondern mir immer wieder halfen, den nötigen Abstand zur Arbeit zu wahren. Zu guter Letzt möchte ich den Menschen danken, denen diese Arbeit gewidmet ist, meiner Großmutter Maria und meinen Eltern. Jeder Versuch, all die Gründe hierfür aufzuzählen, muss fehlschlagen! Danke.

Wuppertal, im Dezember 2007

Jarek Krajewski

Für meine Eltern und meine Freundin Wiebke

INHALTSVERZEICHNIS

TEIL I: THEORETISCHE GRUNDLAGEN DES AKUSTISCHEN MESSANSATZES

1. AUTOMATISCHE DETEKTION VON SCHLÄFRIGKEITZUSTÄNDEN 1
 - 1.1 Zielsetzung und Aufbau der Arbeit
 - 1.2 Unfallrisiko Schläfrigkeit
 - 1.3 Alternative Anwendungsfelder der Schläfrigkeitsmessung
 - 1.4 Messansätze Schläfrigkeit
 - 1.5 Mehrwert des akustischen Messansatzes

2. LINGUISTISCHE, PHONETISCHE UND AKUSTISCHE GRUNDLAGEN. 24
 - 2.1 Phonetische Grundlagen der Sprachproduktion
 - 2.2 Kognitive Sprachproduktionstheorien
 - 2.3 Auditiv-perzeptive Beschreibungskategorien
 - 2.4 Akustische Beschreibungsgrößen

3. SCHLÄFRIGKEITSBEDINGTE STIMMÄNDERUNGEN 60
 - 3.1 Kognitiv-phonetisches Mediatormodell
 - 3.2 Empirische Befunde: schläfrigkeitsinduzierte Stimmänderungen

4. AKUSTISCHER MESSPROZESS ALS MACHINE-LEARNING PARADIGMA . . . 78
 - 4.1 Sprachaufnahme
 - 4.2 Vorverarbeitung
 - 4.3 Merkmalsberechnung
 - 4.4 Dimensionalitätsreduktion
 - 4.5 Klassifikation

TEIL II: EMPIRISCHE VALIDIERUNG DES AKUSTISCHEN MESSANSATZES

1. METHODE	116
1.1 Stichprobe	
1.2 Untersuchungsprozedere	
1.3 Instrumente	
1.4 Acoustic Sleepiness Analysis	
2. ERGEBNISSE: SPRACHSEGMENT: VOKAL	134
2.1 Ergebnisse: Einzelkennzahlen	
2.2 Ergebnisse: Machine-Learning Klassifikation	
3. ERGEBNISSE: SPRACHSEGMENT WORT	157
3.1 Ergebnisse Einzelkennzahlen	
3.2 Ergebnisse: Machine-Learning Klassifikation	
4. ERGEBNISSE: SPRACHSEGMENT PHRASALE EINHEIT	166
4.1 Ergebnisse: Einzelkennzahlen	
4.2 Ergebnisse: Machine-Learning Klassifikation	
5. ERGEBNISSE: SPRACHSEGMENT MULTIPLE PHRASALE EINHEIT	186
5.1 Ergebnisse: Einzelkennzahlen	
5.2 Ergebnisse: Machine-Learning Klassifikation	
6. DISKUSSION	196
6.1 Zusammenfassung der Ergebnisse	
6.2 Fehleranalyse, Datenqualität und Validitätsgefährdung	
6.3 Forschungsdesiderate	
6.4 Anwendungsbezogenes Transferpotenzial	
7. ZUSAMMENFASSUNG	219
8. LITERATURVERZEICHNIS	229
9. ABKÜRZUNGSVERZEICHNIS	260
ANHANG	260

ABSTRACT

Ziel der vorliegenden Arbeit ist die konzeptionelle Begründung und experimentelle Validierung eines Verfahrens zur akustischen Schläfrigkeitsmessung. Potenzielle Anwendungsfelder der Echtzeit-Detektion von Schläfrigkeitszuständen liegen in der unfallpräventiven Fahrerzustanderkennung und adaptiven Optimierung des Arbeits-Erholungs-Zyklus. Die Analyse von Schläfrigkeitszuständen erfordert in diesen Kontexten minimal-invasive, belastungsfreie und den Tätigkeitsvollzug nicht beeinträchtigende Messzugänge. Ein vielversprechender neuer Messansatz für die Echtzeit-Detektion von Schläfrigkeit ist die hier vorgestellte akustische Stimmanalyse. Ihr zentrales Messprinzip stützt sich auf der Vorstellung, dass schläfrigkeitsinduzierte Veränderungen des Sprachproduktionsprozesses zu messbaren Effekten in den Stimmbereichen Prosodie, Artikulation und Stimmqualität führen. Dieser Beitrag gibt einen Überblick zu linguistisch-phonetischen Grundlagen, zentralen auditiv-perzeptiven und akustischen Beschreibungskategorien, schläfrigkeitsbedingten Stimmveränderungen, dem Machine-Learning-basierten Messprozedere sowie empirischen Validierungsbefunden der akustischen Schläfrigkeitsanalyse.

Zur Validierung der akustischen Schläfrigkeitsmessung wurden in einer Schlafdeprivationsstudie (20.00 - 04.00; N=24) alle 40 Minuten sowohl Sprachaufnahmen als auch selbst- und fremdberichtete Schläfrigkeitsratings (Karolinska Sleepiness Scale, KSS) erhoben. Das verwendete Sprachmaterial bestand aus aufsteigend komplexen linguistischen Einheiten (extrahierten Vokalen, Worten und ganzen phrasalen Einheiten). Für alle Spracheinheiten wurden 338 akustische Stimmerkmale aus den Bereichen Prosodie, Artikulation und Stimmqualität berechnet. Die Auswahl der Kennzahlen folgte zum einen theoriegeleitet (aus dem hier formulierten Kognitiv-phonetischen Mediatormodell) und zum anderen orientiert an sprachunabhängigen Signalverarbeitungs-Standards.

Der komplette Messprozess der hier vorgestellten Acoustic Sleepiness Analysis (ASA) wiederum folgt dem Ablauf der Mustererkennung und des Machine Learnings der künstlichen Intelligenzforschung (Aufnahme, Vorverarbeitung, Merkmalsberechnung, Dimensionalitätsreduktion, Klassifikation). Zu den eingesetzten Dimensionalitätsreduktionsverfahren gehören die filterbasierte Vorselektion (Korrelationsfilter), die Dimensionsreduktion mittels Principle Component, die wrapperbasierte Hauptselektion mit Genetic Algorithms und einem K-Nearest-Neighbour Verfahren. Zu den erfolgreich eingesetzten Klassifikationsverfahren zäh-

len Künstliche Neuronale Netze, Support Vector Machines und K-Nearest-Neighbour Verfahren. Die Validitätskorrelationen von dichotomisierten vorhergesagten Schläfrigkeitsszuständen und den selbstberichteten Schläfrigkeitssratings liegen für das erfolgreichste K-Nearest-Neighbour Verfahren über alle Sprachmaterialien gemittelt bei durchschnittlich $r = .55$. Der Klassifikationsgenauigkeit der Schläfrigkeitssvorhersage befindet sich bei durchschnittlich 82.0% (Sensitivität = 63.4%; Spezifität = 87.8%). Trotz einiger mit dem Within-Subject Validierungsdesign und dem verwendeten Selbstreport-Validierungsanker verknüpften Limitationen, kann von einer gelungenen ersten Validierung des hier entwickelten stimmbasierten akustischen Schläfrigkeitssdetektionsinstruments gesprochen werden.

Schlüsselwörter: Schläfrigkeitssmessung, Akustik, Signalverarbeitung, Machine Learning, Mustererkennung

1 AUTOMATISCHE DETEKTION VON SCHLÄFRIGKEITZUSTÄNDEN

1.1 Zielsetzung und Aufbau der Arbeit

Ziel der vorliegenden Arbeit ist die experimentelle Validierung eines Verfahrens zur akustischen Schläfrigkeitsmessung. Die hier beschriebene *Acoustic Sleepiness Analysis* (ASA) basiert auf der Integration von Konzepten der linguistisch-phonetischen, phoniatischen, schlafmedizinischen, arbeitspsychologischen, nachrichtentechnischen und neuroinformatischen Forschung. Das zentrale Messprinzip stützt sich auf der Vorstellung, dass schläfrigkeitsinduzierte Veränderungen des Sprachproduktionsprozesses zu messbaren Effekten in den Stimmbereichen Prosodie, Artikulation und Stimmqualität führen. Die schläfrigkeitsbedingt veränderte Sprache wird über 338 Kennwerte erfasst und über einen aus der Künstlichen Intelligenz Forschung entlehnten Machine-Learning Prozess zur Vorhersage der Schläfrigkeit genutzt.

Der zentrale Neuigkeitswert der Arbeit liegt in der Entwicklung eines akustischen Messalgorithmus zur Erfassung von Schläfrigkeit, sowie seiner erfolgreichen experimentellen Validierung. Tabelle 1 gibt einen Überblick zu Haupt- und Nebenbeiträgen der vorliegenden Arbeit.

Tabelle 1: Haupt- und Nebenbeiträge der vorliegenden Arbeit für die psychologische Forschung

Beiträge der vorliegenden Arbeit	Bedeutung
<i>Hauptbeitrag</i>	
Entwicklung eines akustischen Messalgorithmus zur Erfassung von Schläfrigkeit	Aufgreifen einer wissenschaftlich- und anwendungspragmatisch-relevanten Problemlage; Entwicklung einer die bestehenden Ansätze erweiternden, Mehrwert liefernden Lösung
Erfolgreiche experimentelle Validierung des akustischen Messinstruments	Nachweis der Funktionsfähigkeit und Einsatzmöglichkeit der akustischen Schläfrigkeitsdetektion innerhalb einer Laborsituation
<i>Nebenbeitrag</i>	
Einführung des Kognitiv-phonetischen Mediatormodells schläfrigkeitsinduzierter Stimmänderungen	Theoretische Fundierung des Wirkzusammenhangs von Schläfrigkeitsinduzierten physiologischen Veränderungen auf den Stimmproduktionsprozess und auditiv-perzeptive Stimmmerkmale
Programmierung einer Kennzahlenbatterie für die Erfassung diverser interner Zustände	Grundlage für die Entwicklung diverser stimmbasierter Messinstrumente zur Erfassung von psychologisch relevanten State- und Trait-Konstrukten
Einführen von Mustererkennungsansätzen für die psychologische Modellbildung	Optimierte Prädiktion diverser psychologischer Vorhersageprobleme mittels nicht-linearer Machine-Learning Verfahren

Kapitel 1 gliedert sich in fünf Teile. Zunächst werden die unfallbezogene Bedeutung von Schläfrigkeit für den Schienen-, Luft- und Straßenverkehr skizziert (Abschnitt 1.2) sowie weitere Anwendungsfelder der Schläfrigkeitsmessung angedeutet (Abschnitt 1.3). Anschließend folgt die begriffliche Abgrenzung des Konstrukts Schläfrigkeit und ein kritischer Überblick zu bislang existierenden Schläfrigkeits-Messansätzen (Abschnitt 1.4). In Abgrenzung zu bisherigen Schläfrigkeitsmessinstrumenten zeigt Abschnitt 1.5 den möglichen Mehrwert des akustischen Messansatzes auf. Im Anschluss führt die Arbeit in linguistisch-phonetische Grundlagen der Sprachproduktion (Abschnitt 2.1), in kognitive und phonetische Sprachproduktionstheorien (Abschnitt 2.2), phonetische Beschreibungskategorien (Abschnitt 2.3) sowie in akustische Stimmmerkmale ein (Abschnitt 2.4).

Integriert werden diese Perspektiven im *Kognitiv-phonetischen Mediatormodell* schläfrigkeitsinduzierter Stimmveränderungen (Abschnitt 3.1). In diesem Modell werden die Wirkpfade ausgehend von Schläfrigkeit über physiologische und kognitive Komponenten der Sprachproduktion zu auditiv-perzeptiven und akustischen Parametern beschrieben. Neben diesen Modellvorstellungen unterstreichen auch die empirischen Ergebnisse der akustischen Forschung zu alkohol- und schläfrigkeitsinduzierten Sprachveränderungen unsere Forschungshypothese (Abschnitt 3.2).

Das folgende Kapitel thematisiert das auf die akustische Analyse angewendete Machine-Learning Prozedere der Künstlichen Intelligenz Forschung (Kapitel 4). In der Konstruktions- und Lernphase werden die Machine-Learning Phasen Sprachaufnahme, Vorverarbeitung, Merkmalsberechnung, Dimensionalitätsreduktion und Merkmalsklassifikation durchlaufen. Die anschließende Testphase vergleicht mittels einer 10-fachen Kreuzvalidierung die aus den Stimmkennwerten vorhergesagten Schläfrigkeitswerte mit den tatsächlichen.

Im Teil 2 der vorliegenden Arbeit wird die zur Validierung der Acoustic Sleepiness Analysis (ASA) herangezogene Schlafdeprivationsstudie beschrieben. Die resultierenden Ergebnisse werden getrennt auf Einzelmerkmalsniveau sowie auf Klassifikationsebene deskriptiv und inferenzstatistisch für verschiedene Sprachmaterialien beschrieben (Kapitel 2 - Kapitel 5) und diskutiert (Kapitel 6). In diesem Kapitel wird neben einer Zusammenfassung der gefundenen Ergebnisse auch die Qualität der vorliegenden Daten bewertet. Dazu werden mögliche Fehlerquellen untersucht und die interne und externe Validität bewertet (Abschnitt 6.2). Es folgen ein Vergleich der Ergebnisse mit den Vorhersagen des Kognitiv-phonetischen Mediatormodells sowie ein Vergleich der Befunde mit bisherigen schläfrigkeitsbezogenen Forschungsergebnissen. Im abschließenden Forschungsausblick werden die diskutierten Einschränkungen aufgegriffen, Perspektiven zur Weiterentwicklung der Acoustic Sleepiness Analysis (ASA) aufgezeigt (Abschnitt 6.3) sowie das anwendungsbezogene Transferpotenzial der akustischen Stimmessung umrissen. Die Arbeit schließt inhaltlich mit Kapitel 7, das eine resümierende Zusammenfassung der gesamten Arbeit gibt.

1.2 Unfallrisiko Schläfrigkeit

“Sleepiness is the largest identifiable and preventable cause of accidents in transport operations (between 15 and 20% of all accidents), surpassing that of alcohol or drug related incidents in all modes of transportation” (Akerstedt, 2000).

Der erste Super-GAU in der Geschichte 1986 im sowjetischen Atomkraftwerk von Tschernobyl, der Zwischenfall im Reaktor von Three-mile Island, die offene Bugklappe der britischen Kanalfähre „Harald of Free Enterprise“ 1987 vor dem belgischen Hafen Zeebrugge sowie das Tankerunglück 1989 des Öltankers „ExxonValdez“ vor der Küste Alaskas werden auf Schläfrigkeitszustände der betreffenden Operatoren zurückgeführt (vgl. z.B. Mitler et al., 1988). Das erhöhte schläfrigkeitsbedingte Unfallrisiko von Überwachungs- und Steuerungstätigkeiten wird über eine Reihe von Einbußen der kognitiven Leistungsfähigkeit vermittelt. Betroffen sind dabei die Funktionen des Arbeits- und Kurzzeitgedächtnisses, der selektiven Aufmerksamkeit, der Daueraufmerksamkeit (Kotterba et al., 1998), der verminderten Vigilanz (Barbé et al., 1998), der überwachenden Steuerung (vgl. Nilsson, Soderstrom, Karlsson & Lekander, 2005), der motorischen Kontrolle, der Situational Awareness (vgl. Harwood, Barnett & Wickens, 1988), der räumlichen Orientierung, non-verbale kognitive Verarbeitungsgeschwindigkeit (Cassel et al., 1989), der mathematischen Verarbeitung, des Problemlösens (Horne, 1988; Linde & Bergström, 1992), der Entscheidungsfindung, des divergentes Denken und der verbalen Kreativität (vgl. Harrison & Horne, 1998; Durmer & Dinges, 2005; Rogers, Dorrian & Dinges, 2005; Wesensten, Belenky, Thorne & Kautz, 2004). Neben diesen schläfrigkeitsbedingten kognitiven Beeinträchtigungen sind es vor allem Mikroschlafepisoden, die als ein unmittelbarer kritischer Faktor in der Unfallentstehung identifiziert worden sind. Mikroschlafepisoden sind charakterisiert durch den unwillkürlichen Verlust der Aufmerksamkeit und Reaktionsfähigkeit. Vorläufer dieser einige Sekunden andauernden Episoden sind bei sehr schläfrigen Personen Ereignisse wie langsame zyklische Augenbewegungen (oder völlig eingestellte Kaskadenbewegungen), verlängerte Lidschlüsse und wandernde haluzinative Gedanken (Thorpy & Yager, 1991).

Betroffen von schläfrigkeitsbedingten kognitiven Beeinträchtigungen und Mikroschlafepisodes sind vor allem der Schienen-, Luft-, Raum-, Schiffs- und Straßenverkehr. So werden im Schienenverkehr bei 25% der Zugführer Mikroschlafereignisse während der Fahrt detektiert (Torsvall & Akerstedt, 1997). Entsprechend gehen konservative Schätzungen davon aus, dass durch Mikroschlafepisodes 1% bis 10% aller tödlichen Unfälle verursacht werden (Dinges, 1995; Horne & Reyner, 1995; Lisper et al., 1986). Auch im Luftverkehr kommt es aufgrund langer Time-on-Task Zeiten sowie gestörter zirkadianer Rhythmen zu gefährlichen Müdigkeitszuständen (Wright & McGown, 2001). Entsprechend sind 21% der im Nasa Aviation Safety Reporting System berichteten Vorfälle mit Schläfrigkeitszuständen verknüpft (Roske-Hofstrand, 1995).

Im Straßenverkehr zeichnet sich ein vergleichbares Bild ab. So liegt die Unfallhäufigkeit von krankheitsbedingt schläfrigen Personen (obstruktive Schlafapnoe mit chronisch nicht-erholsamem Schlaf) im Vergleich zur Allgemeinbevölkerung ca. 2.0 bis 6.3-mal so hoch (Findley et al., 1988; George et al., 1987; Lloberes et al., 2000; Teran-Santos et al., 1999, 2000; Wu & Yan-Go, 1996; Young et al., 1997). Auch in Fahr Simulatoruntersuchungen sind die Ergebnisse von Schlafapnoe-Patienten deutlich schlechter als die von Kontrollpersonen (George et al., 1996; Juniper et al., 2000; Risser et al., 2000). Darüber hinaus ist die Unfallrate von LKW-Berufskraftfahrer, die an einer atmungsbezogenen Schlafstörung leiden, im Vergleich zu LKW-Fahrern ohne diese Erkrankungen zweifach erhöht (Stoohs et al., 1994). Weitergehende zielgruppenspezifische Zusammenhänge zwischen Schläfrigkeit und Unfallrisiken liefern Hinweise auf die Bedeutung von Schläfrigkeit im Unfallgeschehen (vgl. Tabelle 2) (Crum & Morrow, 2002; Harrison, 2006; Hartley & Arnold, 2001; Lam, 2003; Ma, Williamson, & Friswell, 2003; Morrow & Crum, 2004; National Transport Commission, 2005; Smith, Carrington, & Trinder, 2005; Swann, 2002; Tse, Flin & Mearns, 2006; Tzamalouka, Papadakaki, & Chliaoutakis, 2005).

Tabelle 2: Unfallarten und ihre auf Schläfrigkeit zurückzuführenden Ursachenanteile

Unfallcharakteristik	Schläfrigkeits- bedingter Anteil	Referenz
Alle Unfälle	1 bis 40%	Fell, 1994; Garbarino et al. 2001, Gardner, 1998; Knippling & Wang, 1995; Laube et al., 1998; Lyznicki et al., 1998; MacLean et al. 2003; Sagberg, 1999
Alle Unfälle, männliche Fahrer	9 bis 10%	Maycock, 1997
Unfälle mit Personenschaden	7 bis 30%	Sagberg, 1999; UK Department of Transport, 2002
Alle tödlichen Unfälle	3 bis 15%	Akerstedt, 2001; Fell, 1994; Pack et a., 1995, Summala & Mikkola, 1994
Tödliche Unfälle auf Landstraßen	30%	Fell, 1994
Unfälle auf Autobahnen	16 bis 20%	Connor, Norton, Ameratunga et al., 2002; Horne & Reyner, 1995; Lindberg et al., 2001; Masa et al., 2000; Melamed et al., 2002; Reyner & Horne, 2002; Zulley et al., 1995
Tödliche Unfälle auf Autobahnen	25%	Zulley et al., 1995
Frontalzusammenstöße	6%	Amundsen & Christensen, 1986
Fahrbahnverlassende Unfälle	8%	Sagberg, 1999
Unfälle mit LkW Beteiligung	2 bis 41%	McCartt, Hammer & Fuller, 1998; Arnold et al., 1997; Williamson et al., 2001; European Transport Safety Council, 2001
Tödliche Unfälle mit LkW Beteiligung	4 bis 31%	Haworth, Heffeman & Horne, 1989; US National Transportation Safety Board, 1990; Summala & Mikkola, 1994

Auch außerhalb der krankheitsbedingten Schläfrigkeit werden vor allem schwere Verkehrsunfälle auf Autobahnen durch Einschlafen am Steuer verursacht (Bun, Slavova, Struttman, & Browning, 2005; Connor, Norton, Ameratunga et al., 2002; Dobbie, 2002; Gander, Marshall, James, & Le Quesne, 2006; Lindberg et al., 2001; Masa et al., 2000; Melamed et al., 2002; Queensland Travelsafe Committee, 2005; Sagberg, 1999; Williamson, 2005; Williamson, Feyer, Mattick, Friswell, & Finlay-Brown, 2001; Zulley et al., 1995). So konnten Philip et al. (2001) zeigen, dass es bei Urlaubsfahrten, nach verkürztem Nachtschlaf und längerer Fahrzeit, in den frühen Morgenstunden häufig zu schläfrigkeitsbedingten Unfällen kommt. Die Wahrscheinlichkeit von Verkehrsunfällen mit Todesfolge um 4:00 Uhr nachts ist sogar

11-mal so hoch wie am Morgen (Åkerstedt et al., 2001). In einer retrospektiven Analyse über einen Zeitraum von 6 Jahren wurde der Anteil von schläfrigkeitbedingten Unfällen auf 16 % geschätzt (MacLean et al., 2003). Eine prospektive Studie brachte 23 % aller Unfälle mit Schläfrigkeit am Steuer in Verbindung, wobei beinahe 25 % dieser Verkehrsunfälle zu schwerwiegenden Verletzungen des Fahrers führten (Horne & Reyner, 1995). In einer britischen Studie an 4621 befragten männlichen Lkw- und Pkw-Fahrer lag die Jahresprävalenz von beinahe Mikroschlafepisoden bei 29 % (Maycock, 1997). Garbarino et al. (2001) schätzen, dass bis zu 21.9 % der Unfälle mit einem Einschlafen am Steuer oder erhöhte Schläfrigkeit verknüpft sind (Garbarino et al., 2001). Zulley et al. (1995) schätzen, dass ein Viertel aller Verkehrstoten auf bayerischen Autobahnen bei Unfällen starben, die durch das Einschlafen des Fahrers ausgelöst wurden. Schätzungen von nicht-wissenschaftlichen Organisationen in Deutschland gehen von deutlich höheren Risiken zwischen 20% bis 40% aus (Deutscher Verband der Versicherungswirtschaft, Statistisches Bundesamt, HUK-Verband). Dementsprechend wird der durch schläfrigkeitbedingte Unfälle im Straßenverkehr resultierende volkswirtschaftliche Schaden von der Deutschen Gesellschaft für Schlafforschung und Schlafmedizin auf 10 Mrd. Euro jährlich geschätzt. Nicht berücksichtigt sind hierbei Zwischenfälle und Unfälle, die sich in anderen mit monotonen Überwachungs- und Steuerungsaufgaben verknüpften sicherheitssensitiven Tätigkeitsfeldern, wie z.B. in Kraftwerken, chemischen Fabriken, Krankenhäusern ereignen (Melamed, 2002). Gerade die in diesen Monitoringaufgaben benötigte Vigilanz, d.h. die Fähigkeit seltene und zufällig auftretende geringe spezifische Veränderungen in der Umwelt zu entdecken und auf sie zu reagieren (Mackworth, 1948), ist es, die schläfrigkeitbedingt große Einbußen verzeichnet. Zusammenfassend lässt sich daher festhalten, dass die Entwicklung von Messinstrumenten zur Erfassung von sicherheitskritischen Schläfrigkeitzuständen sowohl aus individueller Gesundheitsschutz als auch übergeordneter volkswirtschaftlicher Perspektive eine lohnenswerte Herausforderung darstellt.

1.3 Schläfrigkeitsmessung: Anwendungsfelder

Die Hauptanwendungssituation der Schläfrigkeitsmessung stellt die Bedienerzustandserkennung und Unfallprävention dar. So besitzt die automatische Echtzeit-Erfassung von Schläfrigkeitszuständen in den Anwendungsfeldern (a) Fahrerzustandserkennung und Unfallprävention, (b) ressourcen-adaptive Anforderungsgestaltung im Arbeits-Erholungs-Zyklus, (c) eigungsdiagnostische Anwendungen, (d) Erfolgsbewertung medizinisch-psychologischer Schlafstörungstherapien und (e) grundlagenwissenschaftliche Untersuchungen des Konzeptes Schläfrigkeit ein vielversprechendes Einsatzpotenzial.

(a) Fahrerzustandserkennung und Unfallprävention

Ein wichtiges Einsatzszenario stellt die rechtzeitige Erfassung von sicherheitskritischen Fahrerzuständen, wie Schläfrigkeit dar. Neben einer entsprechenden Zustandsrückmeldung könnten Fahrerassistenzsysteme die zustandsabhängige Modifikation der Fahrzeuginnenwelt einleiten (z.B. Einspielen befindlichkeitsangepasster Musikauswahl, aktivierende Frischluft oder angenehme Geruchsstoffe). Zusätzlich liefern an die Schläfrigkeit des Fahrzeugführers angepasste Sprachausgaben einen sicherheitsrelevanten Beitrag (Nass et al., 2005). Diese unfallpräventiven Überlegungen gelten primär für den Straßenverkehr. Vergleichbare Lösungen sind aber auch für den Wasser-, Schienen- und Luftverkehr sowie andere Überwachungs- und Steuerungsarbeitsplätze denkbar (z.B. Kraftwerk-Operateure, Dispatcherzentralen oder Fluglotsen).

(b) Ressourcen-adaptive Anforderungsgestaltung im Arbeits-Erholungs-Zyklus

Das aktuelle Anforderungs-Kapazitäts-Verhältnis bestimmt im wesentlichen Umfang das Ausmaß der Beanspruchung und das Niveau der Aufgabenerfüllung (siehe Mehrkomponenten Modell für psychische Belastung/Beanspruchung; Wieland-Eckelmann, 1992). Die Anpassung der mentalen, emotionalen und motivationalen Anforderungen an die momentan verfügbaren energetischen (und deshalb schläfrigkeitsabhängigen) Ressourcen beinhaltet z.B. eine Steigerung der Anforderungen in energetisch-wachen Phasen und eine Reduktion in schläfrig-ermüdeten Arbeitsphasen (Wieland-Eckelmann & Baggen, 1994). Diese Reduktion kann in eine komplette Arbeitsunterbrechung und anforderungsminimierte Pausenempfehlung münden (Krajewski, 2005; Krajewski & Wieland, 2003, 2004a, 2004b, 2004c, 2006). Die so resultierende Vermeidung von Beanspruchungsspitzen durch die rechtzeitige schläfrigkeitsspezifische Dosierung von Belastungen und Anforderungen könnte einen sowohl ge-

sundheitsförderlichen als auch leistungsoptimierenden Effekt besitzen. Diese ressourcenadaptive Anforderungsgestaltung könnte auch im Anwendungsfeld der Mensch-Maschine-Interaktion (Human-Computer-Interaction) von Bedeutung sein. Hier stellt sich die Aufgabe der Akzeptanzsicherung von maschineller Kommunikation. Die Anpassung der Computer-Reaktion an den energetischen Benutzerzustand kann die kommunikative Natürlichkeit und somit den Komfort erhöhen. Denkbare Anwendungen könnten sich z.B. für E-learning und Edutainment-Programme ergeben, die über die Detektion von Schläfrigkeit ihre Lehrgeschwindigkeit anpassen.

(c) Diagnostische Anwendungen

Innerhalb von eignungsdiagnostischen Beratungs- oder Auswahl-situationen von z.B. Flugpiloten könnten effiziente, vorurteilsfreie automatische Messungen von Müdigkeitsreaktionen auf beispielsweise verkürzten Nachtschlaf stattfinden. Diese Schläfrigkeitsbeurteilung kann im Sinne einer Belastbarkeitsmessung für die Auswahl von Personen in Schichtarbeitsplätzen sowie in monotonen Überwachungs- und Steuerungstätigkeiten hilfreich sein.

(d) Erfolgswertung medizinisch-psychologischer Schlafstörungstherapien

Verfahren zur Behandlung von Schlafstörungen und nicht-erholsamen Nachtschlaf (wie z.B. durch Schlafapnoe) sollten ihre Wirksamkeit in dem klinischen Endpunkt der Tagesschläfrigkeitsreduktion nachweisen. Zu diesem Zweck könnte die Tagesschläfrigkeit der Patienten zu verschiedenen Tageszeiten erfasst werden. Aber auch innerhalb der Diagnose, Interventionsunterstützung und Therapiebewertung von Krankheitsbildern, die mit energetischen Tiefständen assoziiert sind (z.B. Major Depression), könnte eine genaue Erfassung der Schläfrigkeit sinnvoll sein.

(e) Grundlagenwissenschaftliche Schläfrigkeitsuntersuchungen

Die Untersuchung des Phänomens Schläfrigkeit, seiner Antezedenzen sowie seiner Auswirkungen bedarf sicherer, reliabler, valider und verfälschungsresistenter Messinstrumente. Besondere Anforderungen an die Messansätze stellen hierbei psychologische Feldstudien, sowie Längsschnittstudien, die sich jeweils durch einen belastungsfreien und unaufwendigen Messzugang auszeichnen müssten.

1.4 Messansätze Schläfrigkeit

Begriffsbestimmung

Schläfrigkeit befindet sich in einem Konzeptfeld mit „hypersomnia“, „sleep propensity“, „somnolence“, „weariness“, „tiredness“, „fatigue“, „ability to stay awake“ und „drowsiness“ sowie den deutschen Begriffen Hypovigilanz, Drowsiness, Erregungsniveau, Wachheit, Highway Hypnosis, Mattigkeit, Erschöpfung, Ermüdung (Fatigue), Müdigkeit und Einschlafneigung. (vgl. Akerstedt, 2000; Cluydts et al., 2002; von den Berg, 2006). Unter Vigilanz versteht man einen Zustand der Bereitschaft auf Veränderungen, die in zufälligen Zeitabständen in der Umwelt auftreten, zu reagieren (Mackworth, 1948). Nach Tassi, Bonneford, Hoeft, Eschenlauer und Muzetand (2003) sind Arousal und Vigilanz abgegrenzte Entitäten. Vigilanz ist mit aufmerksamkeitsbezogener Verfügbarkeit verbunden. Arousal hingegen basiert vor allem auf neuronaler Aktivität. Arousal beeinflusst die Informationsverarbeitungsgeschwindigkeit, Vigilanz hingegen eher die Genauigkeit (wie z.B. die Anzahl der Auslassungsfehler bei Aufgabenbearbeitung). So wird Leistung in kognitiven Aufgaben in dieser spezifischen Weise beeinflusst. Hypovigilanz und Vigilanzminderung sind Ergebnisse von Ermüdung und durch veränderte äußere Umweltbedingungen aufhebbare Zustände. Des Weiteren wird Hypovigilanz in der DIN EN ISO-Norm 10075 definiert als Zustand einer herabgesetzten Entdeckungsleistung bei Überwachungsaufgaben mit geringer Variationsbreite. Die angebotenen Definitionen zu Hypovigilanz beinhalten einen energetischen und einen kognitiven Ansatz.

Wachheit stellt den energetischen Gesamtzustand dar und lässt sich als eine Funktion des Erregungsniveaus, also dem Grad der Energiemobilisation beschreiben. Veränderung der Wachheit können auf energetische Kosten einer Handlung, psychotrope Substanzen, Krankheit oder Schläfrigkeit zurückgeführt werden (Vollrath & Krüger, 2002). Drowsiness wird häufig synonym zu Schläfrigkeit gebraucht, versteht aber als Ursache dieses Zustandes Erschöpfung, fehlenden Schlaf und psychoaktive Substanzen. Darüber hinaus ist eine wichtige Unterscheidung zwischen Schläfrigkeit und dem angrenzenden Konzept der Ermüdung (Fatigue) zu treffen. Diese Ermüdung tritt in Folge körperlicher oder kognitiver Anstrengung auf. Fatigue beschreibt ein sehr unangenehmes Gefühl der Zerschlagenheit, Schwäche oder Erschöpfung, das sich negativ auf Motivation, körperliche Aktivität und Leistungsfähigkeit auswirkt (Högl & Poewe, 2001; Richter & Hacker, 1998). Ermüdung oder Erschöpfung können durch Ruhen behoben werden, ohne dabei notwendigerweise schlafen zu müssen. Im

Gegensatz dazu setzt Schläfrigkeit (auch Synonym Müdigkeit) keine körperliche oder kognitive Anstrengung voraus und lässt sich nur durch Schlaf aufheben. Shen et al. (2006) unterscheidet verschiedene Arten von Fatigue: akute vs. chronische Erschöpfung, physiologische vs. psychologische Erschöpfung (Physical and Mental Fatigue; Gaillard, 2001) und zentrale vs. periphere Erschöpfung. Erfasst werden kann Ermüdung u.a. mit der Swedish Occupational Fatigue Inventory oder der Fatigue Severity Scale (Krupp et al., 1989). Weitergehende Differenzierungen zum Konstrukt Ermüdung finden sich siehe bei Richter und Hacker (1998) sowie Shen et al. (2006).

Akerstedt (1998) beschreibt in seiner Charakterisierung Schläfrigkeit als einen Versuch des zentralen Nervensystems auf den Schlafmodus umzuschalten. Schläfrigkeit reflektiert in dieser Vorstellung die Bemühung, sich dem Einschlafen zu widersetzen. Galley und Schleicher (2002) unterscheiden drei Schläfrigkeitskomponenten: Deaktivierung, abnehmende Aufmerksamkeit und „Fighting against Sleepiness“. Weitergehende vertiefte Betrachtungen zur Natur des Schläfrigkeitskonzeptes finden sich in zahlreichen neueren Arbeiten (Desmond & Hancock, 2001; Fairclough, 2001; Falleti, Maruff, Collie, Darby, & McStephen, 2003; Falou et al., 2003; Maruff, Falleti, Collie, Darby, & McStephen, 2005; Matthews, 2002; Matthews & Desmond, 2002; Mills, Spruill, Kanne, Parkman, & Zhang, 2001; Oron-Gilad & Hancock, 2005; Peters, 2005; Philip et al., 2005; Rosa, 2001; Thiffault, & Bergeron, 2003; Van der Hulst, Meijman, & Rothengatter, 2001). Die in diesem Beitrag verwendete Definition des Konstruktes „Schläfrigkeit“ orientiert sich an einer operationalen Definition (Johns, 1998), die Schläfrigkeit über folgende Korrelate charakterisiert: (a) Tendenz Einzuschlafen, (b) subjektiv erlebtes Schlafbedürfnis, (c) Gefühl der schnelleren Ermüdbarkeit und reduzierten Leistungsfähigkeit bei körperlichen und kognitiven Aufgaben, (d) Eigenwahrnehmung von körperlichen Müdigkeitssymptomen (z. B. Schwere der Augenlider, Gähnen, Frösteln, Bewegungsdrang herabgesetzte Kommunikationsbereitschaft und Abnahme des Leistungswillens etc.). Eine hypothetische integrative Perspektive auf das Zusammenspiel einzelner oben genannter Konzepte bietet die unten dargestellte an Sagberg et al. (2004) angelehnte Abbildung 1.

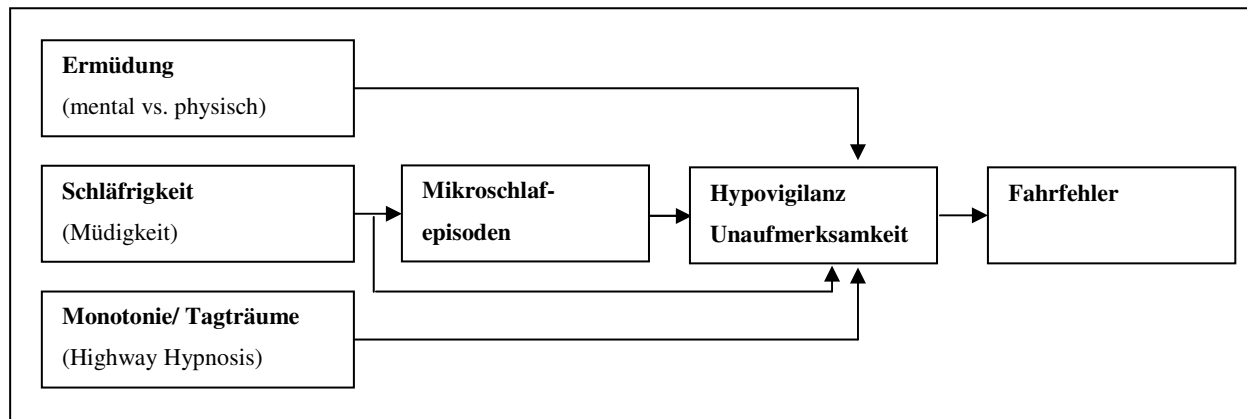


Abbildung 1: Einordnung von Schläfrigkeit und anderen Hypovigilanz-induzierenden Zuständen

Instrumente zur Schläfrigkeitsmessung

Verschiedene konzeptuelle Ansätze zur Schläfrigkeit und unterschiedliche Annahmen über zugrunde liegende Prozesse führten zu einer Vielzahl von Operationalisierungen. Ein Vergleich der Sensitivität, Reliabilität und Nützlichkeit verschiedener Verfahren zur Erfassung von Schläfrigkeit findet sich bei Balkin et al. (2004) und Cluydts et al. (2002). Die zahlreichen Erhebungs- und Messansätze, wie (a) Selbsteinschätzungsverfahren, (b) kognitive Leistungstestverfahren, (c) behaviorale Verfahren, (d) Blick- und Pupillen-bezogene Verfahren, (e) elektrophysiologische Verfahren, (f) EEG-basierte Verfahren sowie (g) Einschlaf latenz-basierte Verfahren sollen im folgenden Überblick vorgestellt werden.

Selbsteinschätzungsverfahren. Zur Erfassung von Schläfrigkeit stehen seit langem Selbstberichtsverfahren zur Verfügung. Dieser Messansatz ist der einfachste und kostengünstigste, um Schläfrigkeit, insbesondere erhöhte Tagesschläfrigkeit, zu messen. Nachteile dieser Methode sind, dass sie wie alle Selbsteinschätzungsmethoden anfällig sind gegenüber kognitiven und motivationalen Verzerrungen, wie der mangelnden Bewusstheit des eigenen Schläfrigkeitsstatus, oder Selbst- und Fremdtäuschungen. Darüber hinaus ist die Korrelation mit objektiven Verfahren häufig gering, was auch darauf zurückgeführt werden kann, dass schlaf- rigkeitsbedingte Leistungseinbußen über- oder unterschätzt werden (Cook et al., 1988). Zu den Verfahren, die den aktuellen Schläfrigkeitsgrad messen, gehört die klassische Stanford Sleepiness Scale (SSS). Dieses 7-stufige Ratingverfahren mit Ankerlabels auf jedem Antwortlevel findet seit vielen Jahren eine breite Anwendung (Hoddes et al., 1973). Die Sensitivität und Validität der Skala wurde in einer Vielzahl von Studien dokumentiert (Williamson et al., 1994; Wylie et al., 1996). So konnten deutliche Zusammenhänge von SSS Werten mit

Aufgabenperformanz (Glenville & Broughton, 1979; Hoddes et al., 1973), Vigilanz und Reaktionszeiten ($r = .69$) (Glenville & Broughton, 1979) beobachtet werden. Hoddes (1973) fand darüber hinaus gute Validitätskorrelationen zu kognitiven Leistungstests wie dem Auditory Vigilance Task ($r = .70$) und dem Wilkinson Additions Task ($r = .67$).

Die derzeit in der schlafmedizinischen Forschung dominierende Skala ist die 9- bzw. 10-stufige Karolinska Sleepiness Scale (KSS; Åkerstedt & Gillberg, 1990; Marsalek et al., 2006). Die Stufen 8 („sleepy, somewhat strenuous to stay awake“) und 9 („very sleepy, great effort to stay awake, fighting sleep“) der KSS sind Mikroschlaf validiert (Marsalek et al., 2006) und mit einem Risiko von 17% für vollständige Fahrbahn Überschreitungen verbunden (persönliche Mitteilung, Akerstedt, 2007). Des Weiteren ist die Karolinska Sleepiness Scale deutlich mit EEG und EOG Indikatoren von Schläfrigkeit assoziiert (Akerstedt & Gilberg, 1990; Gilberg et al., 1994; Kaida, Takahashi, & Akerstedt et al., 2006). Gilberg et al. (1994) fanden darüber hinaus Korrelationen der KSS mit Vigilanzaufgaben und visuellen Signalentdeckungsaufgaben ($r = -.62$). In Luft- und Raumfahrt Forschungskontexten genießt wiederum die Samn-Perelli Skala und auch die Visuelle Analog Skale (VAS) eine große Verbreitung. Gemeinsam ist den Selbstreport Skalen jedoch die Problematik, Symptome von Ermüdung häufig als Schläfrigkeit misszuinterpretieren. Darüber hinaus ist aufgrund der fehlenden Innensicht der eigene Schläfrigkeitsgrad häufig unterschätzt. So fanden Pilcher und Huffcutt (1996), dass je höher der Grad der Schläfrigkeit, desto ungenauer die Einschätzungen des eigenen Schläfrigkeitszustandes.

Kognitive Leistungsverfahren. Da zahlreiche empirische Befunde schläfrigkeitsbedingte kognitive Vigilanzeinbußen belegen (Bonnet, 2000; Dinges & Kribbs, 1991; Pilcher & Huffcutt, 1996), werden Daueraufmerksamkeitsaufgaben zur Schläfrigkeitsbestimmung herangezogen. Ein Vigilanztest erfordert, auf zeitlich und räumlich meist unregelmäßig auftretende Reize zu reagieren. Kurzfristig auftretende Aufmerksamkeitseinbrüche, im Extremfall Mikroschlafepisoden, können während des Tests unentdeckt bleiben, wenn in dieser Zeit keine Reizreaktion erforderlich ist. Verfahren zur Erfassung der Daueraufmerksamkeit hingegen testen die kontinuierlich fokussierte Aufmerksamkeit über einen längeren Zeitraum. Bei der Testung der Daueraufmerksamkeit muss daher häufig reagiert werden, die Dichte der kritischen Reize in der Aufgabe ist darum hier höher. Ein Verfahren zur Messung von Daueraufmerksamkeit ist der *Psychomotor Vigilance Task* (PVT). Hierbei handelt es sich um eine einfache psychomotorische Reaktionsaufgabe mit hoher Reizdichte. Über 10 Minuten muss, so

schnell wie möglich, auf wiederkehrende visuelle oder akustische Reize mittels Tastendruck reagiert werden. Sie hat sich als sensitiv für zirkadian- oder schläfrigkeitbedingte Veränderungen der Wachheit (Alertness) erwiesen (Dinges et al., 1997; Jewett al., 1999; Van Dongen et al., 2003). Als mögliche Messgrößen kommen in diesem Leistungstest Lageparameter (Häufigkeit verschiedener Fehlerarten z.B. Auslasser vs. Falscher Alarm; Mittelwert der RT; Median der RT; mittlere RT der 10 % langsamsten Reaktionen), Streuungsparameter (Streuung der RT; Interquartile Distanzen der RT; Differenz zwischen den 10 % langsamsten und schnellsten Reaktionen) und Verlaufparameter (Anstieg der Fehlerrate; Anstieg der RT; Leistungsdifferenz Anfangsphase vs. Endphase der Testung) zum Einsatz. Ein anderer vielversprechender Leistungstest ist der Unstable Tracking Test (Marsalek et al., 2006), der in der Pocket PC Version ein 10 minütiges Tracking eines zufällig horizontal bewegenden Punktes mittels Touchscreengriffels beinhaltet.

Performanz in Fahraufgaben. Weitere Messgrößen zur Bestimmung von Schläfrigkeit können aus der Performanz realer Arbeitsaufgaben abgeleitet werden. Erste Zusammenhänge von Fahrgeschwindigkeitsschwankungen, Lenkradbewegungen, Spurabweichungen und Schläfrigkeit wurden bereits von Safford und Rockwell (1967), Brown, Tickner und Simmonds (1970), Sussman und Morris (1970), Hulbert (1972), Wierwille und Muto (1981) sowie Williamson et al. (1994) dokumentiert (vgl. Tabelle 3). In Fahraufgaben können so, abgeleitet aus Geschwindigkeits-, Lenkwinkel- und Querabweichungssignalen, wichtige Kennzahlen bestimmt werden. Erfasste Kennzahlenbereiche sind z.B. die Präzision der Geschwindigkeitsregelung, Lenkwinkelgeschwindigkeiten (Mittelwert, Standardabweichungen, Mittelwert der Amplituden über 10 Grad/s, Anzahl der überschwelligen Amplituden pro Minute), Querabweichungsindikatoren (Kennzahlen wie die mittlere oder maximale horizontale Fahrbahnabweichung (Standard Deviation Lane Deviation, oder Ereignisse wie Schwerpunkt, 2-Räder oder 4-Räder des Fahrzeugs über der Spurbegrenzung) und Gierwinkelfehler (Abweichungen der Fahrzeuglängsachse um die aktuelle Fahrbahntangente). Aber auch die Anzahl kritischer Zwischenfälle (Critical Incidents) wie dem Überfahren eines Stoppschildes oder roten Ampel, oder einem Beinaheunfall) können gemessen werden. Die verkehrssicherheitsrelevante Bedeutung betreffend muss jedoch ein Unterschied zwischen simulationsbezogener und realer Performanz gezogen werden. Da Reservekapazitäten nicht vollständig genutzt werden, treten die Verschlechterungen der Leistungsfähigkeit in der Simulation früher auf als in realen Kontexten, in denen sie bei professionellen Fahrern erst in sehr späten Schläfrigkeitsstadien erscheinen (Tornros, 1998).

Tabelle 3: Beispiele fahrverhaltenbezogener Maße der Schläfrigkeitsbestimmung

Fahrverhaltenskategorie	Referenz
Spurhaltemaße, laterale Position	Bittner et al., 2000, Boverie et al., 2002, Boverie, 2004, Flescher et al, 2003, Grace et al., 1998, Pilutti & Ulsoy, 1995, Rimini-Doering et al., 2001, Santana Diaz et al., 2002
Fahrzeuggeschwindigkeit	Boverie et al., 2002, Santana Diaz et al., 2002
Beschleunigungsverhalten, Bremsverhalten	Grace et al., 1998
Pedal-Bewegung	McCall & Trivedi, 2004

Behaviorale Verfahren. Die Erfassung von verhaltensbezogenen Größen wie der Kopfneigung und Kopfbewegung (Wright & McGown, 2001), der Körperhaltung, der reduzierten Mimik (Wierwille & Ellsworth, 1994), dem Gähnen (Baenninger, Binkley & Baenninger, 1996), der Schreckreaktion sowie dem Reiben und Kratzen von Gesicht und Augen (selbstaktivierendes Verhalten) ermöglichen einen Rückschluss auf Schläfrigkeit. Diese Mannierism Maße münden in Scoringverfahren wie der Observer Rating Scale (Kecklund, Anund, Portin & Akerstedt, 2007), dem Wierwille Rating (Vöhringer-Kuhnt, Baumgarten, Karrer & Briest, 2004; Wierwille & Ellworth, 1994), oder dem von Muttray et al. (2007) weiterentwickelten Indikatorensystem, die Ereignisse erfassen wie: große Oberkörperbewegungen und -verschränkungen, das Wegnicken des Kopfes, das Haarkratzen und –glattziehen, das Nacken- und Augenreiben-, halten, -kratzen, das Augenbrauen heben und Augen weit öffnen, das Lippen benetzen und –beißen, das nicht sprachgebundene auffällige Bewegen der Zunge oder die über Sekunden fast geschlossenen Augen. Nachteilig bei dieser Messmethode ist der mit der Rating Prozedur verbundene größere Aufwand. Fraglich bleibt auch, in wieweit alle Müdigkeitsformen über äußerlich zu beobachtende Indikatoren abgebildet werden können. Diese Einschränkung betrifft auch die automatische Erfassung dieser Größen, die mittels computer-gestützter Objekterkennungssoftware („Seeing Machine“), die das Tracking von Gesichts-, Kopf-, Extremitäten- und Körperbewegungen übernehmen, möglich ist (Kapoor et al., 2003).

Weitere Entwicklungen könnten, wie bereits für die Erfassung von Basisemotionen angewendet, weiteres Bewegungsverhalten wie die Mouse-, Keyboard- oder berührungssensitive Touchpad Interaktion (Schuller, 2006; Zimmermann et al., 2003) oder die aus der Fahrzeugbedienung abgeleiteten Lenkrad- und Pedalbewegung (Sandberg & Wahde, 2007) zur Messung von Schläfrigkeit nutzen. Erfasst werden könnten hierzu die räumliche Ausdehnung, Menge, Amplitude, Dauer, Flüssigkeit, Rhythmik und Dynamik der Bewegungen (Kollias et al., 2004). Erste Versuche aus Körperpostur (Andreeva et al., 2004) sowie Limb- und Körperbewegungen Schläfrigkeit zu erfassen, zeigen bereits vielversprechende Erfolge (Bergasa et al., 2006; Boverie, 2004; Grace et al., 1998; Graumann et al., 2002; Heitmann et al., 2000; Zhu & Li, 2004).

Blick- und Pupillenbezogene Verfahren. Zur Erfassung von mentaler Beanspruchung oder Schläfrigkeitszuständen kommen gegenwärtig die Analyse von Sakkadenbewegungen, Lidschlussverhalten (Caffier, 2002; Erdmann et al. 2006; Galley & Schleicher, 2002; Zils, 2005) - sowie Instabilitätsmaße der Pupillengröße zum Einsatz (Heitmann et al, 2001; Lütcke et al., 1998; Schwalm, 2007; Wang et al., 2003; Wilhelm et al., 2001a; Wilhelm et al., 2005). *Lidschlussverhalten.* Ein Einschlafsmerkmal ist das langsame Absinken der Augenlider (Slow Eyelid Droop). Beim Lidschlussmaß „Perclos“ (Percentage of Eyelid Closure) wird Schläfrigkeit über den Zeitanteil gemessen, in dem die Augen bezüglich der Augenlidspalte zu mindestens 80 % geschlossen sind. Ausgeschlossen werden Blinzelvorgänge, d. h. Lidschüsse mit einer Dauer von weniger als 300 ms. Ein kritischer Perclos-Wert liegt bei 15 %, was bei einem Beobachtungszeitraum von 60 Sekunden, einer Zeitdauer von 4 Sekunden entspricht, bei der die Augen zu 80-100 % geschlossen sind (Dinges et al., 1998; Bergasa et al., 2006; Grace et al., 1998; Graumann et al, 2002; Wierwille et al., 1994). Häufig wird jedoch eingewandt, dass das langsame Schließen der Augen, ein ziemlich spätes Phänomen im Übergang vom Wachzustand zum Schlaf darstellt und es an der nötigen Sensitivität für frühere Phasen schläfrigkeitsbedingter Schwankungen des Leistungsvermögens fehle (Mallis, Maislin, Powell et al. 1999; Wright & McGown, 2001).

Zur Beschreibung des Blinzelverhalten der Augen werden daher Maße wie „Delay of Lid Reopening“, „Blink Interval“ und „Blink Duration“, „Blink Amplitude“, „Lid Closure Speed“, „Peak Closing Velocity“, „Lid Opening Speed“, „Peak Opening Velocity“, „Closing Time“, „Opening Time“ und „Standardised Lid Closure Speed“ verwendet. Der mit subjektiven Vigilanzmessungen am besten korrespondierende Kennwert ($r = .50$) ist die Dauer der

Eye Blinks (Bergasa et al., 2006; Boverie et al., 2002; Heitmann et al., 2001; Schleicher, Galley, Briest & Galley, submitted). Problematisch erscheint bei diesem Messansatz seine mögliche Konfundierung mit nicht zustandsrelevanten Personen- und Umgebungsvariablen wie Zugluft, Blickbewegungen, Blickrichtung, Rauchen, Luftfeuchtigkeit, soziale Situation und Gespräche (Hargutt, 2003).

Okulomotorische Aktivität. Mit zunehmender Schläfrigkeit kommt es zu spezifischen Variationen spontaner okulomotorischer Aktivität. Augenbewegungs-Parameter wie etwa Sakkadenbewegungen, reagieren sensitiv auf erste Anzeichen von Schläfrigkeit (Bergasa et al., 2006; Boverie, 2004; Caffier, 2002; Porcu, Ferrara & Urbani, 1998; Rimini-Doering et al., 2001; Wahlstrom, 2003; Wang et al., 2003; Zhu & Li, 2004). Die Schläfrigkeitskomponente Deaktivierung ist sowohl durch eine wachsende Dauer der Blinks und Sakkaden, als auch durch deren abnehmende Geschwindigkeit charakterisiert. Abnehmende Aufmerksamkeit ist durch abnehmende Blink Intervalle und „Fight against Sleepiness“ durch eine steigende Standardabweichung von vielen okulomotorischen Parametern gekennzeichnet (Hagenmeyer et al., 2006; Hargutt, 2003; Schleicher, Galley, Briest & Galley, submitted).

Pupillographie. Der Pupillographische Schläfrigkeitstest (PST, Wilhelm et al., 2006) erfasst Schläfrigkeit über die Detektion von Spontanfluktuationen der Pupillengröße im Dunklen mittels Infrarot-Videographie (vgl. O’Neill, Oroujeh & Merritt, 1998; Micieli, Manni & Tassorelli et al., 1995). Im Wachen und bei hoher zentralnervöser Aktivierung bleibt die Pupillenweite unter Ausschluss von Lichteinfluss für lange Zeit stabil. Bei erhöhter Schläfrigkeit treten hingegen bereits nach wenigen Minuten deutliche Schwankungen der Pupillenweite auf. Als Maß für die niederfrequenten Pupillenoszillationen wird beim PST der Pupillenunruhe-Index (PUI) herangezogen. Darüber hinaus gelang Golz et al. (2001) die Vorhersage von Mikroschlafepisoden über den während einer Fahrsimulatorfahrt aufgenommenen zeitlichen Verlauf des Pupillendurchmessers mit einer Klassifikationsgenauigkeit von 82 %. Gemeinsam ist den Augen- und Pupillen-basierten Schläfrigkeitsverfahren jedoch bislang ihre Anfälligkeit für Messausfälle durch wechselnde Beleuchtungen, Tag- und Nachtunterschiede, ungünstige Kopforientierungen oder das Tragen von Brillen.

Elektrophysiologische Verfahren. Die Aktivität des zentralen Nervensystems sowie vegetative und motorische Funktionen können mittels elektrophysiologischer Messungen untersucht werden. Ein zur Bestimmung der Hypovigilanz genutztes Verfahren ist die *elektrodermale Aktivität* (EDA). Als Schläfrigkeitsindikatoren dienen der steigende Hautleitwiderstand bei Schläfrigkeit und Spontanfluktuationen im EDA-Signal (Boucsein, Haarmann, & Schaefer, 2007; Nishimura & Nagumo, 1985; Boucsein & Ottmann, 1996). Individuelle Variationen und Umweltfaktoren wie Lärm, Physische Bewegung und Feuchtigkeit können diese Messung jedoch konfundieren. Dennoch ist das „Engine Driver Vigilance Telemetric Control System (Neurocom, 2006) derzeit innerhalb des russischen Bahnverkehrs im Einsatz. Ähnliches gilt für die Indikatorfunktion des *Skin Potential Level*. Erwin, Hartwell, Volow und Alberti (1976) berichten von Zusammenhängen zwischen dem Skin Potential Level (SPL) und dem Arousalniveau. Das SPL bestimmt sich hier in der Differenz des elektrischen Potentials zwischen der äußeren Hautschicht (Stratum Corneum) und der direkt darunter liegenden Hautschicht (Stratum Lucidum). Einige Minuten nach größeren Veränderungen im SPL zeigen Probanden deutliche Schläfrigkeitsindikatoren wie reduzierte Leistungsmaße, lange Eye-lid Closures und veränderte Elektroenzephalogramm (EEG) basierte Spektralmaße (Erwin *et al.*, 1976).

Polysomnographie. Die Polysomnographie – bestehend aus EEG, EOG (Elektrookulogramm) und EMG (Elektromyogramm) – kann wertvolle Informationen liefern, um nicht nur Schlafstadien zu differenzieren, sondern darüber hinaus auch Mikroschlaf oder Vigilanzeinbrüchen vorherzusagen (Berrichi *et al.*, 1999). Ferner ist die Nutzung der Herzraten Variabilität als Schläfrigkeitsmarker denkbar (Törnös, Peters, Östlund, 2000; Volow & Erwin, 1973). Neben elektrophysiologischen Verfahren ist auch der Einsatz biochemischer Marker denkbar. Erste Entwicklung basieren hier auf der Schätzung des Alphaamylase Niveaus und sind bereits in mobilen Geräten implementiert (Yamaguchi *et al.*, 2006).

EEG-basierte Verfahren. *EEG Frequenzband Analyse.* Zu den EEG-basierten Kennzahlen, die als Schläfrigkeitsindikatoren zählen, gehören Beta-Wellen, Alpha-, Theta- und Delta-Wellen bzw. ihre Bandenergie Quotienten (z.B. (Theta + Alpha) / Beta) (Bittner *et al.*, 2000; Cantero & Atienza, 2000; Eoh *et al.*, 2005; Golz *et al.*, 2007; Gonzales-Mendoza *et al.*, 2003; Gottlieb, Galley, Schleicher, Galley & Churan, 2004; Jung, Makeig, Stensmo & Sejnowski, 1997; Lal *et al.*, 2003; Makeig *et al.*, 2000; Sayed & Eskandarian, 2001; Verwey & Zaidel, 2000; Wu, Lin, Liang & Jung, 2004). Problematisch ist jedoch die große interindividuelle

Variabilität im EEG bei schläfrigen Personen (Santamaria & Chiappa, 1987). Damousis & Cester (2007) extrahierten mittels Power-Spektralanalyse aus Alpha- und Thetabändern EEG Features. Golz et al. (2007) fanden, dass die Detektion von Mikroschlafphasen, während des Fahrens am erfolgreichsten ist, wenn alle 15 EEG-Kanäle gemeinsam ausgewertet werden. Des Weiteren sind EEG Datenmodelle auch effizient in der Vorhersage von schläfrigkeitsbedingten Simulatorunfällen (Damousis & Cester, 2007). So steigt die spektrale Energiedichte (Power Spectral Density, PSD) im Theta-Alpha-Range (Cajochen et al., 1995; Horne & Baulk, 2004). Lal und Craig (2002) fanden einen Anstieg der PSD im Delta-Theta-Alpha-Beta-Range um einen Faktor von 22%, 26%, 9% und 5%.

Alpha Attenuation Test (AAT). Der Alpha Attenuation Test (AAT) (Michimori, et al., 1994; Stampi et al., 1995) basiert auf der Beobachtung, dass die Alpha-Frequenz des EEGs sich verändert, wenn Testpersonen vom wachen in den schläfrigen Zustand übergehen. Bei geschlossenen Augen nimmt die Bandbreite der Alpha-Frequenz ab und steigt wiederum an, wenn die Augen geöffnet werden. Der Alpha Attenuation Coefficient (ACC) wird ermittelt, indem das Verhältnis der durchschnittlichen Alpha-power Werte, der Bedingungen „Augen-geschlossen“ zu „Augen-geöffnet“, berechnet wird. Untersuchungen zeigten, dass der ACC bei Narkolepsipatienten im Vergleich zur Kontrollpersonen signifikant kleiner ist (Alloway et al., 1997; Gizycki et al., 1997).

Karolinska-Drowsiness-Scoring. Das Karolinska-Drowsiness-Scoring von Gillberg, Kecklund und Åkerstedt (1996) entstand in Anlehnung an die von Rechtschaffen und Kales (1968) entwickelte EEG-basierten Bewertung der Schlafstadien. Die aufgenommenen EEG-Daten werden in 20-Sekunden Epochen unterteilt, die wiederum in zwei Sekunden Einheiten (Bins) segmentiert werden. Der resultierende Karolinska-Drowsiness Score (KDS) beschreibt die Anzahl der Bins, die langsame Alpha-Aktivität oder Theta-Aktivität beinhalten und mit langsamen Augenbewegungen verbunden ist. Jeder zwei Sekunden Bin, der diese Zeichen von Schläfrigkeit aufweist, erhält 10 KDS Punkte. Der maximale KDS ist 100 und entspricht dem Auftreten von Schläfrigkeitszeichen in allen zwei Sekunden Bins. Die Epochen mit KDS > 50 werden als Schlaf-Onset beurteilt.

Einschlaflatenz basierte Verfahren - Multipler Schlaflatenz Test (MSLT). Die einfachste operationale Definition von Schläfrigkeit bezieht sich auf die Stärke der Einschlafneigung. Der Multiple Schlaflatenz Test (MSLT) misst in diesem Sinn daher unter kontrollierten Bedingungen wiederholt die benötigte Einschlafzeit. Ein schnelleres Einschlafen spiegelt dabei eine größere Schläfrigkeit wider. Das Verfahren wurde von Carskadon und Dement (1977) vorgestellt und gilt häufig als Standardmethode der Schläfrigkeitsmessung (Bonnet & Arand, 1998; Thorpy, 1992).

Maintenance of Wakefulness Test (MWT). Der Maintenance of Wakefulness Test (MWT) oder Multiple-Wachbleibe-Test erfasst den Einschlafdruck unter monotonen, einschlaf-fördernden Bedingungen (Doghranji et al., 1997; Martin, Engleman & Deary et al., 1996; Mitler et al., 1982; Richardson et al., 1978). Das Einschlafverhalten wird wie beim MSLT über die polysomnographisch registrierte Einschlafatenz erfasst. Die zahlreichen verfügbaren Messverfahren korrelieren häufig nur in moderater Größenordnung. Das gilt auch für den direkten Vergleich der schlafmedizinisch etablierten Verfahren MSLT und MWT. Die geringen Korrelationen können darauf zurückgeführt werden, dass verschiedene Verfahren nur in bestimmten Schläfrigkeitsbereichen sensitiv sind oder einfach verschiedene Facetten des Schläfrigkeitskonstrukts erfassen. Einen ausführlichen Überblick über standardisierte Verfahren, die im deutschsprachigen Raum breite Anwendung finden, liefert ein Übersichtsartikel der Deutschen Gesellschaft für Schlafforschung und Schlafmedizin (Weeß et al., 2000).

1.5 Mehrwert des akustischen Messansatzes

Der Einsatz akustischer Informationen zur Bestimmung des Befindlichkeitszustandes ist schon seit vielen Jahren anvisiert (Hecker, Stevens, von Bismarck & Williams, 1968; Liebermann & Michaels, 1962; Scherer, 1974; William & Stevens, 1972). Aber erst die Fortschritte der Prozesstechnik und die Entwicklung leistungsfähiger Sprachanalyse-Software (z.B. Praat, Matlab) machten den breiten Einsatz von Sprachverarbeitung auf hohem Niveau möglich. Bausteine zur automatischen Bestimmung von internen Zuständen aus Stimmäußerungen liefern hierzu eine Reihe wissenschaftlicher Disziplinen (siehe Tabelle 4).

Tabelle 4: Wissenschaftsdisziplinen mit signifikantem Beitrag an der Entwicklung der stimmbasierten akustischen Schläfrigkeitmessung

Beteiligte Disziplinen	Beitrag zur Acoustic Sleepiness Analysis
Linguistik und Phonetik	<ul style="list-style-type: none"> • Sprachproduktionstheorien • Audiv-perzeptive Beschreibungskategorien der Stimme • Segmentierung des Sprachmaterials in phonetische Einheiten
Phoniatrie	<ul style="list-style-type: none"> • Konzepte, Merkmale und Kennwerte verschliffener Artikulation • Neuronale Artikulationsmodelle
Schlafmedizin	<ul style="list-style-type: none"> • Konstruktion und Validierung von Schläfrigkeitmessinstrumenten • Grundlagenforschung zu physiologischen und zentralnervösen Korrelaten von Schläfrigkeit
Arbeitspsychologie	<ul style="list-style-type: none"> • Grundlagenforschung zu verwandten Beanspruchungs-, Aufmerksamkeit-, Ressourcen- und Aktivierungskonzepten • Anwendungsfelder der Schläfrigkeitmessung zur Optimierung des Arbeits-Erholungs-Zyklus • Arbeitspsychologische Befunde zur unfallbezogenen Relevanz von Schläfrigkeit
Nachrichtentechnik, Signal und Audio Processing	<ul style="list-style-type: none"> • Signalvorverarbeitung (Filter, Framing, Windowing) • Sprachunabhängige Signalverarbeitungskennzahlen
Künstliche Intelligenzforschung	<ul style="list-style-type: none"> • Machine-Learning Prozedere • Merkmalsselektionsverfahren • Automatische Merkmalsgenerierung mit genetischen Algorithmen • Merkmalsklassifikation- und Mustererkennungsverfahren

Wissenschaftliche und anwendungspraktische Bewertungskriterien. Wichtige Kriterien für die Bewertung von Schläfrigmessverfahren sind (a) die Validität (auch Sensitivität und Spezifität), (b) die Ökonomie in Messvorbereitung, -betreuung u. -auswertung (bzgl. des zeitlichen, personellen und finanziellen Aufwands), (c) die messvorgangsbezogene Belästigung des Messobjekts (durch Trainingsphasen zur individuellen Baseline-Bestimmung, durch Re-Kalibrierungsphasen vor jedem Messbeginn oder durch Sensorapplikationen am Körper) und (d) die Entwicklungsmöglichkeiten des Messansatzes in Richtung einer automatischen Echtzeit-Erfassung von Schläfrigkeit in realen Messanwendungssituationen.

Anders als bei Selbsteinschätzungsverfahren und kognitiven Leistungsverfahren ermöglicht der akustische Stimmanalyseansatz eine beliebig wiederholbare, den Tätigkeitsvollzug nicht beeinträchtigende und verfälschungsresistente Erfassung von Schläfrigkeit. Im Gegensatz zu physiologischen Ansätzen bietet der akustische Stimmanalyse Ansatz, die Vorzüge eines *berührungs- und sensorapplikationsfreien, ökonomischen sowie belästigungs- und kalibrierungsarmen Messzugangs*. Der laufende Tätigkeitsvollzug wird nicht beeinträchtigt, da eine „Hands-free“ und „Eyes-free“ Messsituation vorliegt. Des Weiteren ist im akustischen Ansatz eine aus Komfortgründen unerwünschte, unmittelbar vor der Messung stattfindende Kalibrierungsphase vermeidbar (vgl. dazu z.B. die in der videobasierten Erfassung der Okulomotorik notwendigen Referenz-Fixationen). Schließlich kann durch ein *sprecherunspezifisches Vorhersagemodell* oder durch eine automatische Anpassung des Vorhersagemodells während der Bedienung (die erkannte Schläfrigkeit dient als rückgekoppelte Überwachungsgröße) auf eine entsprechende Trainingsphase verzichtet werden.

Die Vorzüge des akustischen Ansatzes gegenüber raterabhängigen behavioralen Ansätzen liegen in der zeitlichen, personellen und finanziellen Ökonomie des Messansatzes, seines geringeren Belästigungswertes (Mikrophonaufnahmen im Gegensatz zu teilnehmenden Beobachtungen) sowie seines Entwicklungspotenzials in Richtung automatischer realtime Erfassung. Im Vergleich zu rechnergestützten, objekterkennungsbasierten behavioralen Ansätzen („Seeing Machines“) vermeidet die akustische Analyse über seine dezenten Mikrophonaufnahmen das Gefühl einer permanenten, Privatheit reduzierenden Überwachung. Auch ist die benötigte Mikrofon-Standardhardware kostengünstig und in Fahrzeugsystemen serienmäßig installiert, insbesondere im Luftfahrtsektor liefert die funkverkehrgestützte Kommunikation, die zur Spracherfassung nötige Sensor-Infrastruktur. Für wachsende Verfügbarkeit verwertbarer Sprachäußerungen in anderen Bereichen sorgen die in Zukunft sprachgesteuerten

Mensch-Maschine Schnittstellen, wie z.B. Spracherkennungstechnologie in Fahrerassistenzsystemen (z.B. Navigationssysteme, sprachgesteuerte Mobiltelefon Einwahl und Multimedia Anwendungen- „In-Car-Dialog-Systems“) und Computerarbeitsplätzen. Des Weiteren könnte die Integration von akustischen Systemen in den Verbund einer übergeordneten multimodalen Systemarchitektur zur Schläfrigkeitswarnung wiederum den Nachteil temporär nicht verfügbarer Spracheingaben kompensieren und über eine wechselseitige Kreuzvalidierung die Akkuratheit und Robustheit erhöhen. Weitere Berücksichtigung bei der Bewertung von Messgeräten sollten Faktoren wie Raumerfordernisse und Portabilität der Messapparatur oder sonstige Begleiterscheinungen der Messung wie das Erzeugen von biohazardösen Müll finden. Die nachfolgende Tabelle 5 fasst in einem Kurzüberblick die wissenschaftliche und anwendungsorientierte Nützlichkeit der Schläfrigkeitsmessansätze zusammen.

Tabelle 5: Vergleichender Kriterienkatalog zur Bewertung der wissenschaftlichen und anwendungsorientierten Nützlichkeit von Schläfrigkeitsmessansätzen. („+“ entspricht einer günstigen, „o“ einer durchschnittlichen und „-“ einer ungünstigen Ausprägung; SEB = Selbsteinschätzungsverfahren, KOG= kognitive Verfahren, PHY= physiologische Verfahren, AUG= Augen- und Pupillenbezogene Verfahren, BRA= behaviorale Raterverfahren, BSY= behaviorale systemgestützte Objekterkennungsverfahren, ELA= Einschlaf latenz-basierte Verfahren, AKU= akustische Messverfahren)

Bewertungskriterium	SEB	KOG	PHY	AUG	BRA	BSY	ELA	AKU
Validität	o	o	+	+	o	o	+	?
Ökonomie in Messvorbereitung, -betreuung u. –auswertung								
zeitlicher Aufwand	+	+	-	+	+	+	-	+
personeller Aufwand	+	+	-	o	-	+	-	+
finanzieller Aufwand	+	+	-	-	+	-	-	+
Belästigung des Messobjekts								
individuelle Adaptation	+	-	o	+	+	+	+	o
Kalibrierungsphase	+	o	o	o	+	o	-	+
während der Messung	o	-	-	+	+	+	-	+
Eignung für Online Monitoring Realanwendungen	-	-	o	+	-	+	+	+

2 LINGUISTISCHE, PHONETISCHE UND AKUSTISCHE GRUNDLAGEN

2.1 Phonetische Grundlagen der Sprachproduktion

Nachdem im zurückliegenden Kapitel 1 Messansätze zur Detektion von Schläfrigkeit vorgestellt und potenzielle Mehrwerte des akustischen Messzugangs herausgearbeitet worden sind, vertieft das nachfolgende Kapitel 2 die linguistischen, phonetischen und akustischen Grundlagen der Sprachproduktion. Zunächst wird ein Überblick über phonetische Prozesse der Sprachproduktion gegeben. Anschließend werden die auditiv-perzeptiven Stimmkategorien beschrieben sowie akustische Parameterklassen erläutert. Der Sprechapparat des Menschen kann sowohl funktional als auch anatomisch in drei Segmente unterteilt werden. Ein Initiations-, ein Phonations- und ein Artikulationssystem. Die Lunge stellt im Initiationssystem die aerodynamische Energie (Luftdruck und Luftstrom) zur Verfügung (Catford, 1988), die im Kehlkopf in ein akustisches Signal geformt (Phonationssystem) und im Ansatzrohr (Rachen-, Mund- und Nasenraum) moduliert wird (Artikulationssystem) (Kröger, 2007). Der Sprechapparat kann in diesem Sinne in den sublaryngalen (auch: subglottalen) Bereich (Lunge, Bronchien und Luftröhre), die Larynx (Kehlkopf) und den supralaryngalen (auch: supraglottalen) Bereich (Rachen-, Mund- und Nasenraum) unterteilt werden (siehe Abbildung 2 und Abbildung 3).

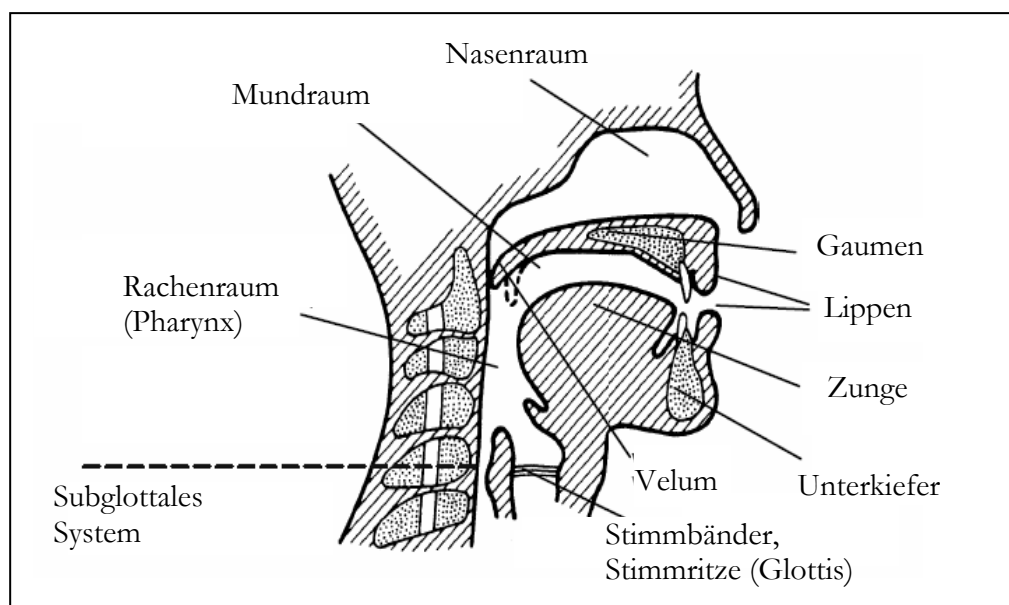


Abbildung 2: Der Sprechapparat des Menschen (aus Hess, 2002)

Initiation

Die Lungenaktivität während des Sprechens erfüllt aus der Perspektive der Sprachproduktion zwei wichtige Funktionen: Zum einen ist sie verantwortlich für die Herstellung und Aufrechterhaltung eines zeitlich konstanten Luftüberdrucks in der Lunge, zum anderen für die Bereitstellung einer bestimmten Luftmenge als Basis für den zum Sprechen nötigen Luftstrom. Die Atmung liefert den für die Rohschallgenerierung nötigen subglottalen Luftdruck, durch den die Stimmlippen (Vocal Cords) im Kehlkopf in klangerzeugende Schwingungen versetzt werden (siehe den als in Kolbenfunktion dargestellte Abbildung 3). Die Atmung basiert auf der mit der Ausdehnung bzw. Verengung des Brustraumes verbundenen Veränderung des Lungenvolumens und damit des Luftdrucks innerhalb der Lunge. Aufgabe der Atmungsmuskulatur ist es, während des Sprechens den Lungenluftdruck annähernd konstant zu halten (Ohala, 1990). Erreicht wird dieses über zwei Atmungsformen, die Brustatmung (Hebung und Senkung des Brustkorbes) und die Bauchatmung (mit Zwerchfellaktivität verknüpfte Bauchwandbewegung). Beteiligt an der Einatmung sind der Zwerchfellmuskel (Diaphragma) und die äußeren Zwischenrippenmuskeln (Mm. intercostales externi). Bei forcierter Einatmung können des Weiteren Teile der Hals-, Schulter- und Rückenmuskulatur bei der Einatmung beteiligt sein. Wichtige Muskeln der Ausatmung sind die inneren Zwischenrippenmuskeln (Mm. intercostales interni) aber auch Teile der Bauch- und Rückenmuskulatur. Die Ausatmung erfolgt jedoch bei Ruheatmung im Wesentlichen passiv durch elastische Rückstellkräfte des Gewebes: Die Lunge stellt sich bei fehlender Kontraktion des Diaphragmas und dem entgegenwirkenden Druck des Bauchhöhleninhalts elastisch auf ihr Ruhevolumen zurück, das im Ruhezustand durchschnittlich vier Liter beträgt. Ebenso bewirken bei nichtaktiven äußeren Zwischenrippenmuskeln die Verdrehungskräfte an den knorpeligen Rippenteilen eine Rückstellung. Ferner senken sich aufgrund der Schwerkraft die beim Einatmen angehobenen Rippen wieder in ihre Ruhelage (Pétursson & Neppert, 2002; Pompino-Marschall, 1995).

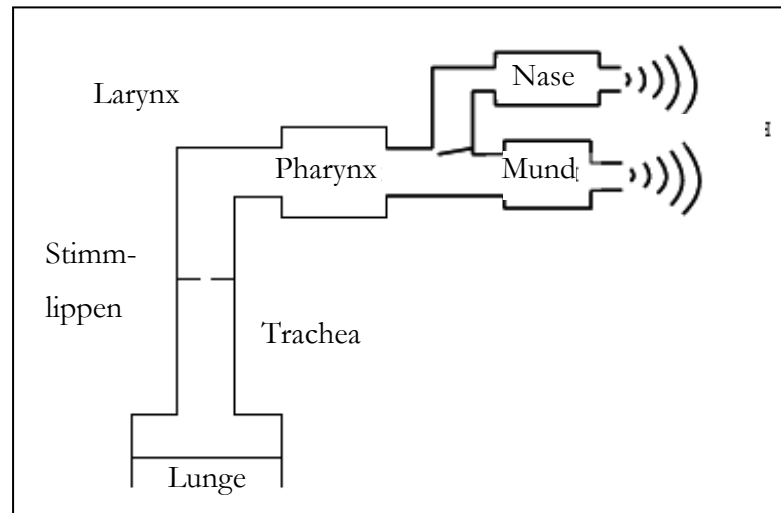


Abbildung 3: Schematisches Modell des Sprachproduktionssystems (nach Lindgren, 2003). Dargestellt sind die kolbenartig Luftdruck generierende Lunge (Initiation), die anschließende Verengung der Luftröhre an den Stimmlippen (Phonation), der Kehlkopfresonanzraum, sowie die abschließende Schallabstrahlung über Mund- und Nasenöffnungen

Phonation

Im Folgenden werden die den ausgeatmeten Luftstrom in einem ersten Schritt modifizierenden Mechanismen näher erläutert. Es handelt sich hierbei um Vorgänge innerhalb des Kehlkopfs, der bei stimmhaften Lauten den Rohschall generiert. Ein Teil der aerodynamischen Energie wird im Kehlkopf in Schwingungen der Stimmlippen und somit in ein akustisches Signal (Schwingungen der Luftsäule im Ansatzrohr) umgesetzt. Die Produktion eines akustischen Signals im Kehlkopf (Stimmton) über Stimmlippenschwingungen wird als Phonation bezeichnet. Der *Kehlkopf* (Larynx) bildet den oberen Abschluss der Luftröhre (Trachea) gegenüber dem Rachen- und Mundraum (siehe Abbildung 2). Die Primärfunktion des Kehlkopfes ist der Schutz der Luftwege (Pétursson & Neppert, 2002). Für die lautsprachliche Kommunikation besteht seine Funktion in der kontrollierten Stimmtoneerzeugung (Phonation) durch schwingende Stimmlippen. Die Kehlkopfmuskulatur bestimmt im Zusammenspiel mit dem Druck der ausgeatmeten Luft das Auftreten, die Geschwindigkeit, die Stärke und die Form der Stimmlippenschwingung. Die Stimmhaftigkeit, Stimmtonehöhe, Lautstärke und Stimmqualität kann über diese Prozesse vermittelt kontrolliert werden (Pompino-Marschall, 1995).

Der Kehlkopf besteht aus fünf gelenkig miteinander verbundenen Knorpelstrukturen, fünf intrinsischen Muskelgruppen, Bändergewebe und auskleidenden Schleimhäuten. Er ist durch extrinsische Muskeln oberhalb mit dem Zungenbein (Hyoid) und unterhalb mit dem Brustbein (Sternum) verbunden. Dadurch ist er in seiner vertikalen Lage veränderbar. Zwischen den Ary-Knorpeln (Stellknorpeln) und der mittleren Innenkante des Thyroids (Schildknorpel) spannen sich *die Stimmbänder* (L. vocale) und *Stimmlippenmuskeln* (siehe Abbildung 4). Der Zwischenraum zwischen den Stimmlippen und den Ary-Knorpeln wird *Glottis* genannt. Die beweglich an der inneren Vorderkante des Thyroids ansetzende blattförmige Epiglottis schließlich bildet den zurückklappbaren Verschlussdeckel des Kehlkopfes.

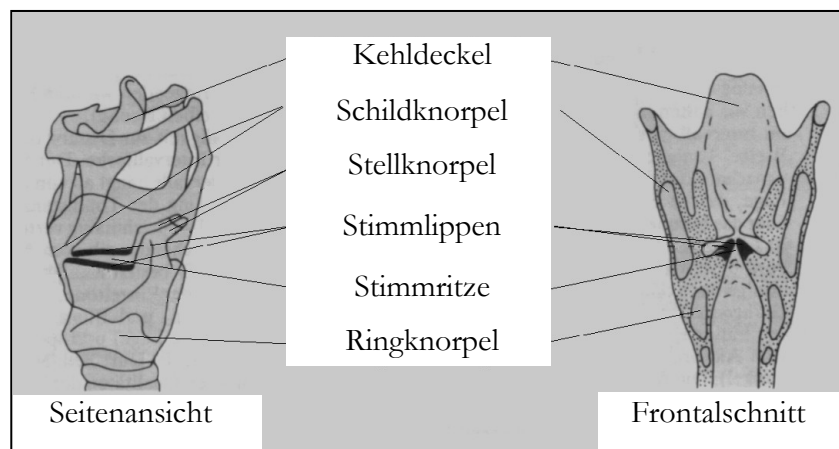


Abbildung 4: Seitenansicht und hinterer Frontalschnitt des Kehlkopfes (nach Kröger, 2007)

Der Phonationsvorgang. Die *Stimmtonerzeugung* im Kehlkopf erfolgt durch die Schwingung der Stimmlippen und basiert auf einem komplexen Zusammenspiel von neuraler Aktivität, Muskelspannung, Regulation des subglottalen Luftdrucks und aerodynamischen Prozessen. Neurale Aktivierung und Muskelspannung kontrollieren die Larynx, die Stellung der Kehlkopfknorpel sowie die Spannung und Elastizität der Stimmlippen, während subglottaler Luftdruck und aerodynamische Prozesse für die eigentliche Phonation, das schnelle und regelmäßige Schwingen der Stimmlippen, verantwortlich sind.

Um in Schwingung geraten zu können, müssen die Stimmlippen durch die Ary-Knorpel in die geschlossene Position gebracht (adduziert) werden. Die muskulären und elastischen Kräfte der Stimmlippen und die durch die Glottis fließende Luft wirken in Form eines Kreisprozesses (siehe Abbildung 5). Zu Beginn dieses Prozesses ist die Glottis geschlossen und die Stimmlippen gespannt. Der subglottale Luftdruck bewirkt, dass dieser glottale Verschluss gesprengt wird. Luft beginnt durch den Spalt zu fließen. Die Verengung der Durchflussöffnung an der Glottis führt mit der damit verbundenen Erhöhung der Fließgeschwindigkeit der ausströmenden Luft zu den senkrecht zur Fließrichtung wirkenden sogenannten Bernoulli-Kräften. Diese Kräfte lassen über ihre Sogwirkung die elastischen Stimmlippen wiederum einen Verschluss bilden. Der gesamte Prozess startet somit wiederum von vorne (Pompino-Marschall, 1995).

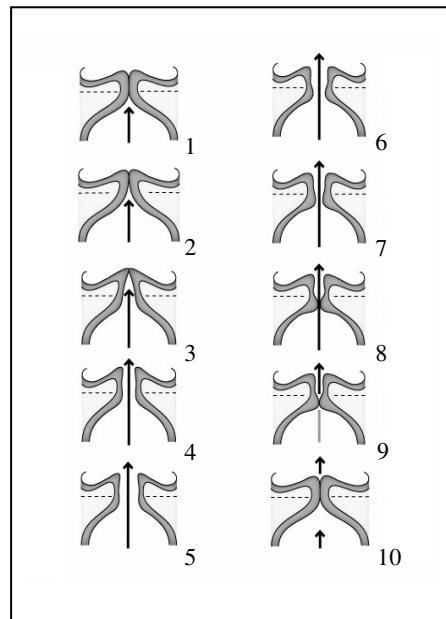


Abbildung 5: Zyklus der Stimmlippenschwingung

Die *Stimmtonfrequenz* ist abhängig von der Rate der Stimmlippenschwingung, die *Stimmqualität* u.a. von der Form dieser Schwingung. Die Länge der Stimmlippen ist für die Stimmtonfrequenz (Grundfrequenz) von grundlegender Bedeutung: Bei Frauen variiert die Stimmlippenlänge zwischen 13 und 17 mm, die mittlere Grundfrequenz liegt bei ca. 230Hz. Bei Männern variiert die Länge der Stimmlippen zwischen 17 und 24 mm, die Grundfrequenz beträgt im Mittel 120Hz. Weitere wesentliche Einflussfaktoren auf die Stimmtonfrequenz sind die muskuläre Einstellung der Stimmlippen und die Stärke des Ausatemdrucks.

Laryngale Stimmkontrolle: Grundfrequenz, Lautstärke und Stimmqualität. Neben der Kontrolle der suprasegmentalen Eigenschaften des Grundfrequenz- und Intensitätsverlaufs einer lautsprachlichen Äußerung und der Kontrolle der Stimmqualität kommen dem Kehlkopf auch eher als artikulatorisch zu bezeichnende Aufgaben zu: Eine der wichtigsten ist die Differenzierung zwischen stimmhaften und stimmlosen Lauten. Bei der Produktion von Sprachsignalen ist die Phonationsstellung, d.h. aneinanderliegende Stimmlippen, als die Normalstellung zu betrachten. Für die Artikulation stimmloser Laute hingegen muss die Glottis für die freie Luftpassage geöffnet werden. Ermöglicht wird dieses über eine ballistische Öffnungsbewegung der Stimmlippen, die durch die Kontraktion des Cricoaerytenoideus posterior bewirkt wird.

Zwei Stimmlippenparameter sind aus funktional phonetischer Sicht wichtig: die Längsspannung der Stimmlippen und der Abstand der Stimmlippen zueinander. So bewirkt ein durch die Aktivität des M. vocalis produziertes Kippen des Schildknorpels eine veränderte Längsspannung der Stimmlippen und eine veränderte Höhe des Stimmtons. Die Positionierung von Schild- und Stellknorpel wird durch die innere Kehlkopfmuskulatur reguliert. Im Wesentlichen wird die *Grundfrequenz (Tonhöhe)* der Stimme durch die Schwingungsrates der Stimmlippen bestimmt. Hierfür ist, wie bereits ausgeführt, die natürliche Länge der Stimmlippen von Bedeutung: kürzere Stimmlippen schwingen schneller als längere. Für die Variation der Stimmtongfrequenz sind eine Reihe von Determinanten ausschlaggebend: die Masse der schwingenden Stimmlippen, ihre Steife, die durch unterschiedlich starkes Gegeneinanderdrücken (mediale Kompression) veränderbare Länge ihres schwingungsfähigen Teils sowie die Stärke des subglottalen Drucks. Im Rahmen der schläfrigkeitsbedingten Stimmveränderungen besitzt dieser Sachverhalt eine besondere Bedeutung, da zum einen dünnere (und damit masseärmere) und steifere (und damit in ihrem Schwingungsbereich eingeschränkte) Stimmlippen und zum anderen aber auch durch einen höheren Druck angeregte Stimmlippen schneller schwingen und somit einen höheren Ton produzieren. Der Cricothyroid-Muskel besitzt durch die Streckung und damit Verdünnung und Masseverminderung der Stimmlippen, unterstützt durch die adduzierende Funktion des lateralen Cricoaerytenoids, eine zentrale Bedeutung. Vergleichbar hierzu ergibt sich auch die *Lautstärke* des Stimmtons aus dem Zusammenspiel zwischen subglottalen und glottalen Mechanismen und ist hauptsächlich durch die Höhe des Lungenluftdrucks (und damit des Luftdrucks unterhalb der Stimmlippen) determiniert.

Ausschlaggebend für die akustische Artikulation ist vor allem die Abruptheit, mit der der transglottale Luftstrom durch das glottale Schließverhalten abgeschnitten wird. Letzteres wird auch durch die Steife und geometrische Einstellung der Stimmlippen bestimmt. Durch das veränderbare Unterbrechen des Luftstroms entstehen unterschiedlich scharf abgegrenzte akustische Impulse. Bei einer lauten Stimmgebung treten schärfere und spektral ausgeprägtere Impulse auf, die durch eine schnellere Verschlussphase und eine damit insgesamt verkürzte offene Phase von 30 – 50% des Schwingzyklus gegenüber 50 – 70% bei normaler moderater Lautstärke charakterisiert sind (Pompino-Marschall, 1995).

Der Abstand der Stimmlippen hingegen determiniert die *Stimmqualität*. So wird die Stimme als „knarrend“ oder „gepresst“ bezeichnet, wenn starke Verschlusskräfte senkrecht zur Stimmritze wirken. Die Stimmqualität ist somit von der Form des glottalen Schwingungsverhaltens abhängig. Sind bereits vor Einsetzen der Schwingung der Stimmlippen die Stimmlippen leicht geöffnet, ergibt sich die Stimmqualität „behaucht“. Eine ‘behauchte’ Stimme wird demnach verursacht durch unmodulierte Atemluft aufgrund eines mangelhaften Stimmlippenverschlusses (Seidner & Wendler, 1982), wodurch ein geräuschbehaftetes, spektral flaches und intensitätsschwaches Rohschallsignal entsteht. Sind die Stimmlippen jedoch zu Beginn zu weit geöffnet sein, können keine Stimmlippenschwingung mehr auftreten. Diese An- bzw. Abwesenheit von Stimmlippenschwingungen unterscheidet genau die Lautkategorien *stimmhafte und stimmlose Laute* (stimmlos vs. stimmhaft: z.B. [k] vs. [g], [t] vs. [d] und [p] vs. [b]). So ist im Fall der stimmlosen Laute eine weite Öffnung der Stimmritze gegeben. Eine ‘rauhe Stimme’ entsteht durch Unregelmäßigkeiten im Stimmlippenschwingungsverhalten, also durch Frequenz- und/oder Amplitudenschwankungen, verbunden mit starker laryngealer Gespanntheit (Vieregge, 1989). Damit verbundene akustische Auswirkungen sind ein variabler Intensitätsverlauf (Shimmer) bzw. Grundfrequenzverlauf (Jitter). Die ‘heisere Stimme’ ist durch das gemeinsame Auftreten der Merkmale der behauchten wie der rauhen Stimme gekennzeichnet. ‘Pressen’ wird durch Überfunktionen im Glottisbereich verursacht, ‘Knarren’ kommt durch Irregularitäten bei den Stimmlippenschwingungen zustande oder durch tieffrequente Stimmlippenschwingungen, bei denen nur ein Teil der Glottis schwingt. Im Gegensatz zu den beschriebenen Stimmqualitäten wird eine modale (normale) Stimme bei locker verschlossenen Stimmlippen produziert (Seidner & Wendler, 1982; Vieregge, 1989).

Artikulation und Akustogenese

Bei der Artikulation wird der im Kehlkopf erzeugte Rohschall durch die Veränderung der Ansatzrohrgeometrie modifiziert, wodurch die für die lautsprachliche Kommunikation notwendigen relevanten lautlichen Differenzierungen erzeugt werden. Als *Ansatzrohr* wird die Gesamtheit der Hohlräume oberhalb des Kehlkopfes bezeichnet, in denen die Artikulation realisiert wird (Pétursson & Neppert, 2002, S. 76). Das Ansatzrohr umfasst anatomisch den von Mund- bzw. Nasenöffnung begrenzten Luftraum oberhalb des Kehlkopfs bzw. der Glottis, der aus dem Rachen-, Mund- und Nasenraum gebildet wird. Er ist mit Ausnahme der Zähne mit Schleimhaut und weichem Gewebe ausgekleidet.

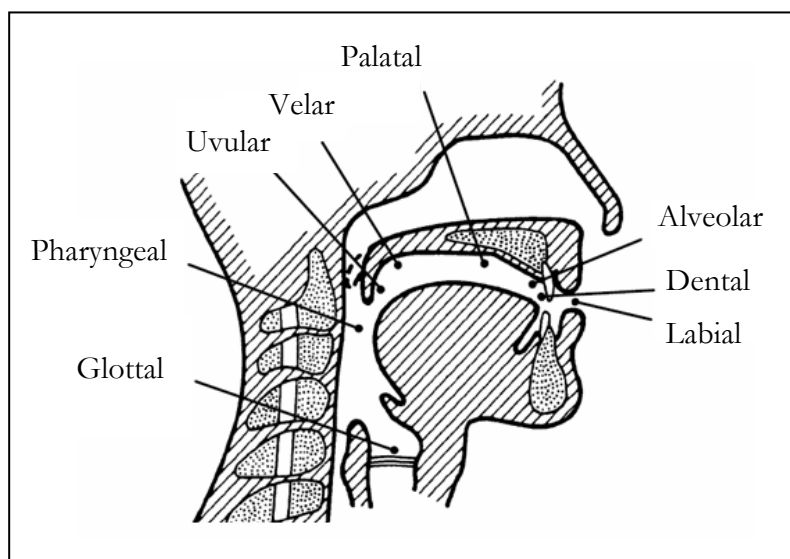


Abbildung 6: Schematische Darstellung der wesentlichen Artikulationsorte (nach Hess, 2002b)

Die *Artikulatoren*, d.h. die beweglichen Teile des Ansatzrohres, bestehen aus Zunge (Lingua), Lippen (Labia), Unterkiefer (Mandibulum), Gaumensegel (Velum) mit dem Zäpfchen (Uvula) sowie in eingeschränkter Form dem Rachen (Pharynx) und der Glottis. Beschrieben werden können diese Artikulatoren z.B. in ihrer horizontalen Zungenlage, vertikalen Zungenhöhe, dem Zungenwinkel, dem Zungenspitzenwinkel, dem Grad der Lippenvorstülpung oder Lippenspreizung, der relativen Lippenhöhe, dem Unterkieferöffnungswinkel, der Velumhöhe und der horizontalen und vertikalen Hyoid Location (Kröger, 2007). Die momentane Stellung der Artikulatoren determiniert den Querschnittsverlauf und somit die Filterfunktion des Vokaltrakts (vgl. Abbildung 7). Bei Veränderung dieser Artikulatorenstellung resultiert eine ebenfalls jeweils charakteristische, zeitlich variierende Filterwirkung. Beschreibungskategorien der Artikulatoren-Bewegung sind Amplitude, Dauer, Glattheit (Smoothness) und Fluidität. Veränderte Artikulatoren-Bewegungsmuster sollten sich daher in einer entspre-

chenden charakteristischen Veränderungen der Transferfunktion und somit auch Veränderung des resultierenden Stimmsignals widerspiegeln.

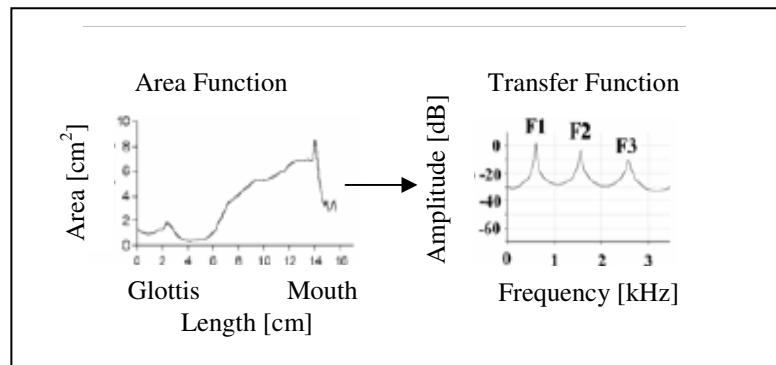


Abbildung 7: Zusammenhang der Vokaltraktquerschnittsfunktion und der hieraus resultierenden Filterwirkung (Transferfunktion) (nach Kröger, 2007)

Das Ansatzrohr kann zusätzlich auch mittels der durch die extrinsische Kehlkopfmuskulatur kontrollierten Kehlkopfhöhe verändert werden. Die Bewegung der Artikulatoren, der artikulierenden Organe, werden in der deskriptiven Phonetik in Bezug zu relativ fixen Strukturen des Ansatzrohres, den sogenannten *Artikulationsstellen*, beschrieben, die in Abbildung 6 dargestellt sind. Die Artikulationsstellen von außen nach innen sind gegeben durch (Pompino-Marschall, 1995): die Oberlippe (*labial*): z.B. [p], [b], [m] die oberen Schneidezähne: (*interdental*): [f], [v] den Zahndamm, Alveolen (*alveolar*): [t], [d], [n], [s], [z], [l] den harten Gaumen, Palatum (*palatal*): [ç], [j] den weichen Gaumen, Velum (*velar*): [k], [g], [ŋ], [x] das Zäpfchen, Uvula (*uvular*): [R] die Rachenwand, Pharynx (*pharyngal*): [ħ] die Glottis (*glottal, laryngal*) (vgl. Abbildung 7 und 8).

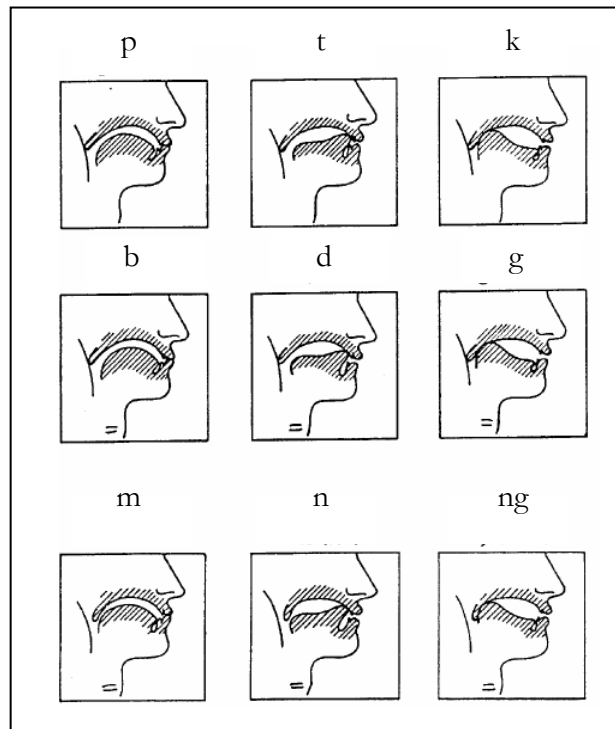


Abbildung 8: Schematische Darstellung der Artikulationsgesten der englischen Plosive und Nasale (*General American*) (aus Hess, 2002b)

Durch die Artikulatoren werden geometrische Veränderungen des Ansatzrohres bewirkt und verschiedene *Artikulationsmodi* realisiert. Sie bestehen in:

- globalen Formveränderungen bezüglich der Länge (durch Kehlkopfhebung/-senkung) und Lippenvorstülpung (-spreizung) und des Querschnittverlaufs (je nach Grad der Kieferöffnung und der vertikalen Zungenhöhe) sowie der horizontalen Zungenlage (vokalischer Artikulationsmodus)
- lokalen Engebildungen (Artikulationsmodus Frikativ: z.B. [s] mit alveolarer Engebildung) bzw. Verschlussbildungen (total bei Artikulationsmodus Plosiv: z.B. [t] mit alveolarem Verschluss)
- der Zuschaltung (Artikulationsmodus Nasal) bzw. Abschaltung des Nasenraums (durch abgesenktes bzw. gehobenes Velum: z.B. bei [n] mit oralem alveolarem Verschluss und gesenktem Velum).

Quelle-Filter-Theorie

Das Produkt der Artikulation ist der Sprachschall. Den für das Funktionieren der lautsprachlichen Kommunikation wichtigsten Aspekt der Sprachschallproduktion bildet die Modifikation des Rohschalls durch die Filterwirkung des Ansatzrohres im Sinne der Quelle-Filter-Theorie nach Fant (1960). Die Quelle-Filter-Theorie der Sprachproduktion beinhaltet die Vorstellung, dass das von den Stimmlippen produzierte Quellsignal im Sprechtrakt gefiltert wird und dadurch seine Formung und spezifische Klangfarbe erhält (Greisbach, 2003). Das Anregungssignal kann entweder über einen Impulsgenerator (stimmhafter Ton) oder aber über einen Rauschgenerator (stimmloser Ton) generiert werden (siehe Abbildung 9). Das Quellen- oder Primärsignal (Primärschwingung) regt im Ansatzrohr die Luftmoleküle im Vokaltrakt zum Mitschwingen an (Sekundärschwingung). Bei der Übertragung der Schallwellen im Ansatzrohr fungiert das Ansatzrohr sowie weitere Parameter des Artikulationstrakts in Abhängigkeit von ihrer jeweiligen geometrischen Form und ihrer momentanen Position (vgl. Abbildung 9) als variabler Filter bzw. Resonator. Die Theorie postuliert insbesondere für Vokale die Unabhängigkeit von Schallquelle (Stimmlippenschwingungen bei Vokalen) und Filter (Resonanzeigenschaft der Luftsäule oberhalb der Schallquelle bis zu den Lippen) sowie die Möglichkeit einer Zerlegung des aufgezeichneten Sprachsignals in diese beiden Komponenten (Greisbach, 2003). Das folgende Blockschaltbild illustriert die oben beschriebenen Anregungs- und Filterphänomene der Sprachgenerierung (vgl. Fellbaum, 1984; Schuller, 2006).

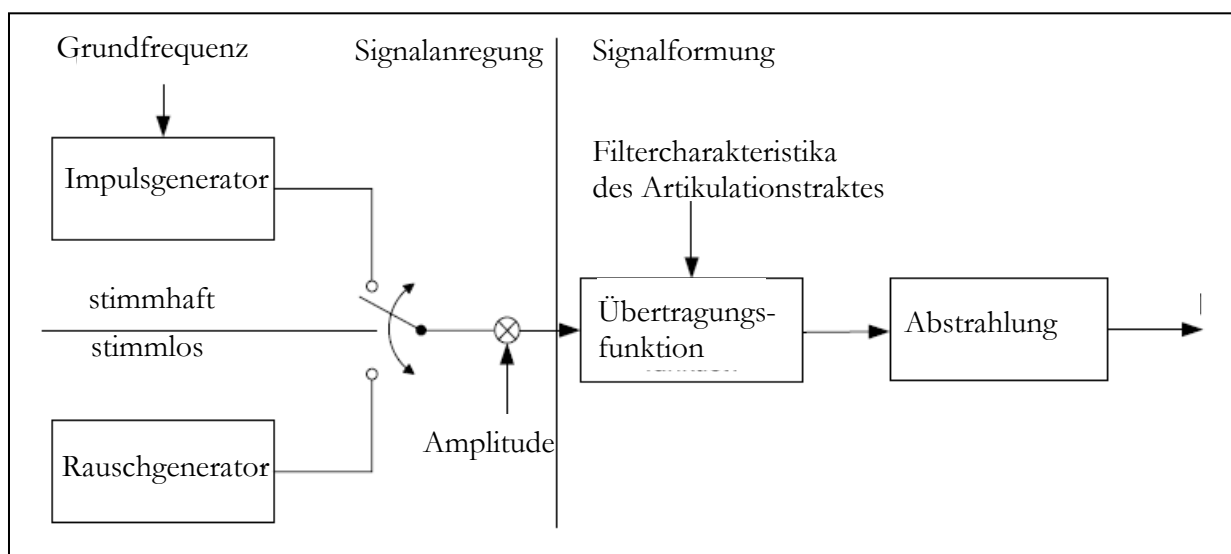


Abbildung 9: Sprachgenerierung als zweistufiger Prozess von Anregung (stimmhaft vs. stimmlos) und Formung (Filterung des Signals durch Parameter des Artikulationstraktes) (nach Schuller, 2006)

Filterfunktion des Ansatzrohres, Vokal- und Formantenenstehung. Der menschliche Vokaltrakt von der Glottis bis zur Mundöffnung kann bei der Vokalerzeugung in erster Näherung als Rohr mit vollständig reflektierenden Wänden betrachtet werden. Am geschlossenen Ende des Ansatzrohres befindet sich als Schalldruckquelle die Glottis, das obere Ende bildet die Mundöffnung. Funktionell gesehen ist das Ansatzrohr ein schwingungsfähiger Hohlkörper mit mehreren Eigenfrequenzen (Resonanzfrequenzen) (Friedrich et al., 2000). Frequenzen, die im Bereich der Resonanzfrequenzen liegen, werden im Ansatzrohr besonders gut übertragen bzw. verstärkt, während dazwischen liegende Frequenzen abgeschwächt oder ausgelöscht werden. Auf diese Weise wird der Rohrschall gefiltert, und es entsteht der spezifische Klang der Stimme. Nach Ungeheuer (1962) ist die Form des Querschnitts des Ansatzrohres akustisch nicht relevant; lediglich die Querschnittsfläche ist entscheidend. Daher wird nun ein 'neutrales' Rohr mit kreisrundem Querschnitt angenommen. Die Glottis erzeugt im Idealfall eine sich kugelförmig ausbreitende Schalldruckwelle. Diese wird an den schallharten (d.h. vollständig reflektierenden) Wänden des Rohres reflektiert. Die Reflektionen überlagern sich additiv, da es sich um ein lineares System handelt. Am verschlossenen Rohrende ist aufgrund der totalen Reflexion der Schalldruck maximal, am offenen Ende hingegen herrscht der normale Umgebungsluftdruck.

Die erste Welle, die diesen Randbedingungen genügt, ist ein Cosinus mit einer Wellenlänge von viermal der Rohrlänge L , die zweite Resonanzfrequenz passt bezüglich ihrer Wellenlänge zu drei Viertel in das Rohr (mit einem weiteren Knoten), die dritte schließlich besitzt eine Wellenlänge von vier Fünftel der Rohrlänge (d.h. sie passt in dieses mit insgesamt drei Knoten eineinviertel Mal). Aufgrund der im einseitig geschlossenen Rohr vorliegenden Randbedingung bezüglich der Lage der Maxima (geschlossenes Ende) und Knoten (offenes Ende) ergeben sich als stehende Wellen jeweils die ungradzahligen Vielfachen der Schwingung, die eine Wellenlänge von viermal der Rohrlänge hat. Der Stimmklang ist ein aus periodischen Einzelschwingungen (Teilschwingungen) zusammengesetzter komplexer Schall. Durch Überlagerung von Teilschwingungen, deren Frequenzen im ganzzahligen Verhältnis zueinander stehen, entsteht bei Vokalen das harmonische Sinusschwingungsgemisch, der Primärklang. Durchläuft der primäre Kehlkopfklang (Primärklang) nun ohne Behinderung, d.h. ohne Geräuschbildung, das Ansatzrohr, filtert dieser allein das Signal. Als Sprachsignal resultiert in diesem Fall die für Vokale (Öffnungslaute) charakteristische spektrale Verteilung.

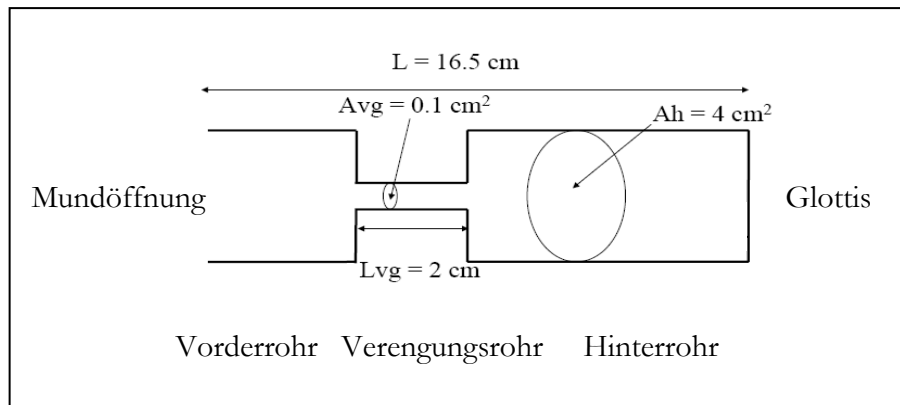


Abbildung 10: Akustisches Drei-Rohrmodell des Vokaltrakts (nach Harrington, 2007)

Eine weitergehende Verfeinerung des Vokaltraktmodells ergibt sich bei Berücksichtigung der Verengungsstelle im Mundraum (siehe Abbildung 10 und Abbildung 11). Aufgrund dieser Verengungsstelle wird der Vokaltrakt auch mit einem Drei-Rohrmodell angenähert. Entsprechend ergeben sich ein Hinterrohr mit geschlossenen Enden, ein Verengungsrohr mit zwei offenen Enden und ein Vorderrohr mit je einem geschlossenen und einem offenen Ende.

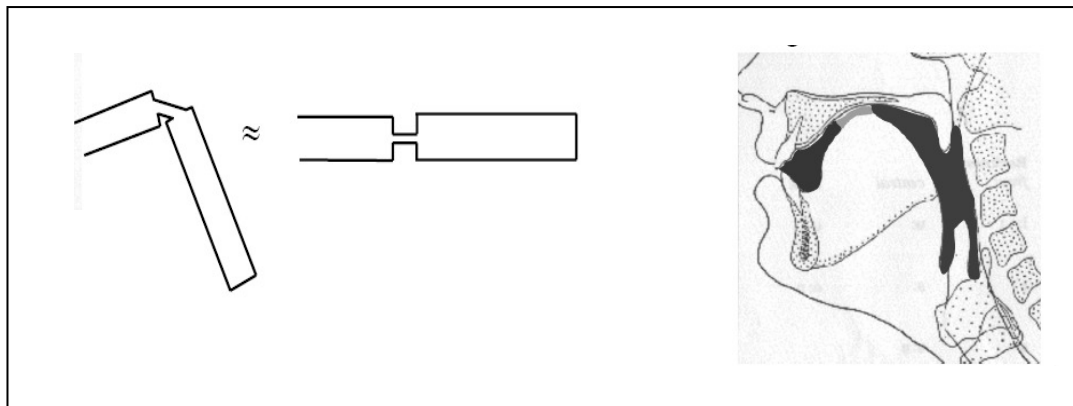


Abbildung 11: Akustisches Drei-Rohrmodell des Vokaltrakts (nach Harrington, 2007)

In Abhängigkeit von Artikulationsbewegungen wird die geometrische Konfiguration der Rohrkonfiguration (oder vereinfacht: des Ansatzrohres) verändert, wodurch sich die Resonanzfrequenzen verschieben. Folge ist eine veränderte Frequenzlage der vokalspezifischen Teiltonmaxima, der Formanten, die im Resonanzbereich des Ansatzrohres liegen. Die Formantstruktur der einzelnen Vokale ist im Wesentlichen unabhängig von der Tonhöhe und sonstigen Eigenschaften des primären Kehlkopfklanges. Die Lage der ersten und zweiten Formanten ist entscheidend für die Vokalerkennung, die weiteren Formanten charakterisieren eher individuelle Sprechermerkmale. Der Zusammenhang zwischen Artikulation und Fre-

quenzlage bzw. Frequenzänderung der Vokalformanten kann durch die in Tabelle 6 berichteten vier Regeln schematisiert werden (Pétursson & Neppert, 2002.). Die wichtigsten Zusammenhänge sind hierbei: die Lage des ersten Formanten hängt vom Abstand Kehlkopf-Zungenenge ab; die Lage des zweiten Formanten vom Abstand Zungenenge-Mundöffnung, wobei die Formantfrequenz tiefer liegt, wenn dieser Abstand größer wird.

Tabelle 6: Vokaltraktformänderungen und ihre Wirkung auf Formantfrequenzlagen

Vokaltraktform Veränderung	Verschiebung von Formantfrequenzlagen
Länge-Regel	Je länger das Ansatzrohr, desto niedriger die Formantenmittenfrequenz.
Verengung-im- Ansatzrohr-Regel	Verengung in der vorderen Hälfte des Ansatzrohrs: Frequenz von F_1 sinkt. Je größer der Grad der Engebildung, umso niedriger die Frequenz von F_1 . Verengung im pharyngalen Teil: Frequenz von F_1 steigt. Je größer der Grad der Engebildung, desto höher wird die Frequenz von F_1 . Verengung im mittleren bis vorderen Teil des Mundhohlraumes: Frequenz von F_2 steigt. Verengung im hinteren Teil des Mundhohlraumes (unter dem hinteren Gaumen): Frequenz von F_2 sinkt.
Lippenrundungs- Regel	Verengende Lippenrundung (meist mit Vorstülpung verknüpft): Alle Formant-Frequenzen sinken. Je größer der Grad der Lippenrundung, desto größer der Einfluss auf die Formanten. Der Einfluss der Lippenrundung ist größer auf F_2 und F_3 als auf F_1 .
Kieferwinkel- öffnungs-Regel	Je größer der Kieferwinkel (Abstand zwischen oberen und unteren Schneidezähnen), desto höher liegt die Frequenz von F_1 .

Konsonantenentstehung. Das konsonantische Artikulationsprinzip besteht in der Bildung von Hemmstellen für den Luftstrom an verschiedenen Stellen des Ansatzrohres. Die artikulatorische Beschreibung von Konsonanten erfolgt in einem zweidimensionalen Schema nach Art und Ort dieser Luftstrombehinderung. So werden nach der *Artikulationsart* eine mediale oder laterale Engebildung (Approximanten, Frikative), eine vollständige Verschlussbildung (Plosive), eine Umlenkung des Luftstromes in den Nasenraum (Nasale) sowie ein intermittierender Verschluss (Vibranten) unterschieden. Als Artikulationsorte werden u.a. bilabial (an Unter- und Oberlippe gebildet, z.B. [p], [b], [m] in „pellen“, „bellen“, „melden“), labiodental (an

Unterlippe und oberer Zahnreihe gebildet, z.B. [f], [v] in „falten“, „walten), alveolar (am Zahndamm gebildet, z.B. [t], [d], [n], [s], [z], [l] in „Tank“, „Dank“, „Name“, „reißen“, „reisen“, „Lamm“), postalveolar (an Zahndamm bis hartem Gaumen gebildet, z.B. in „Schein“, „Genie“), palatal (am harten Gaumen gebildet, z.B. in „ich“, „jagen“), velar (am weichen Gaumen gebildet, z.B. [k], [x] in „Kohl“, „ach“), uvular (am Zäpfchen gebildet, z.B. gerolltes Zäpfchen-r in „Radio“) und glottal (an der Stimmritze gebildet, z.B. [h] in „Einer“, „Heiner“) unterschieden (Kroeger, 2007). Als *Artikulationsort* wird demnach die Stelle beschrieben, an der durch einen beweglichen Artikulator (wie Lippen oder Zunge) und einen relativ unbeweglichen Artikulator (wie Zähne, Gaumen oder Rachen) eine Hemmstelle gebildet wird (siehe Abbildung 6 und Abbildung 8). Wird der Luftstrom beispielsweise durch die beiden Lippen behindert (wie in den Phonen p, b, m, so spricht man von einem bilabialen Laut. Bilden Unterlippe und obere Schneidezähne eine Hemmstelle (wie in [f]), resultiert ein labiodentaler Laut. Akustisch gesehen werden an diesen Hemmstellen Luftturbulenzen und damit Geräusche erzeugt. Je nach Glottiskonfiguration kann zusätzlich der primäre Glottisklang (Primärklang) vorhanden sein; man spricht in diesem Fall von *stimmhaften Konsonanten*. Fehlt hingegen eine Stimmtonbeteiligung bzw. Klangbeimischung, spricht man von *stimmlosen Konsonanten* (Pétursson & Neppert, 2002).

2.2 Kognitive Sprachproduktionstheorien

Im vorhergehenden Teil wurden die anatomisch-physiologischen Grundlagen der Sprachproduktion erläutert. Das nachfolgende Unterkapitel beschäftigt Im Folgenden sollen kognitive Grundlagen der Sprachproduktion erklärt werden. Sprachproduktionstheorien beschäftigen sich mit der Frage, welche bewussten und unbewussten mentalen Prozesse von der Kommunikationsintention bis zur Artikulation stattfinden. Sprachproduktionsmodelle lassen sich in zwei Gruppen einteilen: *Stufenmodelle* und *Aktivationsausbreitungsmodelle*. Die Stufenmodelle der Sprachproduktion gehen von diskret-serieller Verarbeitung aus, wobei sich die autonomen Verarbeitungsstufen nicht überlappen, und keine Rückkopplung zwischen den Ebenen stattfindet. Informationen werden erst dann an die folgende Stufe weitergegeben, wenn die Prozesse auf der vorhergehenden Ebene abgeschlossen sind. Zu dieser Gruppe gehören die Modelle von Levelt (1989), Garret (1982) und Motley, Baars und Camden (1983). In *Aktivationsausbreitungsmodellen* (auch Kaskadenmodelle, vgl. Harley, 1993) wird hingegen eine direkte Rückkopplung zwischen Verarbeitungsebenen durch das sogenannte ‘backward spreading’ postuliert. Dazu gehören die Modelle von Harley (1984), Stemberger (1985), Dell (1986), MacKay (1987), Berg (1988) und Schade (1992). Beide Ansätze sind wissenschaftlich viel diskutiert, es konnte bisher allerdings keine Entscheidung zu Gunsten einer Klasse von Modellen getroffen werden, da beide Modelltypen genug Flexibilität aufweisen, sich neueren Befunden immer wieder anzupassen (vgl. Cutting & Ferreira, 1999; Dell & O’Seaghdha, 1991, 1992; Harley, 1993; Jenseniak & Schriefers, 1998; Levelt et al. 1991a, 1991b; Levelt, Roelofs & Meyer, 1999). Im Folgenden sollen beispielhaft Modelle beider Theorienrichtungen kurz vorgestellt werden.

Levelts Stufenmodell der Sprachproduktion

Das Sprachproduktionsmodell von Levelt et al. (1999) setzt sich aus drei Grundkomponenten – Konzeptuell-Semantisches System, Phonologisch-Phonetisches System und Artikulator System zusammen (siehe Abbildung 12). Nach Levelt (1989) finden auf der pragmatisch-konzeptuellen Ebene, *Konzeptualisierer* genannt, kontrollierte, mit Aufmerksamkeit verknüpfte Prozesse statt. Im Konzeptualisierungsprozess wird ein Sprechplan in Form einer propositionalen Repräsentation erstellt. Dabei greift der Konzeptualisierer mit Hilfe des Arbeitsgedächtnisses nach Baddeley (1986) auf das Langzeitgedächtnis zurück, in dem das Wissen über Situationen und Personen sowie allgemeines Weltwissen gespeichert sind. Die kommunikative Intention des Sprechers wird in eine vorsprachliche Botschaft umgesetzt, in der die zu verbalisierenden Konzepte als lexikale Konzepte - konzeptuelle Repräsentationen,

für die ein Eintrag im lexikalen Gedächtnis (mentalem Lexikon) vorliegt - spezifiziert sind. Output des Konzeptualisierers ist die präverbale Botschaft, die ihrerseits Input des Formulators ist. Die Botschaften an den Formulator werden von einer Kontrollinstanz (Monitor) innerhalb des Konzeptualisierers überwacht. Davon zeugen z.B. Selbstkorrekturen (Selbstmonitoring) während des Sprechens (Levelt, 1989). Im Formulierungsprozess wird dieser Sprechplan in eine adäquate sprachliche Form gebracht.

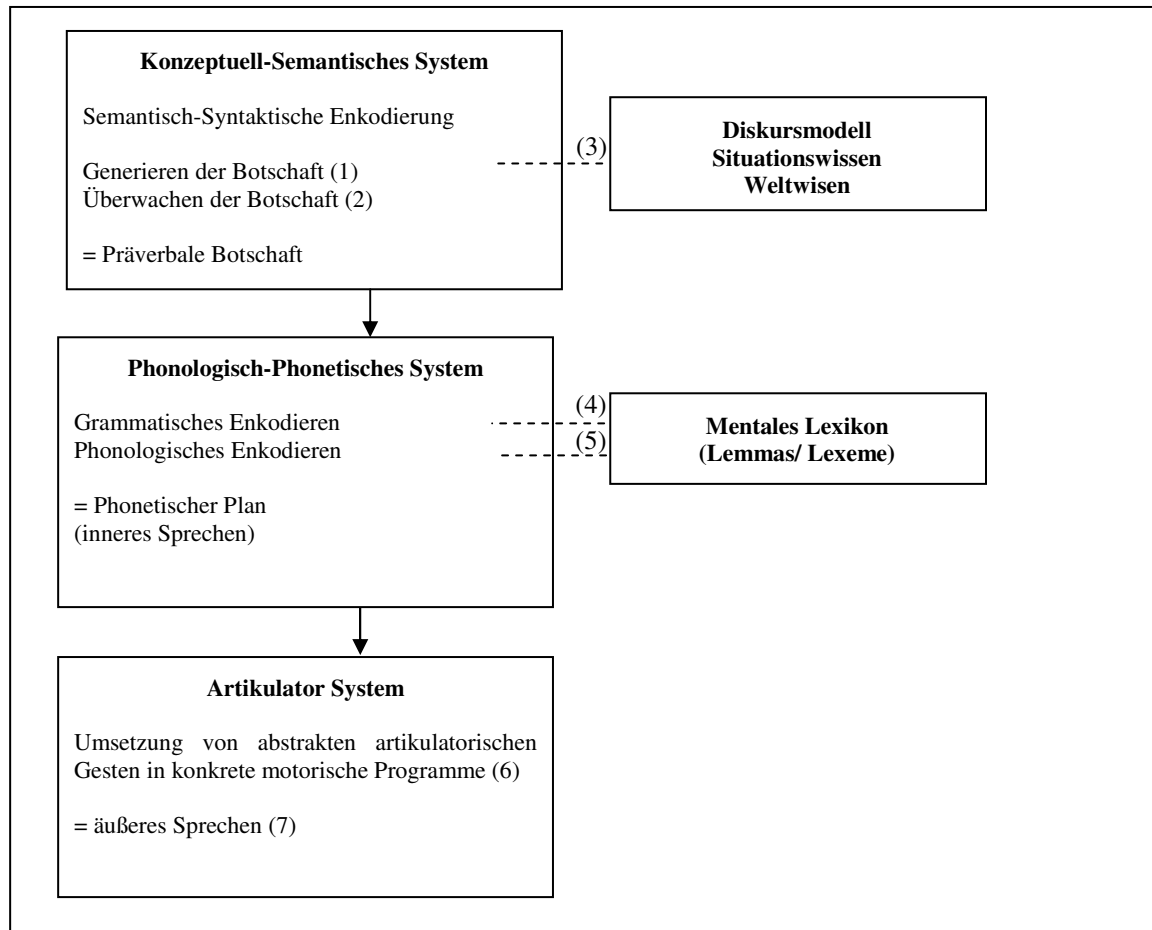


Abbildung 12: Modell der Sprachproduktion nach Levelt et al. (1999). Die aufgeführten Ziffern stellen die im weiteren Verlauf relevanten Bezüge zum Kognitiv-phonetischen Mediatormodell her (vgl. Abbildung 17)

Im Formulator werden die konzeptuellen Strukturen in linguistische übersetzt. Die Bearbeitungsprozesse im Formulator sowie im Artikulator finden anders als im Konzeptualisierer automatisch und nicht-intentional statt. Die erste Verarbeitungsphase des Formulierungsprozesses ist das grammatische Enkodieren. Unter Verwendung der Informationen des mentalen Lexikons werden Lemmas (Worteinträge im mentalen Lexikon, die semantische und syntaktische Informationen eines Wortes beinhalten) abgerufen. Hieraus werden Phrasen gebildet (Nominal-, Präpositional-, Verbalphrasen etc.) und zu einer Oberflächenstruktur zusammengesetzt. Im syntaktischen Puffer (wird in der Abbildung nicht dargestellt) können Zwischen-

ergebnisse gespeichert werden. Der Selektionsprozess auf der Lemmaebene muss abgeschlossen sein, bevor das phonologische Enkodieren beginnt (z.B. Butterworth, 1989; Fromkin, 1971; Garrett, 1980; Levelt, 1989; Levelt et al., 1999). Während der phonologischen Enkodierung wird auf Grundlage metrischer und segmentaler Informationen die generierte Oberflächenstruktur (die semantisch-syntaktische Repräsentation) der Äußerung in einen phonetischen Plan umgesetzt. Die metrische Information beinhaltet die Betonung der Silben und die Silbenanzahl der Morpheme, während die segmentale Information die spezifische Silbenstruktur (Positionen von Anfangs- und Endkonsonanten sowie von Vokalen) enthält. Die nachfolgende phonetische Enkodierung setzt diese Repräsentation in abstrakte artikulatorische Gesten um, die bei der Sprachproduktion erreicht werden müssen (Browman & Goldstein, 1992; Levelt, 1989). Diesen abstrakten artikulatorischen Gesten liegen jedoch noch keine konkreten Bewegungsabläufe zugrunde (Jescheniak, 2002). Nach Festlegung der artikulatorischen Gesten müssen schließlich auch Dauer, Amplitude und Grundfrequenz jeder Silbe bestimmt werden. Im Artikulator erfolgt die Umsetzung der abstrakten artikulatorischen Gesten in konkrete motorische Programme. Die motorische Ausführung durch die artikulatorischen Sprechwerkzeuge führt schließlich zum wahrnehmbaren Sprechergebnis. Benötigt die Artikulation eine längere Zeitspanne, kann sie im artikulatorischen Puffer zwischengespeichert werden (diese und andere Gedächtniskomponenten sind in der Abbildung 14 nicht dargestellt).

Schläfrigkeitsbedingte Störungen des Sprachproduktionsprozesses lassen sich wie in Abschnitt 3.1 näher dargestellt wird vielen Ebenen des Sprachproduktionsmodells verorten. So können z.B auf der Ebene der Phonologischen Enkodierung beim Zugriff auf die Lexem-Information phonologische Defizite oder auf der Ebene der phonetischen Enkodierung bei Zugriff auf sprechmotorische Programme phonetische Defizite erklären werden. Vergleichbare Beeinträchtigungen lassen sich auch entlang des Sprachverarbeitungs-Modells von Stackhouse und Wells (1997) festmachen. Auch hier werden sprechmotorische Programme erzeugt, die den phonetischen Plan in Artikulation umsetzen. Auf der Ebene des motorischen Programms können hier Lauten, die in der phonologischen Repräsentation korrekt abgespeichert sind, fehlerhafte „gestische Targets“ zugeordnet werden, d.h. Stellung und Bewegung der Artikulationsorgane werden fehlerhaft programmiert. Auf der Ebene des motorischen Planens könnten bei der Zusammensetzung der „gestischen Targets“ (enthalten Informationen über die Stellung und Bewegung der Artikulationsorgane) in die korrekte Abfolge Fehler auftreten und so phonetische Sprechfehler entstehen. Ein für die Vorhersage Schläfrigkeits-

bedingter Sprachveränderungen ebenso anwendbares Sprachproduktionsmodell ist das Modell der sensomotorischen Sprechkontrolle von Van der Merwe (1997). Van der Merwe (1997) postuliert fünf Phasen der Sprachproduktion: (a) kommunikative Intention, (b) linguistisch-symbolische Planungsphase (c) motorischen Planung, (d) motorische Programmierung und (e) motorische Ausführung. Auch entlang dieses Modells können auf allen Phasen schläfrigkeitsbedingte Beeinträchtigungen der Sprachproduktion beschrieben werden. Vertiefende Betrachtungen zur Sprachproduktionstheorien finden sich bei Kröger (1998).

2.3 Auditiv-perzeptive Beschreibungskategorien

Nachdem in Abschnitt 2.1 und 2.2 die phonetischen und kognitiven Grundlagen der Sprachproduktion beschrieben wurden, werden nun phonetische Stimmkategorien vorgestellt, mit denen ein wahrgenommener Höreindruck beschrieben werden kann. Als auditiv-perzeptive Beschreibungskategorien der Sprache dienen die Klassen (a) Prosodie, (b) Artikulation und (c) Stimmqualität.

Prosodie

Unter Prosodie fallen suprasegmentale (über mehrere Einzellaute verlaufende) Stimmphänomene wie Lautstärke, mittlere Sprechstimmlage, Sprechmelodie, Betonungswechsel, Sprechrhythmus, Sprechtempo und Sprechpausen (Pétursson & Neppert, 1996). Dabei werden in der Regel drei Hauptgruppen unterteilt: *Intonation*, *Intensität* und *Dauer*. Anwendung finden können diese Hauptkategorien mit Einschränkungen jedoch auch auf alle sprachlichen Einheiten: Phon, Silbe, Wortteil (Stamm, Präfix, Suffix), Wort, Wortgruppe (Satzglied, Phrase), Satz, Satzgruppe. In den folgenden Abschnitten wird auf die oben genannten prosodischen Merkmale näher eingegangen.

Intonation. In der Phonetik wird der Verlauf der Tonhöhe innerhalb eines Sprechaktes als Intonation oder *Satzmelodie* bezeichnet. Im Deutschen oder Englischen wird diese syntaktisch etwa zur Andeutung von Fragen benutzt. Darüber hinaus spielt sie eine entscheidende Rolle für den Ausdruck von Emotion. Beim sachlichen Sprechen Erwachsener bewegt sich die Tonhöhe selten über eine Oktave hinaus, meist innerhalb einer Quinte oder Sexte (Pahn, 1988). Als *Sprechstimmlage* wird die wahrgenommene durchschnittliche Sprechtonhöhe bezeichnet, um die sich die Tonhöhenvariationen bewegen (Vieregge & Pahn, 1996). Der Tonhöhenverlauf kann hinsichtlich seiner zeitlichen Variabilität sowie hinsichtlich seines Umfangs und seiner Form beschrieben werden.

Die *Grundfrequenz* (auch *Fundamental Frequenz*, F_0) ist das akustische Korrelat zur wahrgenommenen Tonhöhe einer sprachlichen Äußerung. Zwischen der Grundfrequenz, die in der Regel in Schwingungen pro Sekunde (Hz) angegeben wird und der Hörempfindung besteht kein linearer Zusammenhang (Zwicker & Fastl, 1990). So löst das menschliche Gehör durch die logarithmische Frequenzwahrnehmung etwa tiefere Frequenzen feiner auf als höhere Frequenzen. Ferner kann die subjektive Tonhöhenwahrnehmung auch durch die Lautstärke und Dauer beeinflusst werden. Die komplexe Natur einer Verbindung diverser Frequenzanteile in

der menschlichen Stimme begünstigt weiterhin die Divergenz zwischen wahrgenommener Tonhöhe und physikalischer Grundfrequenz. Darüber hinaus kann auch ein Eindruck der Tonhöhenwahrnehmung in unharmonischen Signalen entstehen, wie etwa bandbeschränktem Rauschen (Zwicker & Fastl, 1990). Dies beruht auf der frequenzanalytischen Arbeitsweise des Gehörs: Der größte gemeinsame Teiler einer Frequenzmenge wird vom Menschen als Grundfrequenz empfunden. Um eine Grundfrequenz wahrzunehmen, gelten 30 ms generell als ausreichend, wobei eine größere Messungen auch bei kürzerer Dauer möglich ist.

Intensität. Das physikalische Gegenstück zur wahrgenommenen Lautstärkeempfindung (Lautheit) ist der Schalldruck. Sie wächst mit der physikalisch messbaren Schall-Amplitude, allerdings besteht auch hier ähnlich wie beim Zusammenhang von Fundamentalfrequenz und Tonhöhenwahrnehmung kein linearer Zusammenhang (Neppert, 1999). So liegt eine Reihe komplexer und umstrittener psychoakustischer Zusammenhänge zwischen der Lautstärkeempfindung eines Tones und seiner Frequenz und Dauer vor. Nach Vieregge & Pahn (1996) sind Lautheitswahrnehmungen nicht nur von der physikalisch messbaren Intensität abhängig, sondern darüber hinaus auch vom subjektiven Gesamteindruck der Sprechstimme, von der Dauer des Sprechereignisses, von den Vokalquantitäten und von der auditiv-visuell wahrnehmbaren Sprechanstrengung. Neben der Bestimmung der durchschnittlichen Intensität können Intensitätsveränderungen hinsichtlich ihrer Häufigkeit, ihres Umfangs und ihrer Form skaliert werden. Problematisch an intensitätsbezogenen Größen ist der Umstand, dass bei der Sprachaufzeichnung die Aufnahmerichtung und Distanz zum Mikrofon einen großen Einfluss auf die gemessene Intensität ausüben.

Dauer, Rhythmus, Pausen und Sprechfehler. Die Dauer stimmhafter Laute kann in einer Näherung aus der Länge stimmhafter Abschnitte, die Dauer von Pausen hingegen aus dem Verlauf der Signalenergie geschätzt werden. Die Sprechgeschwindigkeit wird als Realisierung bestimmter sprachlicher Einheiten (Laute/Silben/Wörter) pro Zeiteinheit angegeben (Bose, 1994). Die auditive Wahrnehmung beziehungsweise Beurteilung des Gesprochenen hängt aber auch noch von anderen Faktoren ab, wie der bereits genannten Pausenlänge, der Pausenhäufigkeit und -position, der Akzentuierungsstärke, der Artikulationspräzision und der Lautdehnung. Die Sprechgeschwindigkeit in spontaner Sprache variiert deutlich und schwankt kleinräumig. Sie ist unter anderem abhängig vom Sprechplanungsaufwand, von der Sprechgliederung und der Sprechausdrucksgestaltung (Bose, 2001). Neben der durchschnittlichen Geschwindigkeit kann die Geschwindigkeitsveränderung hinsichtlich ihrer Häufigkeit,

ihres Umfangs und ihrer Form beurteilt werden. Das Zusammenspiel von Intonation, Intensität und Sprechgeschwindigkeit spielt eine entscheidende Rolle für die *Betonung*. Generell sind Merkmale einer starken Betonung erhöhte Lautheit, gehobene Sprachgrundfrequenz sowie eine gestreckte Silbenlänge. Die sogenannte Hauptbetonung einer Phrase weist aus semantischer Sicht auf ihr zentrales Element hin. Daneben existieren Wortbetonungen an der Stelle der meistbetonten Silbe. Die Positionen des Intensitäts- und Intonationsmaximums einer Phrase spielen daher auch für die Interpretation spontansprachlicher Äußerungen eine wichtige Rolle.

In engem Zusammenhang mit Pausen- und Rhythmusphänomenen in der Sprache stehen Redeflussstörungen. Die meisten Redefluss-Störungen haben eine 3-Phasen-Struktur: (1) Reparatur, d.h. das zu ersetzende Sprachmaterial: der Sprecher entdeckt den Fehler und stoppt den Redefluss, (2) Bearbeitungsphase/Editing Phase: gar keine, stille oder gefüllte Pause („uh“/ „um“) oder Einfügungen („I mean“), (3) Reparaturphase/Repair: Wiederaufnahme der flüssigen Rede mit ausgebessertem Sprachmaterial (vgl. Shriberg, 2001; Uhmann, 2001). Bei gefüllten Pausen werden z.B. Füll-Vokale („uh“/“um“ bzw. „äh“/“ähm“) eingefügt, die länger dauern als die gleichen Vokale in flüssiger Rede (vgl. Clark & Tree, 2002). Darüber hinaus sinkt die Intonation und es kommt oft zu Diplophonie (abwechselnd starke und schwache glottale Reize während der Phonation). Zur Illustration verschiedener Redeflussstörungen sei hier ein Auszug aus der Rede eines bayerischen Ministerpräsidenten herangezogen:

„Wenn Sie vom Hauptbahnhof in München [*ungefüllte Pause*] (...) mit zehn Minuten, ohne dass Sie am Flughafen noch einchecken müssen, dann starten Sie im Grunde genommen am Flughafen – [*Wiederholung*] am am Hauptbahnhof in München starten Sie ihren Flug. Zehn Minuten - schauen Sie sich mal die großen Flughäfen an - wenn Sie in Heathrow in London oder sonst wo, meine s..., [*Fehlstart/ Tilgung*] Charles de Gaulle äh [*Längung einer Silbe*] in Frankreich oder in [*gefüllte Pause*] äh, in in äh in äh Rom - wenn Sie sich mal die Entfernungen ansehen, wenn Sie Frankfurt sich ansehen, dann.....“

Diese Redeflussstörungen stellen Phänomene dar, die auch mit wachsender Schläfrigkeit zusammenhängen könnten. Je nachdem, ob es sich dann um einen „Wiederholer“ Sprechertypus (langsame Sprecher, mehr Wiederholungen und Zögerphasen, weniger Fehlstarts) oder einen „Tilger“ Sprechertypus (schnelle Sprecher, mehr Tilgungen, Fehlstarts und Neuanfänge, weniger Zögerphasen) handelt, können mit wachsender Schläfrigkeit unterschiedliche Redeflussstörungen prognostiziert werden.

Artikulation

Artikulationsbezogene Beschreibungskategorien werden auf segmentaler Ebene betrachtet und beinhalten die Komponenten: Lippen- und Kieferbewegung (eingeschränkt vs. übermäßig), Zahnreihenabstand, Zungenbewegung (Vor- und Rückverlagerung der Zunge) und Lautbildung(sfehler) (verwaschen, unverständlich, Zusammenziehen von Worten). Im Folgenden werden Artikulationsphänomene wie Assimilation, Elision vs. Epenthese, Lenisierung vs. Fortisierung und die Zentralisierung beschrieben.

Die *Artikulationspräzision* bezeichnet den Sorgfältigkeitsgrad bei der Ausspracherealisierung. Bei deutlicher und präziser Artikulation werden alle Artikulationsorte erreicht und alle Artikulationsgesten (Gestures) exakt durchgeführt. Gestures sind Konstriktionseinheiten, die im dynamischen Zusammenhang betrachtet werden und die koordinierte Bewegung mehrerer auf ein linguistisches Ziel ausgerichteter Artikulatoren (z.B. labiale Verschlussbildung) umfassen. Diese Gestures sind zusammensetzbar und bilden größere sprachliche Einheiten wie Segmente und Silben (Harrington, 2007). Als Fehlertypen können des Weiteren unterschieden werden die „Gestural Intrusion Errors“ und die „Reduction Errors“. Die Gestural Intrusion Errors umfassen das Einfügen einer Konstriktionsgeste, die zu diesem Zeitpunkt gewöhnlich nicht produziert wird, wie z.B. beim Sprechen von „cop top“ die zusätzliche Konstriktion am Dorsum während der Produktion des [t]). Die Reduction Errors beinhalten eine Reduktion der Intensität einer artikulatorischen Bewegung wie z.B. beim Sprechen von „cop“ die nicht komplette und nur partielle Konstriktion am Dorsum während der Produktion des [k]. Jenseits dieser artikulatorischen Feinprozesse versuchen rezente Entwicklungen, ein globales, automatisch quantifizierbares Maß der Verständlichkeit und Artikulationspräzision zu entwickeln, in dem automatische Spracherkennungssysteme und ihre Worterkennungsrate als Artikulationspräzisionsmaß verwendet werden (Maier et al., 2007).

Als *Assimilation* wird das Artikulationsphänomen der Angleichung von Segmenten bezeichnet. Unterschieden werden regressive vs. progressive Assimilation, d.h. Angleichung in bzw. gegen die Zeitrichtung, partielle (teilweise) vs. totale (vollständige) Assimilation und Kontakt- vs. Fernassimilation, d.h. Angleichung unmittelbar benachbarter vs. nicht unmittelbar benachbarter Segmente. Im Gegensatz zu diesem Angleichungsphänomen steht der Prozess der Dissimilation, der Segmente kontrastiert. Analog zur Assimilation wird auch hier unterschieden zwischen progressiver vs. regressiver Dissimilation sowie der Dissimilation unmittelbar benachbarter oder weiter auseinander liegender Segmente (Gussenhoven & Jacobs, 2003). Das Artikulationsphänomen der *Elision* beinhaltet die vollständige Unterdrückung von Segmenten (Spencer, 1996). Das betroffene Segment verschwindet im Rahmen der Elision vollständig, wodurch Raum für weitere phonologische Prozesse (wie die Assimilation) entsteht. Im Gegensatz hierzu beschreibt die Epenthese das Einfügen eines Vokals oder eines Konsonanten in eine lautliche Äußerung (Spencer, 1996). Sie ist in der Regel die Folge der Fehlsynchronisation zweier oder mehrerer Artikulationsgesten, die gleichzeitig ablaufen müssten.

Als weiteres Artikulationsphänomen ist die *Lenisierung* aufzuführen. Sie umfasst die fortschreitende Erweichung (Schwächung) eines Konsonanten, die bis zur Elision führen kann. Im Falle des Wechsels von stimmlosen zu stimmhaften Plosiven ([p] zu [b], [t] zu [d], [k] zu [g]) kann die *Voice Onset Time*, die den zeitlichen Abstand zwischen Burst und Phonationsbeginn des darauf folgenden Vokals bestimmt, als Maß der Lenisierung herangezogen werden. Die Fortisierung ist das Gegenstück zur Lenisierung. In den meisten Fällen handelt es sich um die Entstimmung stimmhafter Plosive und Frikative. Sie tritt häufig im Silbenauslaut in Form der Auslautverhärtung auf (Ternes, 1999). Denkbar ist eine zukünftige Anwendung von automatischen Spracherkennungssystemen zur Identifikation von Laut-bezogenen Auslassungen (Elisionen) und Einschüben (Epenthesen). Auch der generelle Grad der Dissimilierung im Vergleich zu einer normierten phonetischen Verschriftung eines neutral gesprochenen Wortes könnte mittels dieser Systeme geprüft werden. Dies ist jedoch für eine automatische Extraktion besonders schwierig, da es nicht nur eine automatische Erkennung von Sprache, sondern auch eine exakte phonetische Verschriftung im Falle von ausgelassenen oder eingeschobenen Lauten verlangt.

Als *Zentralisierung* von Vokalen (auch Vokalabschwächung) bezeichnet man die Verlagerung der Zungenlage in einen mittleren Bereich in Richtung des Zentralvokals [ə] (Schwa). Diese mittlere Position des Schwa-Vokals drückt sich auch in der zentralen Position des Schwas innerhalb des Vokaltrapezes aus. Eine Schwa-Reduktion zeichnet sich dementsprechend durch eine reduzierte erste Formant- und eine erhöhte zweite Formantfrequenz aus (siehe Abbildung 13), was graphisch im Einnehmen einer zentralen Position innerhalb des Vokaltrapezes entspricht. Eine weitergehende artikulatorische Aufwandsreduktion wird erzielt, wenn dieser Schwa-Vokal vollständig gelöscht wird (Schwa-Elision).

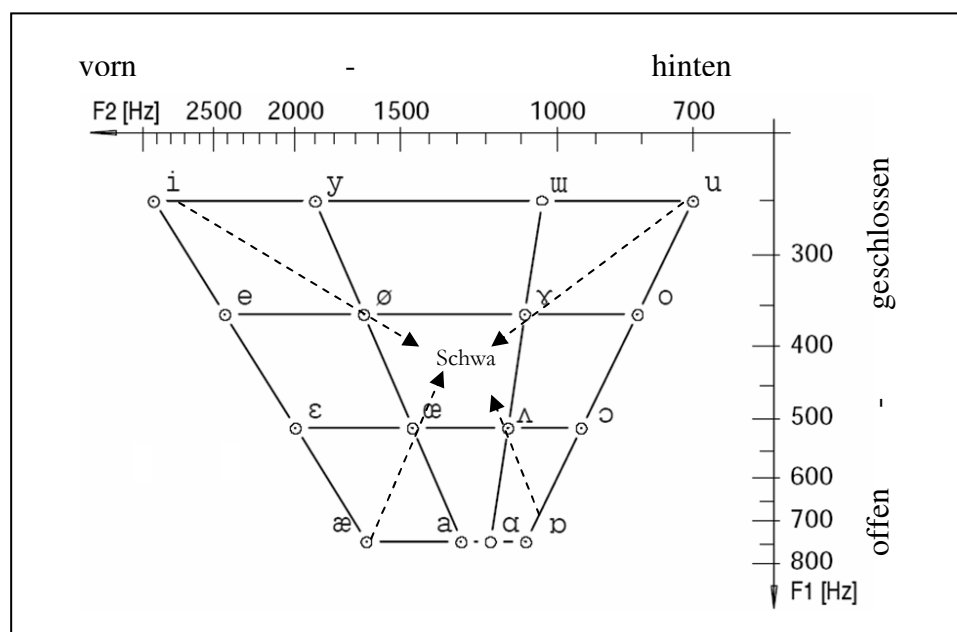


Abbildung 13: Akustischer und artikulatorischer Vokalraum. Darstellung der Zentralisierung von Vokalen (Schwa-Reduktion) im Vokaltrapez (vgl. Kohler, 1977, S. 75)

Stimmqualität

Stimmqualität wird über Merkmale des Stimmklangs (fest, klangvoll, wohlklingend, modulationsfähig, kräftig, dumpf), der Hyperfunktionalität (rau brüchig, knarrend, kratzend, gedrückt, gepresst, blechern, kloßig), der Hypofunktionalität (leise, matt, verhaucht, belegt, zittrig, klangarm) und der Stimmein- und -absätze (weich, fest, hart, verhaucht) beschrieben. Eine zusätzliche Randkategorie bildet die Sprech- und Ruheatmung (angestrengt, schnappen, ziehen vs. mühelos, fließend, ausgeglichen). Gebräuchlich ist auch die gröbere Differenzierung in modale, falsetto, whispery (flüsternde), breathy (behauchte), creaky (knarrende) und rough (raue) Stimmqualitäten (Laver, 1980).

Eine weitergehende Sprachproduktions-bezogene Betrachtung zeigt, dass der Stimmklang durch die Tätigkeit der Glottis, die Klangbildung und Resonanzwirkung im Ansatzrohr, die Artikulation und die Funktion der Atmung beeinflusst wird. Der modulierte Stimmklang kann hinsichtlich verschiedener Merkmale beurteilt werden: Das Merkmal *Klangfülle* ist physiologisch bedingt durch die Regelmäßigkeit im Schwingungsablauf der Stimmlippen und durch Resonanzeigenschaften des Ansatzrohres, die die Teiltonstruktur des Spektrums bestimmen. Als 'sehr klangvoll' wird in dieser Hinsicht eine resonanz- und auch teiltonreiche, ökonomisch, günstige Stimmgebung bezeichnet (Bose, 2001).

Das Stimmqualitäts-bezogene Merkmal *Klangfarbe* (auch Timbre) kann über die Kategorien „Helligkeit“ und „Schärfe“ beschrieben werden. Dunklere Klänge haben ihre stärkeren Teilschwingungen überwiegend im Bereich der niedrigen Frequenzen. Der spektrale Energieschwerpunkt hellerer Klänge hingegen liegt im Bereich hoher Frequenzen (Neppert, 1999). Die Klangfarbe oder das Timbre einer Stimme entspricht der individuellen Perzeption der Stimmqualität. Diese ist kein Träger linguistischer Information und unmittelbar mit dem emotionalen Charakter korreliert. Darüber hinaus hängt von ihr in erhöhtem Maße die Verständlichkeit des gesprochenen Inhalts einer Äußerung ab. Das Timbre ist maßgeblich vom Verhältnis der spektralen Bänder (Klasmeyer & Sendlmeier, 2000) und der Lage und Bandbreite der *Formanten* geprägt, die wiederum durch unterschiedliche Ansatzrohrformungen determiniert werden.

Weitere Stimmqualitäts-bezogene Kategorien wie die der *faukalen Distanz* werden über die Weite im Bereich der hinteren Gaumenbögen, die Lage der Zungenwurzel und die Lage des Kehlkopfes während der Stimmgebung bestimmt. Stimmein- und -ausschwingphasen wiederum sind abhängig von der Stimmlippenspannung und von der Stärke des subglottischen Drucks (Seidner & Wendler 1982). Als 'weich' werden Stimmein- und -absätze bei geringer Stimmlippenspannung und kleinem, allmählich wirkendem subglottischen Druck empfunden, so dass die Stimmlippen ohne wahrnehmbares Geräusch zu schwingen beginnen. Bei starker Stimmlippenspannung und großem, plötzlich wirkendem subglottischen Druckanstieg werden Stimmein- und -absätze dagegen als 'hart' empfunden. Ein weiteres mit dem perzeptiven Eindruck der Stimmqualität assoziiertes Merkmal ist der *Geräuschanteil* der Stimme. Er bezieht sich auf die Art der Stimmlippenschwingungen. Als klar wird ein Stimmklang empfunden, bei dem die Stimmlippen regelmäßig schließen und schwingen. Hörer mit ausgeprägtem auditivem Differenzierungsvermögen können verschiedene Ursachen von Geräuschhaftigkeit

näher bestimmen, zum Beispiel Hauchen, Pressen, Knarren, Rauigkeit. Unterschieden werden des Weiteren Kategorien der *Hyperfunktionalität* (rau, brüchig, knarrend, gepresst) und *Hypofunktionalität* (matt, verhaucht, belegt, zittrig).

Die Stimmqualitäts-bezogene Beschreibungskategorie des nasalen Stimmklangs (Nasalität) entsteht, wenn, wie im Fall der Nasale [m], [n], [ng] Luft während der Artikulation in den Nasenraum strömt. Laver (1980) unterscheidet zwei Arten von Nasalität, die 'Faukale Nasalität' und die 'Velopharyngeale Nasalität'. Die '*Faukale Nasalität*' ist verbunden mit einer faukalen, pharyngalen und laryngalen Muskelspannung. Diese Muskelkonstriktion tritt nach Scherer (1986, S. 151) als physiologische Reaktion auf einen unangenehmen Reiz auf. Die - im Verlauf unter der Perspektive der Schläfrigkeits-bezogenen Stimmveränderungen Bedeutung gewinnende - 'Velopharyngeale Nasalität' lässt sich auf ein abgesenktes Gaumensegel (Velum) in Folge einer allgemeinen oropharyngealen Entspannung zurückführen. Die Konsequenz dieser Entspannung ist eine sich weitende Öffnung der nasalen Passage (vgl. Laver, 1980, S. 77ff.).

2.4 Akustische Beschreibungsgrößen

Im nachfolgenden Abschnitt werden die wichtigsten zur Quantifizierung der zuvor beschriebenen audiv-perzeptiven phonetischen Phänomene eingesetzten Stimmsignalkennzahlen beschrieben. Mögliche Kategoriensysteme von akustischen Kennzahlen sind entweder von Signalverarbeitungskonzepten (*Time Domain, Frequency Domain und Phase Space Domain*) oder von phonetisch-linguistischen Konzepten abgeleitet. Fokussiert wird in diesem Kapitel auf ein phonetisch-linguistisches Kategoriensystem, das Stimmsignalkennzahlen, den Kategorien Prosodie, Artikulation und Stimmqualität zuordnet. Die Zuordnungen stellen jedoch in vielen Fällen nur eine lose Verknüpfung dar und erfüllen daher im Wesentlichen didaktische Funktionen. Darüber hinaus wird in Abschnitt 4.3 ein generelles Prozedere zu Generierung weiterer Stimmphänomene unabhängiger Stimmsignalkennzahlen dargestellt. Wie die Abbildung 14 verdeutlicht, lassen sich drei Konstellationen unterscheiden: (1) Für ein gegebenes audiv-perzeptives phonetisches Phänomen existiert bislang noch keine stimmsignalbezogene Quantifizierung (z.B. weiche Stimmeinsätze); (2) Für ein gegebenes audiv-perzeptives phonetisches Phänomen existiert eine stimmsignalbezogene Quantifizierung (z.B. Satzmelodie und Fundamentalfrequenz-Verlauf); (3) Für eine gegebene Signalverarbeitungskennzahl (aus dem Zeit-, Frequenz- oder Phasenraumbereich) existiert kein bekanntes audiv-perzeptives phonetisches Phänomen (Delta-Delta Mel Frequency Cepstral Coefficient 8).

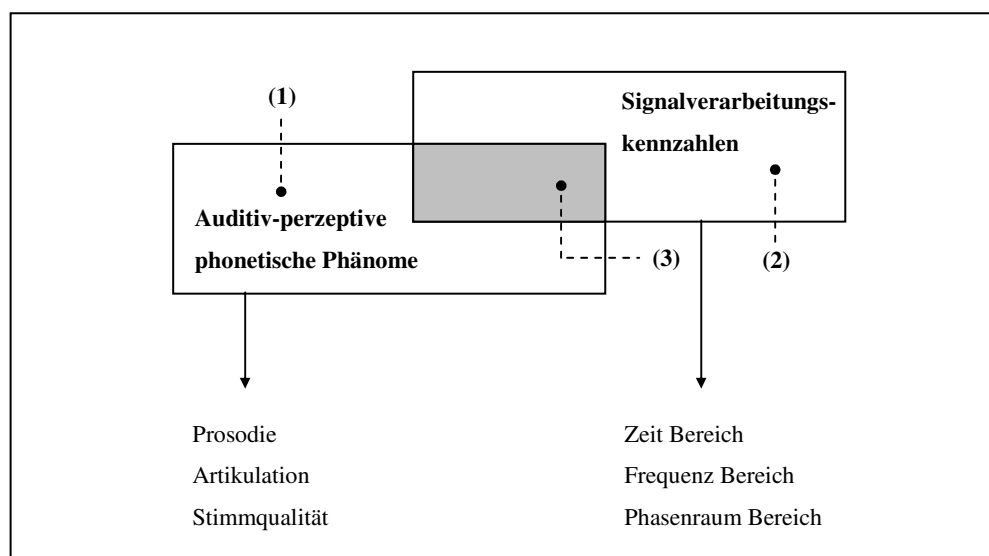


Abbildung 14: Gegenstandsbereich der auditiv-perzeptiven phonetischen Phänomene (1) sowie der die Rohwavenform quantifizierenden Signalverarbeitungskennzahlen (3). Die graue Schnittfläche (2) markiert bereich über Signalkennzahlen erfassbare auditiv-perzeptive phonetische Phänomene

Prosodie-bezogene akustische Beschreibungsgrößen. Wie bereits oben erwähnt, versucht der nachfolgende Abschnitt Signalverarbeitungskennzahlen zu beschreiben, für die es ein ungefähres Pendant im Bereich der auditiv-perzeptiven Phänomene, wie Prosodie, Artikulation und Stimmqualität gibt. Verschiedene sprachliche Segmente wie Phone, Silben, Worte oder phrasale Einheiten (Chunks, Tunes, Turnes) können als Referenzeinheiten akustischer Beschreibungsgrößen dienen. Prosodische Merkmale (Hauptgruppen: Intensität, Intonation, und Dauer) lassen sich jedoch nur bedingt auf die kurzen sprachlichen Segmente wie Phone anwenden, da man im wesentlichen suprasegmentale, also über mehrere Einzellaute verlaufende, Eigenschaften der Melodie und des Rhythmus in dieser Kategorie zusammengefasst.

- *Intensität.* Zur Bestimmung der Lautstärke wird die mittlere quadrierte Amplitude des Rohsignals innerhalb eines Zeitsegments berechnet. Wichtig ist bei der Messung von intensitätsbezogenen Lautstärke-Größen, dass Aufnahme-richtung und Mikrofon-Distanz annähernd konstant gehalten werden (wie in Fahrzeugszenarien umgesetzt) oder nur normalisierte Änderungen der Energie berechnet werden.
- *Intonation.* Zur Bestimmung der Tonhöhe (Vibrationsrate der Stimmlippen) und Sprechmelodie wird der Verlauf der Fundamental Frequenz über die Maxima der Autokorrelationsfunktion berechnet.
- *Dauer.* Als Kenngrößen können z.B. die Sprechrate (Anzahl der Sprachsegmente pro Zeit), die Dauer von Pausen (innersilbisch, zwischen Phrasen), der Anteil von Pausen pro Sprechzeit sowie die Dauer von Vokalen und Konsonanten bestimmt werden.

Fundamental Frequency. Klänge können in zwei Kategorien unterteilt werden: in harmonische und nicht-harmonische. Ein harmonischer Klang besteht aus einer Reihe von bedeutenden Frequenzkomponenten, welche die Grundfrequenz und die ganzzahligen Vielfachen derselben beinhalten. Bei einem Sprachsignal sind die stimmhaften Anteile harmonisch und die stimmlosen Anteile nicht harmonisch. Die Sprachgrundfrequenz (Fundamental Frequenz) ist das akustische Korrelat zur perzeptiv wahrgenommenen Tonhöhe einer Äußerung. Zur Schätzung der Grundfrequenz existieren eine Reihe von Pitch Detection Algorithmen (Mousset, Ainsworth & Fonollosa, 1996).

Grundsätzlich kann zwischen Verfahren unterschieden werden, die auf der Grundlage des zeitlichen oder des spektralen Verlaufs arbeiten. Die als besonders präzise geltende Autokorrelationsmethode nutzt das Maximum der Autokorrelationsfunktion als Schätzer der Grundfrequenz (Boersma, 1993). Folgende Überschlagsangaben zu Grundfrequenzen der Sprech-

stimme lassen sich geben. Die Streubreite der Mittelwerte der individuellen Grundfrequenzlagen betragen bei Männern 100-150Hz, bei Frauen 190-250Hz und bei Kindern 350-500Hz (Neppert, 1999).

Short Time Energy. Die Short-Time Energy Function zeigt die durchschnittliche Energie im Signal pro Zeitausschnitt, und macht somit eine Aussage über die Lautstärke der verschiedenen Signalabschnitte. Mit diesem Feature können über die Energiewerte stimmhafte von stimmlosen Signalanteilen unterschieden werden. Stimmlose Abschnitte weisen deutlich geringere Energiewerte auf (Rabiner & Schafer, 1978). Wenn das Signal-to-Noise-Ratio ausreicht, kann Stille von hörbaren Abschnitten unterschieden werden. Ebenso können Rückschlüsse über die Periodizität und den Rhythmus des Signals anhand der Energiewerte gezogen werden (Bindreiter, 2006). Berechnet wird die Energie über die Summation der quadrierten Beträge der Signalenergie zu allen Messzeitpunkten eines Rahmens.

$$E_t = \sum_{n=0}^{N-1} |s_t[n]|^2 \quad (2.1)$$

Artikulationsbezogene akustische Beschreibungsgrößen. Der artikulatorische Aufwand bei der Lautbildung von Frikativ-Konsonanten wird über die Berechnung des spektralen Zentroids (Frequenzposition und der spektralen Energiewert) genähert (Van Son & Pols, 1999). Die aus dem Verlauf des spektralen Zentroids gebildeten Funktionale sind z.B. Langzeitmittelwert, Standardabweichung, relatives Maximum, mittlere betragsmäßige Steigung, Standardabweichung der betragsmäßigen Steigung und maximale betragsmäßige Steigung. Die Analyse der Vokalartikulation verwendet die Lage der ersten beiden Formanten, sowie ihren Bezug zu phonetischen Normwerten (Formantenpräzision) (Kienast & Sendlmeier, 2000).

Stimmqualitäts-bezogene akustische Beschreibungsgrößen. Das Verhältnis ganzer spektraler Bänder, die Verhältnisse der Energien einzelner Harmonischer zur Gesamtenergie, die Regressionssteigung der Energieverteilung im Frequenzbandbereich über 1 kHz prägen ebenso wie kleine Schwankungen und Irregularitäten der Tonhöhe und Intensität (Jitter, Shimmer) und die Lage und Bandbreite von Formanten (Klasmeyer & Sendlmeier, 2000) den Stimmqualitätseindruck. So ist die Klangfarbe z.B. abhängig von der Teiltonstruktur sowie der Anzahl und der Stärke der im Klang enthaltenen Obertöne. So indizieren eine große Anzahl an Obertönen tragfähige Stimmen. Energiekonzentrationen in hohen Frequenzen hingegen er-

zeugen einen hellen Stimmklang. Zu den wichtigsten Einzelkennzahlen zählen die Resonanzfrequenzen des Vokaltrakts (Maxima im Spektrum), die Formanten. Sie sind sensitiv für kleine Veränderungen der Vokaltraktform, wie sie durch ein Lächeln (Verkürzung des Vokaltrakts, Erhöhung der Formantenposition) oder eine gekrümmte Körperhaltung bewirkt wird. Weitere mit Stimmqualität assoziierte Maße sind Mel Frequency Cepstral Coefficient (MFCC), Roll-Off-Punkte, der spektrale Fluss, Harmonics-to-Noise Ratio und Hammarberg Indizes (vgl. Stevens & Hanson, 1994).

Spectral Centroid. Der Spectral Centroid stellt den Schwerpunkt des Signals dar und ist ein Indikator für die „Fülle“ eines Audiosignals. Bei der Lautbildung von Konsonanten indiziert er den vorgenommenen artikulatorischen Aufwand (Van Son & Pols, 1999). Berechnet wird der spektrale Schwerpunkt über das aufsummierte Produkt der spektralen Energie und der dazugehörigen diskreten Frequenz. Bei gleichbleibender Tonhöhe hat ein Signal mit mehreren oder stärkeren Harmonien einen höheren Schwerpunkt. Der eigentliche Indikator für die Fülle eines Klangs ergibt sich erst aus dem Verhältnis der Spectral Centroids zur Tonhöhe des Signals (Bindreiter, 2006).

$$Spec_cg = \frac{\sum_{f=0}^{(N/2)-20} S_i[f] \cdot f}{\sum_{f=0}^{(N/2)-20} S_i[f]} \quad (2.2)$$

Spectral Flux. Um die Dynamik des gesamten Signalverlaufs bei der Entwicklung der Spektren zu erfassen, kann der sogenannte spektrale Fluss (Spectral Flux) berechnet werden. Der spektrale Fluss beschreibt die lokalen spektralen Frequenzveränderungen über den Signalverlauf (Bindreiter, 2006). Er ist ein Maß dafür, wie stark sich die spektrale Form des Signals verändert. Auf der Seite des perzeptiven Höreindrucks ist mit einem geringen spektralen Fluss ein monotoner Höreindruck verknüpft. Der Spectral Flux ist definiert als Differenz der spektralen Energie zweier aufeinanderfolgender Frames.

Spectral Rolloff. Der Roll-Off-Wert ist wie der Spectral Centroid ein Maß für die spektrale Form eines Signals. Er gibt wieder, bis zu welchem Frequenzband 85% der Gesamtenergie des Spektrums aufkumuliert worden sind (Zölzer, 2002). Viele tieffrequente Anteile reduzieren, viele hochfrequente Anteile im Frequenzspektrum hingegen erhöhen diesen Index.

Spectral/ Formant Bandwidth (spektrale Bandbreite). Die spektrale Bandbreite beschreibt die Frequenzbreite eines Signalpeaks. Bei der Berechnung wird der Frequenzbereich um einen Peak ermittelt, in der das Signal 3 dB an Energie verliert (siehe Abbildung 15). Je kleiner die spektrale Bandbreite ist, desto konzentrierter die Energieverteilung im Signal. So ist bei stimmhaften Lauten die spektrale Bandbreite wesentlich geringer als bei einem Signal, das einen hohen Rauschanteil hat. Darüber hinaus enthält die Bandwidth, wichtige Informationen zur Dämpfung und somit des Energieverlustes eines akustischen Signals im Vokaltrakt. Ein akustisches Signal, wie es bei einem schläfrigen Sprecher zu erwarten ist, verliert um so mehr Energie je entspannter der Vokaltrakt ist, und je mehr die Vokaltraktwände mitschwingen (Yielding Wall Effect; Rabiner & Schafer, 1978).

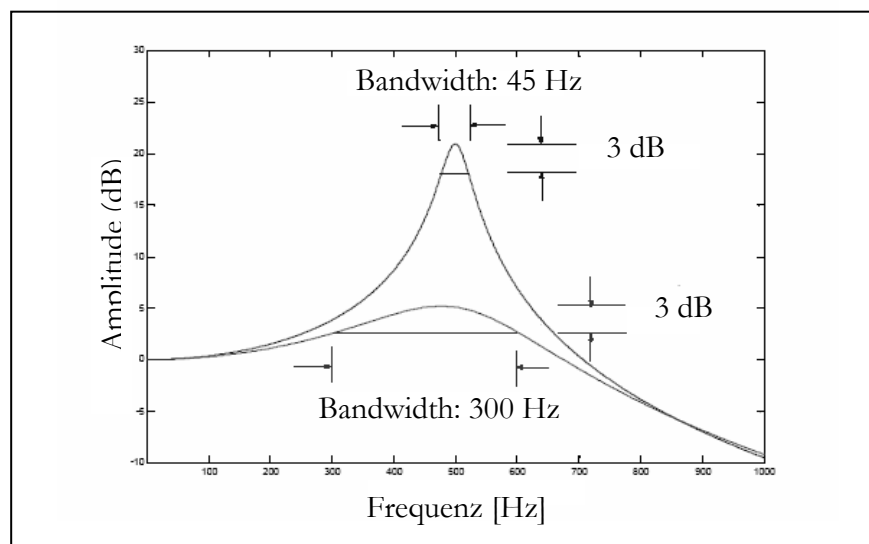


Abbildung 15: Exemplarische Darstellung zweier Frequenzpeaks mit kleiner (45Hz) und großer Bandwidth (300Hz) im Amplitudenspektrum

Harmonics-To-Noise-Ratio. Das in der Signalverarbeitung häufig zur Beurteilung der Signalqualität genutzte Signal-to-Noise Ratio wird aus dem Quotienten der spektralen Leistung des Signals durch die spektrale Leistung des Rauschanteils quantifiziert. Daraus abgeleitet erfasst das die Sprachqualität beurteilende Harmonics-To-Noise-Ratio (HNR) die in den periodischen Anteilen (Harmonischen) enthaltene Signalleistung im Verhältnis zu der Leistung des umgebenden Rauschens. Da periodische Signalanteile nur in stimmhaften Abschnitten auftreten, ist auch das HNR nur in diesen Abschnitten mit einer definierten Grundfrequenz bestimmbar. Das HNR ist gut geeignet, um verschiedene Phonationsarten wie zum Beispiel gehauchte Stimmqualitäten zu erkennen. Geschätzt wird die in dB angegebene HNR aus

normalisierten Größen der Autokorrelationsfunktion. Wichtigste Komponente ist in dieser Bestimmung ist das Maximum der Autokorrelationsfunktion, in dem sich die Leistung der periodischen Sprachanteile widerspiegelt (Boersma, 1993).

$$HNR_t = 10 \cdot \log (AKF_t [T_0] / (AKF_t [0] - AKF_t [T_0])) \quad (2.3)$$

Jitter. Der Jitterfaktor misst die Mikrovariation der Grundfrequenz. Bei der menschlichen Stimme nimmt er ab, wenn die Stimmlippen durch die ansteigende Grundfrequenz angespannter sind und symmetrischer schwingen. Da der Jitterfaktor sensibel auf den Intonationsverlauf reagiert, wurde entsprechend ein "Period Pertubation Quotient" (PPQ) eingeführt.

Shimmer. Shimmer ist definiert als die durchschnittliche Differenz zwischen aufeinanderfolgenden Peak-to-Peak Distances im Rohsignal. Der Shimmerfaktor misst die Mikrovariation der Amplitude. Bei der menschlichen Stimme können diese darauf beruhen, dass die Stimmlippen unfähig sind, periodische (quasiperiodisch) Schwingungen auszuführen, oder dass im Sprachsignal ein hoher Anteil turbulenter Luft vorhanden ist. Da der Shimmerfaktor sensibel auf den Intonationsverlauf reagiert, wurde der "Amplitude Pertubation Quotient" (APQ) eingeführt. Um den echten Shimmer von der Intonationslinie zu trennen, wird der Mittelwert über fünf benachbarte Perioden errechnet. Der Shimmerfaktor ist bei heiserer Stimme erhöht.

Mel Frequency Cepstral Coefficients (MFCC). Die vor allem in der Spracherkennung routinemäßig zum Einsatz kommenden Mel Frequency Cepstral Coefficients (MFCCs) liefern über die Spektralzerlegung des Amplitudenspektrums („Spektrum des Spektrums“) eine holistische Repräsentation des Spektrums. Peaks im Cepstrum korrespondieren mit den Harmonischen des Leistungsspektrums. Die zugrundeliegenden Cepstrum Definition beschreibt das Cepstrum als die Fourier Transformation des logarithmierten Spektrum des Ausgangssignals:

$$\text{Signal} \gg \text{FT} \gg \text{Log} \gg \text{FT} \gg \text{Cepstrum}$$

Bei der Berechnung der MFCCs erfolgt zunächst eine Diskrete Fourier Transformation (DFT) des Eingangssignals, das Amplitudenspektrum wird bestimmt. Anschließend wird das Signal logarithmiert und eine an die menschliche Wahrnehmung angepasste Skalierung des bereits fourier-transformierten Signals vorgenommen. Dieses wird durch die Transformation in die Mel-Skala erzielt, die eine Nachbildung der Frequenz-Orts-Transformation im mensch-

lichen Gehör darstellt. Die Mel-Frequenz-Achse f_{mel} ist eine Koordinatentransformation der linearen Frequenzachse f und ist wie folgt definiert:

$$f_{\text{mel}} = 2595 \log_{10}(1 + (f / 700)) \quad (2.4)$$

Durch die Abbildung der Hz-Skala in die Mel-Skala ist die Auflösung im unteren Frequenzbereich relativ hoch, während sie im oberen Frequenzbereich logarithmisch abnimmt. Im Anschluss daran wird das als Folge von N komplexen Werten darstellbare Spektrum y_j , $j = 0 \dots N-1$ im Spektralbereich mit P Dreiecksfiltern mit der Filterfunktion g , den Filtermittelpunkten f_p und dem Filterstreckfaktor a_p verrechnet (siehe Abbildung 16).

$$m_p = \sum_{j=0}^{N-1} g(a_p | j - f_p) y_j \quad (2.5)$$

Als Resultat ergibt sich eine Folge von P Mel-Frequenz-Koeffizienten $m_0 \dots m_{P-1}$. Diese Koeffizienten m_j werden schließlich mittels einer inversen Diskreten Cosinus Transformation in den Cepstralbereich überführt. Die resultierenden Komponenten o_i entsprechen den Mel Frequency Cepstral Coefficients:

$$o_i = \left[(2 / N) \sum_{j=0}^{N-1} m_j \cos(\pi i / N (j + 0.5)) \right]^{0.5} \quad (2.6)$$

Es handelt sich hierbei um eine strukturerhaltende spektrale Transformation (Rigoll, 2004), deren ursprüngliches in der Spracherkennung angesiedeltes Ziel es war, das abgestrahlte Sprachsignal in das ursprüngliche Stimmlippen-Anregungssignal und in ein die sprecherspezifischen Eigenschaften des Vokaltrakts repräsentierendes Signal zu trennen. Ein wichtiges Charakteristikum der Cepstrum-Koeffizienten ist die Unkorreliertheit ihrer Koeffizienten, die jeweils einen redundanzarmen eigenständigen informativsten Gehalt der Kennzahlen sichert.

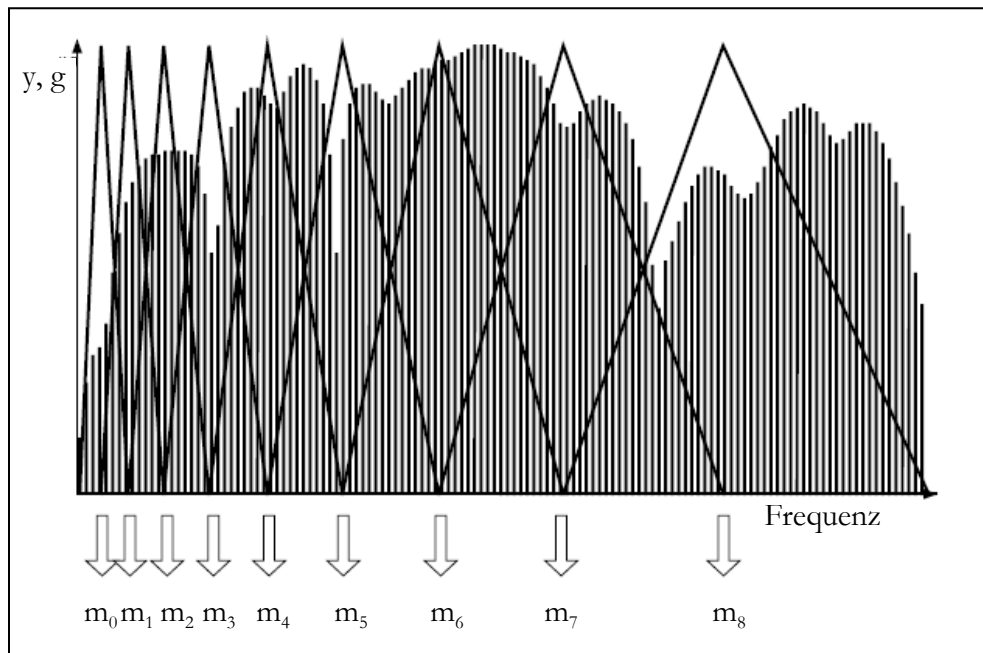


Abbildung 16: Mel-Filterbank Analyse. Das Amplitudenspektrum wird mit den P Dreieckfiltern im Spektralbereich korreliert und resultiert in einer Folge von P Mel-Frequenz-Koeffizienten

Zusammenfassend listet Tabelle 7 die gebräuchlichsten Kennzahlenfamilien auf und umreißt ihre Bedeutung. Die letztendlich berechneten Kennzahlen leiten sich im einfachsten Fall aus dem Mittelwert des jeweiligen zeitlichen Verlaufs der Beschreibungsgrößen ab (z.B. Mittelwert des Intensitätsverlaufs). Denkbar sind aber auch andere verlaufsbeschreibende Kenngrößen wie die deskriptiven Kennwerte: Minimum, Maximum, Median, Standardabweichung, Kurtosis, Schiefe, Steigung Regressionsgerade, Regressionsfehler erster bis n -ter Ordnung, mittlere erste und zweite Ableitung (siehe auch Abschnitt 4.3).

Tabelle 7: Akustische Beschreibungsgrößen aus den Stimmerkmalsklassen Prosodie, Artikulation und Stimmqualität

Beschreibungsgröße	Erläuterung der akustischen Kennzahl
<p>Prosodie</p> <p>Intensität F0 F0 Kontour Sprechrare</p>	<p>Mittlere quadrierte Amplitude innerhalb eines Zeitsegments („Lautstärke“) Fundamental Frequenz (Vibrationsrate der Stimmlippen) („Tonhöhe“) F0 Verlauf („Intonation“, „Sprechmelodie“) Anzahl Sprachsegmente pro Zeit</p>
<p>Artikulation</p> <p>Formant (F1-F5) Spektraler Schwerpunkt</p>	<p>Resonanzfrequenzen des Vokaltrakts (Energie Konz. im Frequenzspektrum) Zentroid der spektralen Energieverteilung</p>
<p>Stimmqualität</p> <p>Jitter Shimmer Spektraler Energie Distrib. Spektraler Slope HNR Formant Bandwidth MFCC</p>	<p>Kleine Schwankungen und Irregularitäten der Tonhöhe Kleine Schwankungen und Irregularitäten der Intensität Relative Stärke einzelner Frequenzbänder (z.B. HF 500) Regressionssteigung der Energieverteilung im Frequenzband über 1 kHz Harmonics-to-Noise Ratio (Anteil der periodischen Signalenergie) Breite des Resonanzpeaks (-3dB Schwelle) 12 Mel Frequency Cepstral Coefficients (sowie 12 Δ und 12 $\Delta\Delta$ MFCC)</p>

3 SCHLÄFRIGKEITSBEDINGTE ÄNDERUNGEN DER SPRACHE

Die in diesem Kapitel angeschnittenen Problemkreise beschäftigen sich zum einen mit den für die spätere Konstruktion eines Messinstruments wichtigen Fragen zu kognitiven und phonetischen Mediatorprozessen der schläfrigkeitsinduziert veränderten Stimme (Abschnitt 3.1). Zum anderen werden empirische Befunde zu auditiv-perzeptuell sowie akustisch erfassten Aktivierungs-bezogenen Stimmänderungen dargestellt (Abschnitt 3.2).

3.1 Kognitiv-phonetisches Mediatormodell

Wie bereits in Kapitel 2 beschrieben, basiert der phonetische Teil der Sprachproduktion auf den drei Funktionskreisen Atmung, Phonation sowie Filter- und Artikulationssystem (vgl. Fant, 1960; Ladefoged, 1975; O'Shaughnessy, 2000). In Anlehnung an das Quelle-Filter-Modell (Abschnitt 2.1) wird hier das Kognitiv-phonetische Mediatormodell schläfrigkeitsinduzierter Stimmveränderung entworfen. Es integriert neben den bei Fant (1960) beschriebenen physiologischen Komponenten der Sprachproduktion auch die kognitive Komponenten der Sprachplanung in ein Modell der Sprachproduktion. Die in diesem Zusammenhang einschlägigen Sprachplanungssysteme werden in Abschnitt 2.3 dargestellt. Des Weiteren werden in dem hier postulierten Kognitiv-phonetischen Mediatormodell auch die aufgrund des schläfrigkeitsinduziert-veränderten Sprachproduktionsprozesses zu erwartenden auditiven Stimmveränderungen angedeutet.

Schläfrigkeitsbedingte kortikale Veränderungen und ihre Wirkung auf die Sprachproduktion. Frith et al. (1991) konnten mittels PET-Scans die Bedeutung des linken präfrontalen Kortex in der Koordination inhibierender und aktivierender Prozesse während der Wortgenerierung nachweisen. Die Funktion der Frontallappen ist durch Schlafentzug beeinträchtigt. So zeigen Untersuchungen mit neuropsychologischen Tests, wie z.B. dem "Tower of London-Test", der bei gesunden Probanden eine signifikante Aktivierung im linken präfrontalen Kortex auslöst (Morris et al., 1993), dass die Aktivität der frontalen Areale durch eine Nacht Schlafentzug reduziert ist (Horne, 1998). Im Gegensatz dazu sind nicht-frontalorientierte Aufgaben ähnlicher Komplexität nicht beeinflusst. Auch Thomas et al. (1993) konnten über PET-Scan (Positron Emission Tomography) Messungen eine Reduktion der Frontallappen Aktivierung nach Schlafentzug feststellen.

Die Wortflüssigkeit ist, wie sich im MRT-Scan demonstrieren ließ (MacCarthy et al., 1993), eine frontal gesteuerte Aufgabe und durch Schlafentzug signifikant beeinträchtigt (Horne, 1998; May & Kline, 1987). Die vorgestellten Ergebnisse der Untersuchungen zum Zusammenhang zwischen Schläfrigkeit und der Aktivierung spezifischer Hirnareale während der Sprachproduktion, lassen die Annahme zu, dass schläfrigkeitsinduzierte Einflüsse auf frontale und temporale Hirnregionen die Sprachproduktion beeinflussen und Aktivierungsprozesse dabei eine wesentliche Rolle spielen. Vertiefende Hinweise zu neuronalen Sprachmodelle finden sich bei u.a. bei Kröger, Birkholz und Neuschaefer-Rube (2007).

Beeinträchtigungen kognitiver Voraussetzungen der Sprachproduktion

Allgemein umfassen schläfrigkeitsbedingte kognitive Beeinträchtigungen folgende Bereiche: Vigilanz, Arbeits- und Kurzzeitgedächtnis, exekutive Funktionen, mathematisches Denken, kognitive Verarbeitungsgeschwindigkeit, räumliches Vorstellungsvermögen, motorische Kontrolle (vgl. Dorrian, Rogers & Dinges, 2005; Durmer & Dinges, 2005; Wesensten, Belenky, Thorne & Kautz, 2004), überwachende Steuerungsfunktionen (vgl. Nilsson, Soderstrom, Karlsson & Lekander, 2005), Problemlösen (Horne, 1988; Linde & Bergström, 1992), divergentes Denken und sprachliche Kreativität (Harrison & Horne, 1998). Von den genannten Forschungsergebnissen sind für die mentalen Bedingungen der Sprachproduktion insbesondere die Befunde zu Aufmerksamkeit, Gedächtnis und kognitiver Verarbeitungsgeschwindigkeit von Bedeutung.

Wirkung reduzierter Aufmerksamkeit auf die Sprachproduktion. Levelt (1989) postuliert in seinem Modell, dass konzeptuelle Sprachproduktionsprozesse kontrolliert ablaufen und mit Aufmerksamkeit verbunden sind. Des Weiteren nimmt er eine aufmerksamkeitsabhängige, für Selbstkorrekturen verantwortliche Kontrollinstanz an (vgl. Levelt, 1989; Stemberger, 1985). Aufmerksamkeit nimmt somit eine wichtige Rolle als kognitive Komponente der Sprachproduktion ein. Die Befunde zu schläfrigkeitsinduzierten Beeinträchtigungen der selektiven Aufmerksamkeit (vgl. Kotterba et al., 1998) lassen die Annahme zu, dass Schläfrigkeit die aufmerksamkeitsfordernden, konzeptuellen Prozesse der Sprachproduktion beeinflusst und Auswirkungen auf die Erstellung des Sprechplans - also der propositionalen Repräsentation dessen, was gesagt werden soll - hat.

Wirkung beeinträchtigter Gedächtnisleistungen auf die Sprachproduktion. Eine weitere für die Sprachproduktion wesentliche kognitive Komponente ist das Gedächtnis, da Gedächtnisinhalte auf allen Ebenen der Sprachproduktion abgerufen werden (vgl. Abbildung 12). Nach Levelt (1989) greift der *Konzeptualisierer* mit Hilfe des Arbeitsgedächtnisses (nach Baddeley, 1986) auf das Langzeitgedächtnis zurück. Im *Formulator* werden im Prozess des grammatischen Enkodierens Lemmas aus dem mentalen Lexikon abgerufen. Während der phonologischen Enkodierung wird auf metrisches und segmentales Wissen zurückgegriffen. Zwischenergebnisse können während des Enkodierens im syntaktischen Puffer (Bestandteil des verbalen Kurzzeitgedächtnisses) gespeichert werden. Im *Artikulator* werden die abstrakten artikulatorischen Gesten in konkrete motorische Programme umgesetzt, die aus dem Langzeitgedächtnis abgerufen werden. Benötigt die Artikulation viel Zeit, kann sie im artikulatorischen Puffer zwischengespeichert werden. Nach den Aktivationsausbreitungsmodellen sind in den Knoten u.a. semantische, syntaktische, lexikale und phonologische Regeln und Informationen gespeichert, die aktivationsabhängig in den Sprachproduktionsprozess einfließen.

Die Befunde zu schläfrigkeitsinduzierten Beeinträchtigungen des Gedächtnisses (vgl. Dorrian, Rogers & Dinges, 2005; Durmer & Dinges, 2005; Wesensten, Belenky, Thorne & Kautz, 2004) lassen die Annahme zu, dass Schläfrigkeit gedächtnisabhängige Prozesse der Sprachproduktion beeinflusst und u.a. Auswirkungen auf die kontextadäquate Reaktion hat, die einen Rückgriff auf relevante Gedächtnisinhalte (Situation, Kontext, Person) erfordert. Insbesondere die Wortfindung, das Abrufen adäquater grammatischer und phonologischer Strukturen und die Speicherdauer des phonetischen Plans im artikulatorischen Puffer können beeinträchtigt sein.

Wirkung reduzierter kognitiver Verarbeitungsgeschwindigkeit auf die Sprachproduktion. Der Begriff der kognitiven Verarbeitungsgeschwindigkeit bezieht sich auf die Geschwindigkeit der Verarbeitung von Reizen und umfasst die Prozesse der Reizwahrnehmung, -aufnahme, -verarbeitung, -speicherung, des Reizabrufs und der Ausgabe der verarbeiteten Informationen (Sturm et al., 2000). Auch wenn das Konzept der kognitiven Verarbeitungsgeschwindigkeit kein integraler Bestandteil der Sprachproduktionsmodelle ist, sind die Befunde zu schläfrigkeitsinduzierten Beeinträchtigungen der kognitiven Verarbeitungsgeschwindigkeit von großem Interesse (vgl. Cassel, et al., 1989; Dorrian, Rogers & Dinges, 2005; Durmer & Dinges, 2005; Wesensten, Belenky, Thorne & Kautz, 2004). So beeinflusst eine verlang-

samte kognitive Verarbeitungsgeschwindigkeit, alle Prozesse der Sprachproduktion - von der Konzeption, über die Formulierung bis hin zur Artikulation. Folgen einer verringerten kognitiven Verarbeitungsgeschwindigkeit könnten allgemein langsames Sprechen sowie häufigere und längere Sprechpausen sein. In Abbildung 17 werden die Zusammenhänge zwischen schläfrigkeitsinduzierten Veränderungen der kognitiven Voraussetzungen der Sprachproduktion und möglichen Auswirkungen auf den Sprachproduktionsprozess und das Sprechergebnis dargestellt.

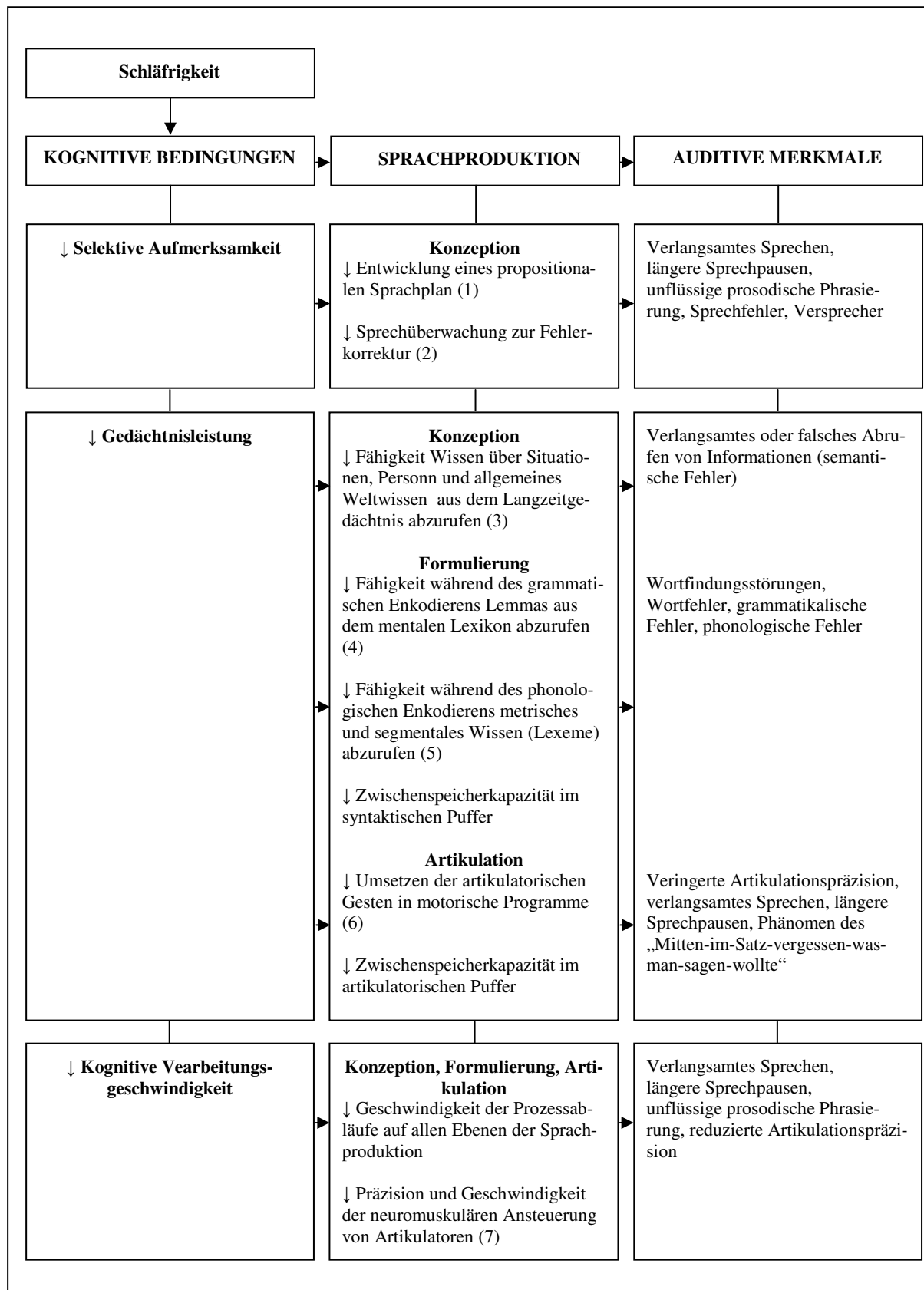


Abbildung 17: Kognitive Komponente des Kognitiv-phonetischen Mediatormodells schläfrigkeitsinduzierter Sprachveränderung. Die Ziffern wurden hinzugefügt und stellen Bezüge zum Sprachproduktionsmodell von Levelt (1989) her (siehe Abbildung 14)

Schläfrigkeitsinduzierende physiologische Prozesse. Nachdem im vorangegangenen Abschnitt die kognitive Komponente des Kognitiv-phonetischen Mediatormodells erläutert wurde, werden in diesem Abschnitt zunächst die Hintergründe der physiologischen Komponente dieses Modells betrachtet. Wie in Abschnitt 1.4 bereits beschrieben, wird Schläfrigkeit auch als das physiologische Bedürfnis des Organismus nach Schlaf verstanden. Die Körperkern-temperatur ist die Leitvariable des zirkadianen Rhythmus und wird über den Hypothalamus gesteuert. Im Schlaf fällt die Kernkörpertemperatur um zirka $0,5^{\circ}\text{C}$ ab. Passend zu diesem Sachverhalt ist auch der zirkadiane Schlafdruck ist am höchsten, wenn die Körpertemperatur ihren tiefsten Punkt erreicht hat (Fietze, 2007). Die Körperkerntemperatur- und Blutdruckrhythmen haben ihre absoluten Minima gegen drei Uhr nachts bei etwa gleichzeitigem Maximum der Melatoninausschüttung. Zunehmende Schläfrigkeit ist allgemein mit einer homöostatischen und zirkadian gesteuerten und über Neurotransmitter und Peptide vermittelten Abnahme der zentralnervösen Aktivierung (Arousal) verbunden. Die physiologische Aktivierung des Organismus sinkt in dieser Phase, der Einfluss des Sympathikus nimmt ab und der parasymphatische Einfluss zu. Daraus resultierend nehmen die Herzfrequenz sowie Atemfrequenz und -tiefe ab, der Blutdruck sinkt, der Muskeltonus wird schwächer und die sensorische Reizschwelle steigt (Peter & Peter, 2007).

Hinweise auf schläfrigkeitsinduzierte Beeinträchtigungen physiologischer Voraussetzungen der Sprachproduktion. Für die Sprachproduktion sind insbesondere die mit Schläfrigkeit verbundene Abnahme der Körpertemperatur sowie der schwächere generelle Muskeltonus von Bedeutung. Darüber hinaus spielen auch die mit veränderten emotionalen Zuständen (verringerte emotionale Aktiviertheits- und Valenzdimension) einhergehenden physiologischen Veränderungen prinzipiell eine Rolle. Eine in Verbindung mit einer generell reduzierten Aktiviertheit und einer sinkenden Muskelspannung stehende, Komponente der Sprachproduktion ist die Respirationsphase. So funktioniert die Regulation der Atmung einerseits homöostatisch, andererseits beinhaltet sie im Wachzustand wesentliche zusätzliche Einflüsse. Sie ist beispielsweise beim Sprechen oder Singen auch über die Willkürmotorik ansteuerbar, so dass die Atmung insgesamt als hybrid geregelte Größe zu betrachten ist (Podszus, 2007). Bei Ermüdung wird mit dem Auftreten von Alpha-Aktivität im EEG eine Abnahme der Atemfrequenz und Atemtiefe gemessen. Vor allem das so reduzierte Tidalvolumens (Atemzugvolumens) spielt für den weiteren Verlauf der Sprachproduktion eine wichtige Rolle. Beeinflusst wird die Atmungsregulation in diesem Sinne durch den fortschreitenden Entspannungsverlauf des allgemeinen Muskeltonus (siehe Abschnitt 2.1). Im Zuge des

fortschreitenden Müdigkeitsverlaufs kommt es zu einer progressiven Abnahme des auch für die Sprachproduktion relevanten Muskeltonus (Mayer, 2007). Gemeinsam mit diesem hier beschriebenen schläfrigkeitsbedingten Respirationsmuster formt die nachfolgende Phonation das Anregungssignal $X(n)$ (siehe Abbildung 18).

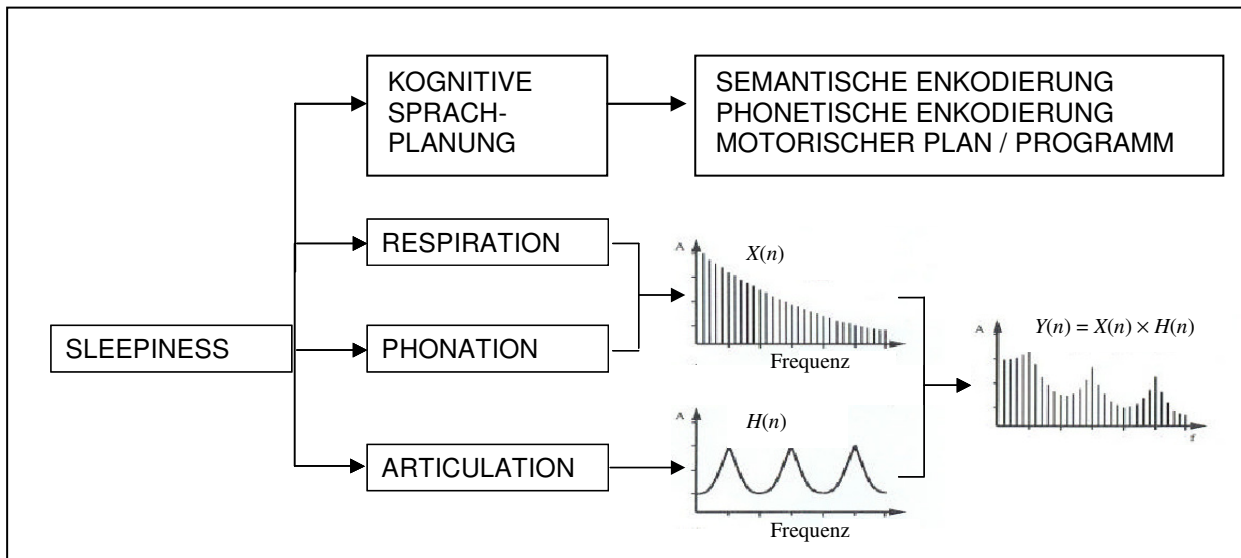


Abbildung 18: Ausdifferenziertes Quelle-Filter Modell mit Ansatzstellen der schläfrigkeitsinduzierten Veränderungen des Sprachproduktionssystems. Angedeutet sind neben den Komponenten der Sprachproduktion kognitive Sprachplanung, Respiration, Phonation, und Artikulation auch die akustischen Wirkungen auf das Anregungssignal $X(n)$, die stilisierte Transferfunktion $H(n)$ sowie das gefilterte Ausgangssignal $Y(n)$

Wie in Abschnitt 2.1 dargestellt wurde, ist die Stimmtonbildung (Phonation) ein komplexes Zusammenspiel von neuraler Aktivität, Muskelspannung, Regulation des subglottalen Luftdrucks und aerodynamischen Prozessen. Im Kognitiv-phonetischen Mediatormodell wird postuliert, dass die schläfrigkeitsinduzierte Abnahme der Atemfrequenz und -tiefe zu einem verringerten subglottalen Druck führt, der mit einer Abnahme von Stimmintensität und Tonhöhe verbunden ist. Schon kleine Änderungen im subglottalen Druck haben große Veränderungen in der Intensität zur Folge (Ladefoged, 1967).

Neben einem flacheren Intensitätsverlauf wird auch ein flacherer F_0 -Verlauf erwartet. Es kann zur Knarrstimme kommen, bei der der hintere Teil der Stimmlippen während der Stimmhaftigkeit geschlossen bleibt, da aufgrund des verminderten subglottalen Drucks der Verschluss der Stimmlippen nicht vollständig gesprengt wird.

Im Modell wird weiterhin postuliert, dass die schläfrigkeitsinduzierte Verringerung des Muskeltonus sich neben der Atmung auch auf die über neurale Aktivierung und Muskelspannung kontrollierte Stellung der Kehlkopfknorpel sowie die Spannung und Elastizität der Stimmlippen auswirkt. Die Stellknorpel nähern die beiden Stimmlippen einander an, und die charakteristische Spannung des Vocalismuskels erlaubt das Schwingen der Stimmlippen, d.h. das schnelle und regelmäßige Öffnen und Schließen der Glottis. Durch eine schläfrigkeitsinduziert verringerte Spannung des Vocalis-Muskels (Stimmlippen) und verringerte Myoelastizität der Stimmlippen kann es zu einem unvollständigen Stimmlippenschluss kommen - die Folge sind weiche Stimmeinsätze und eine behauchte Stimme, bei der die Stimmlippen vibrieren, aber keinen kompletten Kontakt in der Schließungsphase haben. Der akustische Parameter, der eine behauchte bzw. heisere Stimme indiziert, ist eine niedrige Harmonics-to-Noise-Ratio (Jiang et al., 2000) (vgl. Abschnitt 2.4).

Der verringerte Muskeltonus wirkt sich neben der Phonation auch auf die Artikulation aus. Reduzierte Zungen- und Kieferbewegungen sowie eine ausdrucksärmere Mimik verändern die Filter- und Abstrahlungscharakteristika und somit die Artikulationspräzision von schläfrigen Sprechern (vgl. Harrison & Horne 1997; Whitmore & Fisher 1996). Die Artikulationspräzision (vgl. Abschnitt 2.3) lässt nach, die Lautdauer nimmt zu. Denkbare schläfrigkeitsbedingt verstärkte Artikulationsphänomene sollen am folgenden Beispiel „mit dem Wagen“ illustriert werden (vgl. Hess, 2002; nicht erwähnt sind hier darüber hinaus denkbare kompensatorische Hyperartikulationsphänomene):

„mit dem Wagen“

[mit ^h de:m va:gən]	über Zentralisierung des Vokals zu
[mit ^h dəm va:gən]	über Schwa-Elision zu
[mit ^h dm va:gən]	über Unterdrückung der Aspiration vor dem Plosiv zu
[mit dm va:gən]	über regressive Assimilation des Artikulationsortes zu
[mip p bm va:gən]	über Geminantenreduktion zu
[mi bm va:gən]	über regressive Assimilation der Artikulationsart zu
[mi mm va:gən]	

Es kann zur Assimilation und zur Lenisierung bis hin zur Elision kommen. Es wird angenommen, dass die Frequenz des ersten Formanten sinkt und die Bandbreite aller Formanten abnimmt, bei gleichzeitiger spektraler Energieverteilung in Richtung tiefer Frequenzen. Eine weitere mögliche Folge des verringerten Muskeltonus ist die Absenkung des Velums und somit der Zuschaltung nasaler Resonanzräume, die auditiv die Wahrnehmung stimmlicher Nasalität hervorruft. Nach Laver (1980) ist eine Form der Nasalität (Sensuous Nasality) auf eine allgemeine Entspannung im oropharyngalen Bereich (zwischen der Epiglottis und dem Velum) zurückzuführen (vgl. Abbildung 19). Die spektralen Effekte dieser Form von Nasalität sind jedoch bis jetzt jedoch noch nicht hinreichend bekannt (Scherer, 1986). Die allgemeine Muskelspannung nimmt Einfluss auf die Körperhaltung (eingefallener Oberkörper, nach vorn geneigter Kopf). Die Atmung wird dadurch erschwert und der Vokaltrakt gekrümmt. Es wird angenommen, dass sich aus der schläfrigkeitsinduzierten Verkrümmung der Körperhaltung und einer veränderten (verkürzten) Form des Vokaltrakts eine Verschiebung der Formantenlage ergibt (Rabiner & Schafer, 1978).

Auch die Elastizität der Vokaltraktwände wird durch den Muskeltonus beeinflusst. Die Elastizität der Vokaltraktwände bestimmt die Dämpfung und somit den Energieverlust des Signals. Widerspiegeln könnte sich daher eine schläfrigkeitsbedingte Veränderung der Elastizität der Vokaltraktwände in breiteren Formantenpeaks, also einer vergrößerten Formantenbandbreite (vgl. Rabiner & Schafer, 1978).

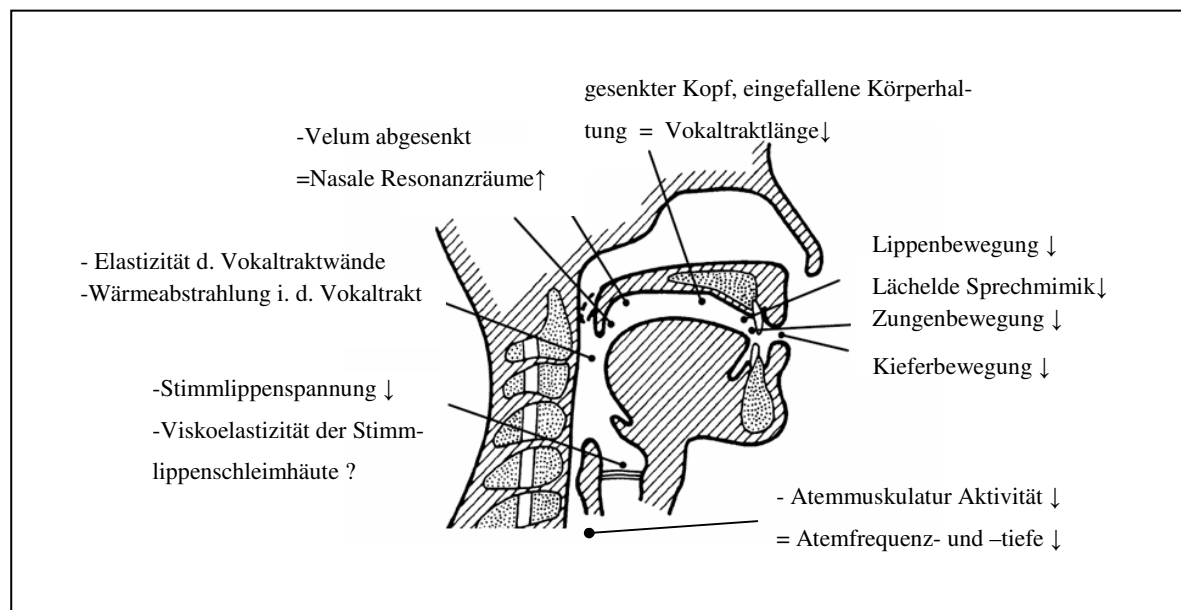


Abbildung 19: Schläfrigkeitsbedingte physiologisch-phonetische Veränderungen des Sprachproduktionssystems

Darüber hinaus kann eine sinkende Körperkerntemperatur, die Luft in den Atemwegen abkühlen. Kältere Luft nimmt weniger Feuchtigkeit auf als wärmere Luft, was die Schleimhäute im gesamten Vokaltrakt trockener werden lässt. Die Stimmlippen-Schleimhaut bewegt sich auf dem Muskelgewebe so, dass aus dem Ausatemstrom in optimaler Weise Energie zum Antrieb der Stimmlippenschwingung entnommen und ein dichtes und sehr schnelles Schließen der Stimmlippen ermöglicht wird, was zu einem rauschfreien Stimmklang führt (Brügge & Mohs, 1996). Die Feuchtigkeit der Stimmbandschleimhäute scheint einen signifikanten Effekt auf die Stimmqualität und die Leichtigkeit der Stimmproduktion zu haben (Verdolini, 1994). Folge trockener Stimmlippen-Schleimhäute könnte für den Stimmlippenverschluss-Zyklus (vgl. Abbildung 6) ein weniger geschmeidiges, reißverschlussartiges Schließen der Stimmlippen sein. Dieser eher „klatschende“ Verschluss könnte sich in einer veränderten auditiv-perzeptiven Stimmqualitäts-Eindruck, sowie einer breiteren Verteilung der Energie des Amplitudenspektrums niederschlagen. Trockene Mundschleimhäute könnten darüber hinaus das Anlegen und Lösen von Artikulatorenbewegung beeinflussen und somit zu veränderten Artikulationsübergängen führen. Eine veränderte Viskoelastizität der Ansatzrohr-Schleimhäute kann auch hier in veränderten Reibungs- und daher Strömungsverhältnissen resultieren (Titze, 1995). Denkbar ist hier die Beeinflussung nicht-linearer Turbulenzphänomene (vgl. auch nicht-laminarer Flow, Jet Formation). Eine schläfrigkeitsbedingte reduzierte

Wärmeabstrahlung der Vokaltraktwände in den Vokaltrakt könnte diese Turbulenzphänomene zusätzlich beeinflussen (vgl. Rabiner & Schafer, 1978).

Die in diesem Kapitel beschriebenen Veränderungen der phonetisch-physiologischen Grundlagen der Sprachproduktion interagieren in einem komplexen Wechselspiel, so dass exakte Prognosen zu schläfrigkeitsinduzierten stimmlichen Veränderungen zum gegenwärtigen Forschungsstand nur schwer zu treffen sind. Vor dem Hintergrund dieser Einschränkungen fasst Abbildung 20 die plausibelsten Vermutungen über schläfrigkeitsinduzierte Einzelprozesse schematisch zusammen.

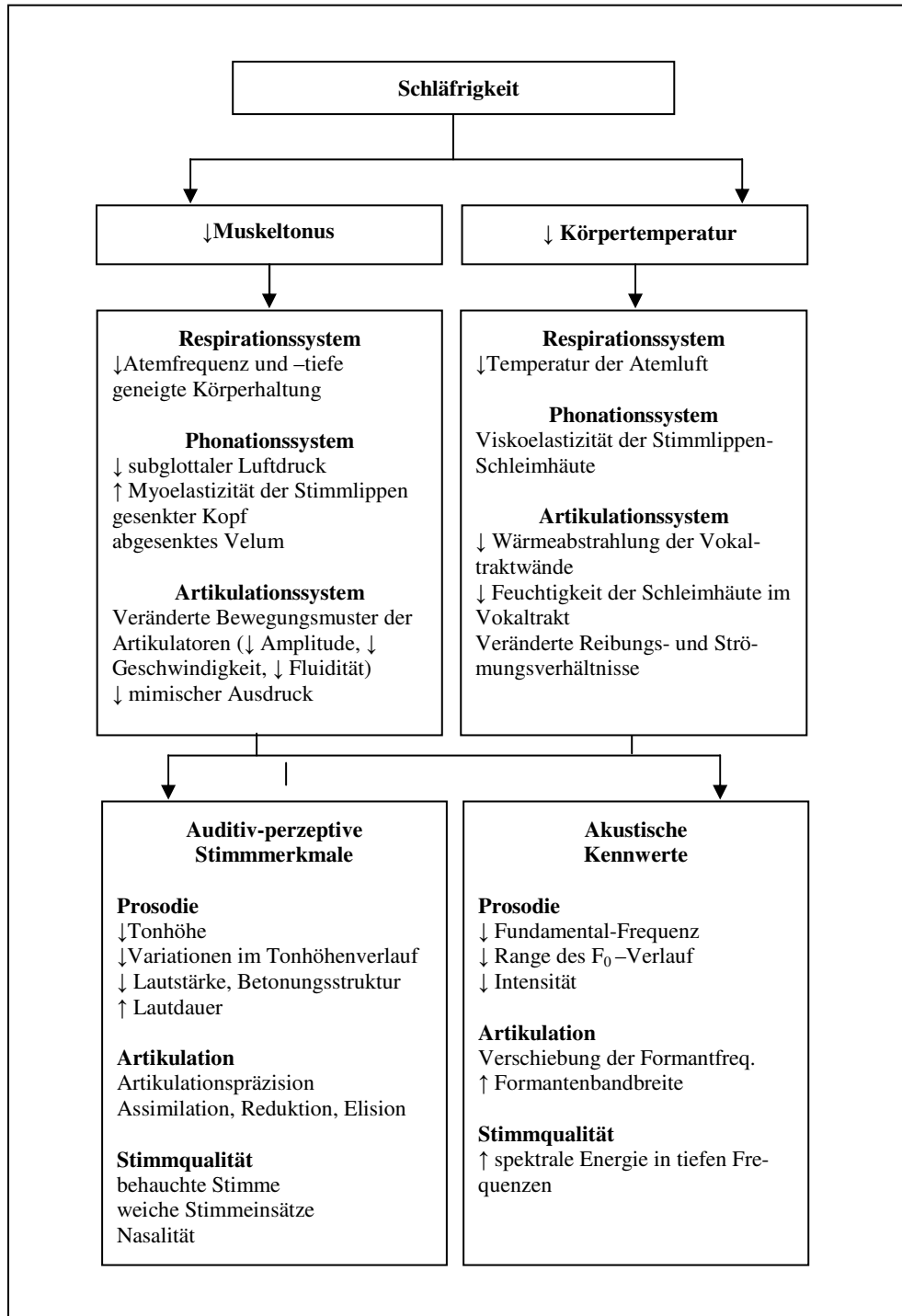


Abbildung 20: Potenzielle Wirkungen von schläfrigkeitsinduzierten physiologischen Veränderungen auf auditiv-perzeptive und akustische Stimmerkmale

3.2 Empirische Befunde: schläfrigkeitsinduzierte Stimmveränderungen

Zahlreiche Studien belegen die Validität der akustischen Stimmanalyse in der Detektion von internen Zuständen wie Basisemotionen, Stress, „User States“ (z.B. Frustration, Belästigung), Alkoholintoxikation und Schläfrigkeit (z.B. Batliner et al., 2006; Cowie et al., 2001; Fernandez, 2003; Juslin & Laukka, 2003; Kienast & Sendlmeier, 2000; Klasmeyer & Sendlmeier, 2000; Paeschke, 2003; Scherer, 2001; Schuller, 2006; Zhou, Hansen & Kaiser, 1999). Beschrieben werden im Folgenden jedoch nur die Befunde, die schläfrigkeitsverwandte Zustände wie die Aktivierungsdimension der Emotion und Alkoholintoxikation untersuchen. Im Anschluss werden Studien mit direktem Bezug zu Ermüdung und Schläfrigkeit vorgestellt.

Akustische Korrelate der Emotionsdimension „Aktiviertheit“. Emotionen rufen neben kulturabhängigen Skripten zur sprachlichen Emotionsdarstellung auch physiologische Veränderungen des Sprachproduktion hervor (Juslin & Laukka, 2003; Scherer, 1986). Diese Veränderungen spiegeln sich vor allem in den Emotionsdimensionen „Aktiviertheit“ und „Valenz“ wider (z.B. Bachorowski, 1999; Cacioppo et al., 2000; Lang et al., 1998; Scherer, 1986). Zusammenhänge zwischen emotionalem Aktivierungsniveau und stimmlicher Veränderung sind, wie aus Tabelle 8 zu entnehmen, gut dokumentiert (Laukka et al., 2005). Eine hohe Aktivierung wird üblicherweise assoziiert mit einer hohen durchschnittlichen Grundfrequenz (F_0) (Breitenstein, Van Lancker & Daum, 2001; Davitz, 1964; Levin & Lord, 1975; Pereira, 2000; Scherer & Oshinsky, 1977; Schröder, Cowie, Douglas-Cowie, Westerdijk & Gielen, 2001), einer hohen F_0 -Variabilität (Breitenstein et al., 2001; Pereira, 2000; Scherer & Oshinsky, 1977; Schröder et al. 2001; Uldall, 1960), einer hohen Sprechgeschwindigkeit (Apple, Streeter & Krauss, 1979; Breitenstein et al., 2001; Davitz, 1964; Kehrein, 2002; Scherer & Oshinsky, 1977), kurzen Sprechpausen (Schröder et al., 2001), einer Zunahme der Stimmintensität (Davitz, 1964; Huttar, 1968; Pereira, 2000; Schröder et al, 2001) und einer Zunahme der Energie in den hohen Frequenzen (Pitam, Gallois & Callan, 1990; Scherer & Oshinsky, 1977; Schröder et al., 2001). Andere Untersuchungen dokumentieren Zusammenhänge zwischen niedriger Aktiviertheit und tiefen F_0 - Werten (Tonhöhe), flachen Intonationsverläufen, niedrigen Intensitäten, geringen Intensitätsschwankungen, niedrigen Positionen des ersten Formanten (F_1), schmaler Bandbreite von F_1 , niedriger F_1 -Präzision (verschleißene Artikulation) sowie geringer spektraler Energie der Frequenzen über 500Hz (Juslin & Laukka, 2001; Scherer, 1986).

Tabelle 8: Akustische Korrelate der Emotionsdimension „Aktiviertheit“, „Valenz“ und „Potenz“ (nach Laukka et al., 2005)

Emotions Dimension	Akustische Merkmale
Hohe Aktiviertheit	Hoher F_0 -Mittelwert, große F_0 -Variabilität, hohes F_0 -Maximum, hohes F_0 -Baselevel, hoher Intensitäts-Mittelwert, große Intensitäts-Variabilität, hoher F_1 -Mittelwert, kleine F_1 -Bandwidth, präzise Articulation, kleine F_3 -Bandwidth, hohe Energie hochfrequenter Anteile im Amplitudespektrum, schnelle Sprechrate, wenig Pausen, langsame Voice Onsets
Positive Valenz	Niedriger F_0 -Mittelwert, große F_0 -Variabilität, niedriges F_0 -Baselevel, niedriger Intensitäts-Mittelwert, kleine Intensitäts-Variabilität, niedrige F_1 -Mittelwert, verschliffene Artikulation, niedrige F_2 -Bandwidth, niedrige Energie hochfrequenter Anteile im Amplitudespektrum, schnelle Sprechrate, schnelle Voice Onsets
Hohe Potenz	große F_0 -Variabilität, niedriges F_0 -Baselevel, hoher Intensitäts-Mittelwert, große Intensitäts-Variabilität, hoher F_1 -Mittelwert, kleine F_1 -Bandwidth, präzise Artikulation, niedriger F_3 -Mittelwert, kleine F_3 -Bandwidth, hohe Energie hochfrequenter Anteile im Amplitudespektrum, langsame Sprechrate, langsame Voice Onsets

Akustische Korrelate der Alkoholintoxikation. Einige Ähnlichkeiten zwischen schläfrigkeitsbedingter muskulärer Relaxation und reduzierter kognitiver Verarbeitungsgeschwindigkeit lassen die Analyse von alkoholisierter Sprache auch für die Schläfrigkeitsdetektion sinnvoll erscheinen. So konnten Brenner und Cash (1991) dokumentieren, dass als Folge von Alkoholintoxikation verlangsamte Sprechgeschwindigkeit, Sprachfehler, Fehlartikulationen von schwierigen Lautkombinationen und Veränderungen in der Stimmqualität auftreten. Cummings et al. (1996) vermuteten das Alkoholintoxikation Einfluss auf die glottale Exzitation hat. Exzitationsparameter wurden daher aus nicht nasalen Vokalen extrahiert, die jeweils unter den Bedingungen „nüchtern“ und „alkoholisiert“ gesprochen wurden. Aus dem Rohsignal wurden fünfzehn Parameter extrahiert, darunter Pitch, Pitch-Kontur, Intensitätsmaße, Jitter und Shimmer. Die größten Unterschiede zwischen nüchterner und alkoholisierter Sprache zeigten sich in Parametern, die die Perturbationen aufeinander folgender Pitch-Perioden (Variabilität von Pitch-Period-to-Pitch-Period) messen. Eine Reduktion des Pitch-Mittelwerts, der aufgrund einer allgemein reduzierten muskulären Spannung innerhalb des Phonationssystems erwartbar gewesen wäre, konnte jedoch nicht nachgewiesen.

Mensík (1999) untersuchte, welche Phone, Phonübergänge und Kennwerte die höchsten alkoholbezogenen Erkennungsraten liefern. Das Klassifikationsproblem beschränkte sich auf die Zuordnung zu zwei Klassen – “nüchtern“ (Blutalkoholwert = 0 Promille) und “alkoholisiert“ (Blutalkoholwert > 0,5 Promille). Als akustische Features kamen Linear Predictive Coding Coefficient Kennwerte (u.a. LPC, LPCC, Log Area Ratio) sowie ihre ersten Ableitungen zum Einsatz. Als Modellierungsverfahren wurde das Dynamic Time Warping eingesetzt. Alle verwendeten Parametrisierungen konnten alkoholisierte Sprache überzufällig genau vorhersagen, die dLAR-Koeffizienten (delta Log Area Ratio) erwiesen im Vergleich zu den anderen Kennzahlen-Blöcken am erfolgreichsten. Die besten Vorhersagekennwerte bezogen sich auf die Konsonanten “r“ sowie deren Übergänge zu Vokalen. Levit et al. (2001) untersuchten in die 2 Klassen “nüchtern“ (Blutalkoholwert = 0 Promille) und “alkoholisiert“ (Blutalkoholwert > 0,8 Promille) eingeteilte Sprachdaten mittels prosodischer Kennwerte und einem MLP (Multi-Layer-Perceptron) Klassifikationsverfahren. Die Resultate zeigten eine durchschnittliche Klassifikationsrate von 69% auf ungesehene Daten. Jitter und Shimmer basierte Charakteristika erwiesen sich hierbei als wichtig für die erfolgreiche Klassifikation. Hollien et al. (2001) untersuchten den Einfluss von Alkohol auf suprasegmentale Merkmale wie Grundfrequenz (F_0), Intensität, Sprechrate und Unflüssigkeiten der Sprache (Disfluencies). Die Ergebnisse zeigten, dass Alkoholintoxikation für 80% der Probanden mit einem signifikanten Anstieg der Grundfrequenz (F_0), der Sprachunflüssigkeiten und der Sprachdauer verbunden ist.

Akustische Korrelate der Ermüdung und Schläfrigkeit. Eine Reihe von Studien beschäftigte sich bislang mit der Identifikation einzelner akustischer Korrelate von Ermüdung und Schläfrigkeit. Die erste dokumentierte Untersuchung von Morris et al. (1960) untersuchte die Wirkung einer 72-98 Stunden langen Schlafdeprivation auf die Sprache. Als Sprachmaterial wurden die spontansprachlichen Berichte der Probanden über ihre persönlichen Erfahrungen während des Schlafentzugs verwendet. Die Ergebnisse zeigten, dass die auffälligsten Veränderungen während des Schlafentzugs im Sprechrhythmus, im Stimmklang und der sprachlichen Klarheit auftraten. Bei fortschreitendem Schlafentzug nahm die Sprachgeschwindigkeit ab, die Stimme wurde weicher, die Aussprache wurde verwaschener und es traten unerwartete Unterbrechungen im Rhythmus auf. Die Mehrzahl der Probanden wiesen ungewöhnliche Variationen in der Lautstärke auf, die Variabilität der Tonhöhe nahm ab. Kritisch ist bezüglich dieser Studie anzumerken, dass es keine objektive akustische Messung der sprachlichen

Merkmale gab und somit lediglich perzeptive Urteile zu müdigkeitsbedingten sprachlichen Veränderungen ungeschulter Rater eingingen.

An die Studie von Morris et al. (1960) anschließend untersuchten Harrison und Horne (1997) die Aspekte der Sprache, die nach Morris et al. (1960) besonders durch Schlafdeprivation beeinflusst werden: Artikulation, Intonation und Wortflüssigkeit. Die Studie wurde an neun Probanden durchgeführt, die zweimal über einen Zeitraum von 36 Stunden getestet wurden - einmal mit einer Nacht Schlafentzug, einmal ohne. In diesem Zeitraum wurden zu drei Messzeitpunkten – nach 6 Stunden, 30 Stunden und 36 Stunden - Messungen über den Thurstone-Wortflüssigkeitstests sowie über gelesene Kindergeschichten durchgeführt. Die Bewertung der anonymisierten und randomisierten Sprachaufnahmen erfolgte an den letzten 10 Sekunden der Aufnahmen. Die Bewertung umfasste die Kategorien: “angemessener Gebrauch der Intonation“, “angemessene Veränderungen in der Lautstärke“, “angemessene Veränderungen in der Geschwindigkeit“ sowie “stockt und macht Fehler“. Es wurde erwartet, dass die Wortflüssigkeit abnimmt und schlafentzugsbedingte Veränderungen der Intonation, der Lautstärke und der Wortfehler auftreten. Die gefundenen Ergebnisse unterstützten die Annahme, dass die Wortflüssigkeit und die Anzahl der gesprochenen Wörter abnimmt, was konsistent mit den Ergebnissen von Horne (1998) und May und Kline (1993) ist. Des Weiteren fanden Harrison und Horne, dass nach 36-stündiger Schlafdeprivation die Intonation abflacht, sowie Wortschwalle bzw. Wortketten auftraten, die durch längere Pausen unterbrochen wurden. Lautstärkebezogene Veränderungen konnten hingegen nicht bestätigt werden. Eine Schwachstelle der Untersuchung ist die geringe Probandenzahl, sowie die lediglich perzeptive Analyse der Sprache.

Whitmore und Fisher (1996) untersuchten den Einfluss von 36-stündiger Schlafdeprivation und Arbeitsbelastung auf Parameter wie Tonhöhe und Wortdauer. Die Untersuchung zeigte, dass bei zunehmender Schlafdeprivation die Tonhöhe signifikant abnahm und die Wortdauer signifikant anstieg. Des Weiteren wird ein Zusammenhang zwischen verringerter Motivation und abnehmendem Muskeltonus angenommen, wodurch die Tonhöhe sinkt (Hockey, 1986). Haslam (1981) dokumentierte eine schläfrigkeitbedingte Verlangsamung der Sprache. Aus schläfrigkeitsinduzierten Wortwiederholungen und langsamerer Sprache resultiert vermutlich eine verlängerte Wortdauer (Fowler & Housum, 1987; Raghavan, 2006).

Zusammenfassend lassen sich folgende schläfrigkeitssensitiven sprachlichen Merkmalen identifizieren: eine verwaschene Aussprache, Wortwiederholungen, von langen Pausen getrennte Wortschwallen, langsames oder murmelndes Sprechen, eine abgeflachte Stimme, eine niedrigere Grundfrequenz (F_0) sowie eine geringe spektrale Hochfrequenzenergie.

Zu den wichtigen neueren Untersuchungen zählt die Studie von Greeley et al. (2006), die in einer 39-stündigen Schlafdeprivationsstudie spezifische Phone mittels automatischer Spracherkennung (ASR) segmentiert und Mel Frequency Cepstral Coefficients (MFCCs) sowie Formantfrequenzen extrahiert. Es wurden jeweils vier Sprachaufnahmen am Tag vor und nach der Schlafdeprivation aufgenommen. Vor der Sprachaufnahme wurden Reaktionszeit, Schlaflatenz und der subjektive allgemeine Schläfrigkeitsgrad gemessen. Einige phonembezogene Formantfrequenzen wiesen gute Korrelationen mit dem Schlaflatenztest auf. Auch die Konsonanten [t] und [p] zeigten gute Zusammenhänge zu den berechneten MFCCs. Unklar bleiben in dieser Studie jedoch eine Reihe wichtiger Fragen, die zur Beurteilung der Aussagekraft der Validitätsergebnisse notwendig wären, wie z.B. welche genauen Maße als Validierungsanker Eingang gefunden haben (Mittelwert, 10tes oder 90tes Perzentil des Reaktionszeittests?). Vor allem aber fehlt die genaue Beschreibung der sprachlichen Referenzeinheiten (zu vermuten sind framebasierte Einheiten), auf die sich die hohe Zahl der berechneten Formantfrequenzen bezieht.

Auch Raghavan (2006) untersuchte mittels automatischer Spracherkennung schläfrigkeitsinduzierte temporale und spektrale Variationen in der Sprache. Im Unterschied zu Greeley et al. (2006) wurden hier jedoch vergleichend drei Familien von Spracherkennungstechniken auf ihre Eignung für die Schläfrigkeitsdetektion untersucht: ein "Sprecher-Verifikationssystem", ein "Word-Spotting-System" und ein "Large-Vocabulary-Speech-Recognition-System". Die Ergebnisse zeigten, dass der Sprecher-Verifikationsansatz und das Word-Spotting-System keine erfolgreiche Schläfrigkeitsprädiktion erlaubten. Eine Stärke der Untersuchung von Raghavan (2006) ist die Erweiterung des Greeley Ansatzes durch die Variation des automatischen Spracherkennungssystems.

Zusammenfassend kann festgehalten werden, dass es eine solide theoretische Fundierung zur Wirkung von schläfrigkeits-veränderten kognitiven und physiologischen Einflussfaktoren auf auditiv-perzeptive und akustische Sprachmerkmale gibt. Diese Wirkungspfade sind in dem hier postulierten Kognitiv-phonetischen Mediatormodell integriert. Ferner existieren auch

eine Reihe von direkten empirischen Hinweisen auf mögliche schläfrigkeitssensitive akustische Korrelate. Hauptsächlich Ursprung dieser Hypothesen sind Untersuchungen im Gegenstandsbereich der akustischen Emotionserkennung, die einen guten sprachlichen Zusammenhang zur Aktiviertheitsdimension der Emotion nachweisen konnten.

4 AKUSTISCHER MESSPROZESS ALS MACHINE-LEARNING PARADIGMA

Die gegenwärtige Leistungsfähigkeit von Rechnern ermöglicht es heute komplexe Abläufe der menschlichen Informationsverarbeitung, wie z.B. das Erkennen von Objekten im Straßenverkehr oder die Detektion des Inhalts gesprochener Sprache zu analysieren und zu modellieren. Vergleichbar hiermit ist die in dieser Arbeit gefasste Zielsetzung, - angelehnt an die menschliche Perzeptionsleistung von stimmlicher Schläfrigkeit - über ein rechnerbasiertes und Machine-Learning-gestütztes System Schläfrigkeit aus sprachlicher Interaktion zu erfassen.

Machine-Learning und Mustererkennung. Der Begriff des Machine-Learning und der Mustererkennung fasst die mathematisch-technische Analyse von Aspekten der automatischen Auswertung und Verarbeitung von Mustern zusammen (Niemann, 2003). Der Einsatz von Verfahren der automatischen Mustererkennung im Rahmen dieser Arbeit begründet sich durch ihre Fähigkeit unbekannte Muster einer Klasse nach einer Lernphase zuzuordnen. Dabei entsprechen den *Mustern* die quantifizierten akustischen Stimmmerkmale eines Sprechers und den *Klassen* die jeweiligen tatsächlichen Schläfrigkeitszustände.

Der Messprozess der Acoustic Sleepiness Analysis orientiert sich am allgemeinen Ablauf des Machine Learnings und der Mustererkennung und untergliedert sich daher in 5 Verarbeitungsstufen (siehe Abbildung 21): (a) *Sprachaufnahme*, (b) *Vorverarbeitung*, (c) *Merkmalsberechnung*, (d) *Dimensionalitätsreduktion* und (e) *Klassifikation*. Im Anschluss an die Sprachaufnahme, folgt eine filternde und segmentierende Vorverarbeitung. Die anknüpfende Extraktion von prosodischen, stimmqualitativen und artikulatorischen Merkmalsverläufen wird von der Berechnung statistischer Merkmalsverlaufs-beschreibender Größen gefolgt. Die nachfolgende automatische Selektion und Reduktion von Merkmalen sind ein weiterer wichtiger Bestandteil zur Steigerung der Erkennensleistung. Der Einsatz von automatischen Klassifikatoren sowie ihre abschließende Kombination durch Metaklassifikatoren setzen den Schlusspunkt des Messprozederes.

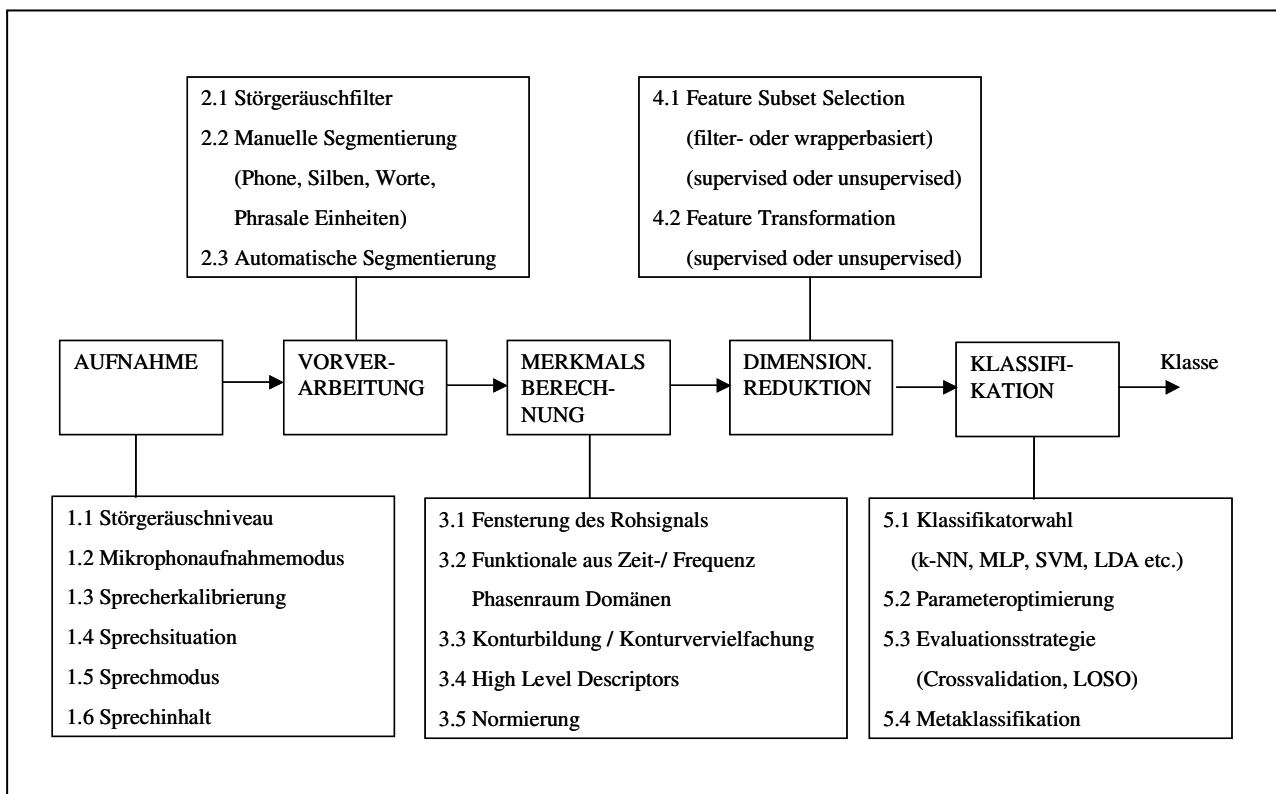


Abbildung 21: Darstellung der Lernphase des akustischen Messprozesses im Machine-Learning Paradigma

Die Entwicklung eines Machine Learning, bzw. Mustererkennungssystems besteht aus drei Phasen: (a) Lernphase, (b) Testphase und (c) Anwendungsphase (siehe Tabelle 9). Bevor ein Klassifikator zur Klassifikation eingesetzt werden kann, muss er zunächst trainiert werden. Dieses Training anhand bekannter Eingabemuster (Merkmalsvektoren, die dem Netz wiederholt zusammen mit dem Klassifikationsergebnis präsentiert werden) wird als überwachtes Lernen (Supervised Learning) bezeichnet. Das rechenzeitintensive Training des Klassifikators erfolgt in der Regel einmalig im Rahmen des Aufbaus des Analysemoduls. Nach durchgeführtem Training erfolgt ein Test des Klassifikators (ebenfalls mit bekanntem Klassifikationsergebnis), um die Güte des Klassifikators zu überprüfen. Ist die Klassifikationsqualität ausreichend, kann der Klassifikator eingesetzt werden, um unbekannte Eingabemuster zu klassifizieren (Anwendungs- bzw. Recallphase). Eine erneute Trainings- und Testphase ist erst erforderlich, wenn sich die Dimension des Merkmalsvektors (Anzahl der zu verwendenden Merkmale) durch Hinzufügen oder Selektion ändert.

Tabelle 9: Phasenbezogene Input und Outputgrößen innerhalb des Machine-Learning Paradigma

	Input der Phase	Output der Phase
Lernphase	-Merkmalsvektoren -Klassifikationslabel	-Klassifikations-Algorithmus
Testphase	-Merkmalsvektoren -Klassifikationslabel -Klassifikationsalgorithmus	-Vorhersagegüte des Modells
Anwendungsphase	-Merkmalsvektoren -Klassifikationsalgorithmus	-Klassifikationsvorhersagen

Anwendungsfelder des Machine Learning. Die Techniken des Machine-Learning (siehe auch Data Mining) sind bereits in einer Vielfalt von Anwendungsfeldern erfolgreich eingeführt, so z.B. im Banking, Biotechnologie/Genomics, Kredit Scoring, Marketing (Fokus der Marketing Bemühungen auf potenzielle Neukunden), Customer Relationship Management (Vorhersage von Kunden mit Kündigungsabsicht, auch Studienabbrecher Vorhersage), automatische Music Retrieval Systeme (Moerchen, 2005), Kreditkarten-Betrugsdetektion, Steuerbetrugsrisikogruppen, Betrug bei der Krankheitsabrechnung, Risiko Management von Versicherungen, Homeland Security („Rasterfahndung“ Detektion von verdächtigen Transaktionen), Investment und Wertpapierhandel, Spam Mail Detektion und Retailing (Geschäftsstandortanalyse, Online Produktempfehlungen, Personalrekrutierung). So werden beispielsweise Machine-Learning Verfahren in der biomedizinischen Forschung dazu verwendet, Gensequenzen zu identifizieren, die mit Krebserkrankungen assoziiert sind. In Geowissenschaften werden mittels Machine-Learning und Mustererkennungsverfahren Vorhersagemodelle für z.B. Hangabrutsch-Risiken berechnet. In der forensischen Forschung sind denkbare Anwendung die Identifikation von Tätermerkmals- und Tatvorgangsmustern zur Kriminalitätsprävention und –aufklärung (Mokros, 2006). Im Audio Processing Bereich werden user-spezifischen Präferenzen gelernt und selbständig auf große Musik Datenbanken angewendet (Mierswa, 2004). In der akustischen Emotionsforschung werden aus Sprachdaten und Computer-Mouse Bewegungen aktuelle Befindlichkeiten des PC-Nutzers bestimmt (Schuller, 2006). Im privatwirtschaftlichen Einsatz finden Data Mining Verfahren in der Finanzwirtschaft und in Versicherungsunternehmen Einsatz, um Betrugs- und Risikoabschätzungen (Credit Scoring) zu treffen. So können z.B. Kundendaten eingesetzt werden, um die Betrugswahrscheinlichkeit einer Kreditkartentransaktion abzuschätzen. Im Marketing und Re-

tailingbereich werden Analysen zur Identifikation von erfolglosen Kundenbindungsstrategien durchgeführt (Cahlink, 2000).

4.1 Sprachaufnahme

Vergleichbar zu den oben genannten Einsatzfeldern des Machine Learnings, wird auch in der hier vorgestellten Arbeit, das allgemeine fünf-stufige Mustererkennungsprozedere angewendet und für die für Klassifikationsaufgabe der stimmbasierten Schläfrigkeitvorhersage nutzbar gemacht. Das in Innerhalb der Sprachaufnahmephase lassen sich verschiedene Variationen des Aufnahmeumfeldes und des Sprechinhalts unterschieden. Je nach Zielsetzung des Messprozesses z.B. erster laborexperimenteller Proof-of-Concept Nachweis oder aber realistische Abbildung aller Rahmenbedingungen eines Anwendungsszenarios, sollten die Parameter der Aufnahmeumfeldes und Sprechinhalts gewählt werden.

Aufnahme-Rahmenbedingungen. Zu unterscheiden sind die in Tabelle 10 aufgeführten Parameter (a) *Störgeräuschniveau* (z.B. Hintergrund-Lautstärke, simultane Nebensprecher, Motorengeräusche, Hall in kleinen Räumen), (b) *Mikrophonaufnahmemodus* (z.B. Digitalisierungsauflösung des Analog-Digitalwandlers, Aufnahme über Freisprecheinrichtung, Funkverkehr- oder Telefonqualität) und (c) *Sprecherkalibrierungsmöglichkeiten* (Gelegenheit vielfache Sprachtrainingssamples aufzunehmen, mit oder ohne Kenntnis der Schläfrigkeit). Hinsichtlich der Mikrophonaufnahmequalität sollte berücksichtigt werden, dass die Quantisierung der Sprachaufnahme in Abhängigkeit von der Abtastfrequenz (z.B. 22.05 kHz = 22050 abgetastete Messwerte pro Sekunde) und des Auflösungsgrades (z.B. 8 Bit Auflösung = 256 unterschiedliche Amplitudenwerte) allen Sample Points Spannungswerte zuordnet. Die Abtastfrequenz sollte hierbei größer als das Doppelte der höchsten zu analysierenden Signalfrequenz sein. Im Falle der menschlichen Sprache mit ca. 5000Hz Maximalfrequenz also mindestens 10 kHz Abtastfrequenz. Für ein Sprachsignal in Telefonqualität hingegen, das Frequenzen bis 3.4 kHz überträgt, reicht auch schon eine Abtastfrequenz von 8 kHz.

Tabelle 10: Determinanten der Realitätsnähe von Sprachaufnahmen

	Realitätsnähe-Faktoren	Beispiel
Aufnahmerahmenbedingungen	(a) Störgeräuschniveau (b) Mikrophonenaufnahmemodus (c) Sprecherkalibrierungsmöglichkeiten	(a) simultane Nebensprecher, Motoren-geräusche (b) Funkverkehr- oder Telefonqualität, Freisprecheinrichtung (c) vielfache Sprachtrainingssamples pro Person
Sprechinhalt und Sprechmodus	(a) Sprechsituation (b) Sprechmodus (c) Sprechinhalt	(a) Mensch-Computer Kommunikation, Monolog, informelle Sprechsituation (b) Lese- oder Spontansprache, Artikulations- u. Lautstärkemodus (c) Phonem-, Silben-, Wortreihen, Phrasale Einheiten

Sprechinhalt und -modus. Bezüglich des Sprechinhalts können folgende Parameter differenziert werden (a) *Sprechsituationen* (z.B. Mensch-Computer Kommunikation, informelle Sprechsituation, Monolog, Mental Workload induzierenden Nebenaufgaben; vgl. Tabelle 11), (b) *Sprechmodus* (Lese- oder Spontansprache; Artikulationsklarheits-, Lautstärke-bezogene Sprechmodusvariationen, z.B. betont deutliche, flüsternde, modale, laute oder schreiende Stimme) und (c) *Sprechinhalt* (Beschränkung des verwendeten Sprechinhalts auf vorgegebenes, restringiertes Sprachmaterial). Bezüglich des eingesetzten Sprachmaterials sind als linguistische Komplexitätsstufen das Sprechen von Phonen (z.B. der Langvokal [a:]), Silbenlisten (Vokal-Konsonant-Vokal, Konsonant-Vokal-Konsonant Einheiten), Wortreihen und phrasale Einheiten (siehe auch Intonation Unit, Prosodic Unit, Chunks, Tunes, Turnes) zu nennen. Gesprochen werden können Wortreihen entweder im Lesemodus oder aber - verknüpft mit Denk- und Rechenaufgaben - im spontansprachlichen Modus (Wochentage rückwärts aufzählen, Lösung arithmetischer Aufgaben sprechen). Ähnliche hinsichtlich Lese- oder Spontansprache liegende Variationen sind auch auf der Ebene ganzer Äußerungen möglich, hier können z.B. phonetisch ausbalancierte Texte gelesen werden („Der Kater hat den Biber gefangen“, siehe Siegert, 2004; Die Sonne und der Nordwind), Bilder beschrieben werden, autobiographische Monologergählsituationen oder auch spontansprachliche Dialogsituationen Verwendung finden. Beispielhaft seien hier einigen Sprachmaterialien aufgeführt:

- Eckvokal ausbalancierter Lesetext „Der Kater hat den Biber gefangen“

„Der Kater hat den Biber gefangen.
Diesen Biber hat der Kater gefangen.
Die Dame hat die Nudeln gegessen.
Diese Nudeln hat die Dame gegessen.
Timo hat den Boden gefegt.
Diesen Boden hat Tim gefegt.
Der Bube hat eine gute Note bekommen.
Diese Note hat der Bube bekommen.“
- Zungenbrecher-Lesetexte

„Dicke Nichten dichten im dichten Fichtendickicht.
Im dichten Fichtendickicht dichten dicke Nichten.“

„Weil die Paten um den Braten baten,
taten Puten wir den Paten braten.
konnten Paten Putenbraten bieten,
weil wir Puten - wie die Paten baten - brieren.“
- Wortliste aus Freiburger Sprachverständlichkeitstest (nach DIN 45621)

„Fink Schlauch Reh Grad Floß Hirn Fuchs Bein Napf Teer Stolz Art Wurm Ding Trab Bett Kleid
Schatz Wut Pflock“

„Schnitt Frau Land Helm Bock Flucht Scherz Keil Rast Gruß Wohl Plan Krieg Ast Pfiff Weib Sturm
Fang Tee Mord“

„Furcht Schlitz See Schar Gold Leib Wunsch Fraß Stier Ton Heer Dachs Bauch Kreuz Akt Pfund Sekt
Glück Molch Rad“

„Fleisch Welt Rohr Park Flut Gieß Saum Krebs Hand Gott Schuh Film Damm Zelt Koch Hanf Leid
Bier Spruch Axt“
- Silbenreihe

“bok, pok, dok, tok, gok, kok, koka, koga, koba, kopa, koda, kota”

Sprachkorpora (Datenbanken mit Sprachsamples). Als Alternative zu der eigenständigen Aufnahme von Sprachmaterial kommt der Rückgriff auf Sprachdatenbanken in Frage. Sie beinhalten sowohl Sprachaufnahmen als auch die dazugehörigen validierten Emotionszuordnung (Annotationen). Datenbanken mit emotionalen Sprachbeispielen sind im Rahmen des Humaine Projekts kategorisiert worden (Campbell, 2002; Douglas-Cowie, et al., 2000; Ververidis & Kotropoulos, 2003). Wichtige Datenbanken sind AEC (Batliner et al., 2004), Sympafly (Steidl et al., 2004), EMO-DB (Burkhardt et al., 2005), IFA (Van Son et al., 2001) und DES (Engberg & Hansen, 1996). Güteigenschaften dieser Datenbanken bzw. des verwendeten Sprachmaterials sind: (a) Stichprobengröße (hohe Gesamtzahl an Lern- und Testbeispielen): In der Regel ist mit steigender Zahl an Lernbeispielen auch eine höhere Gesamtleistung eines Erkennungssystems auf Grund exakterer Modellierung der Problemstellung zu erwarten. Um die Bevorzugung einer überrepräsentierten Klasse zu vermeiden, sollten für alle Kategorien gleichmäßig verteilte Daten vorliegen; (b) ökologische Validität: Die Sprechaufnahmen sollen der intendierten Zielpopulation und Anwendungssituation entsprechen. Eine angepasste Diversivität bezüglich der Altersklassen, Geschlecht, Bildungsgrad, kulturellen und sozialen Hintergründen vermeidet die Über- oder Unterschätzung der berechneten Zielbefindlichkeits-Detektionsraten; (c) Validität der Validierungsanker: Die zur Validierung häufig herangezogenen Perzeptionstests sollen über oder eine möglichst hohe Zahl von individuellen Annotatoren oder Experten abgesichert werden. Berücksichtigt werden sollte auch die Validitätsproblematik von geschauspielerten Emotionen.

Die für die Validierungsphase notwendigen schläfrigen Sprachbeispiele können derzeit lediglich selbst erhoben werden, da keine frei zugänglichen Schläfrigkeitsdatenbanken existieren. Bezüglich der eigentlichen Anwendungsphase der akustischen Schläfrigkeitsmessung stellt sich die Frage, wie sichergestellt werden kann, dass die nötigen Sprachinputs regelmäßig zur Verfügung stehen. Eine inhärente Schwierigkeit der sprachgestützten Schläfrigkeitsmessung ist die regelmäßige Gewinnung von Sprachäußerungen. Im Fahrzeugkontext ein regelmäßiger (z.B. 30-minütiger) Anfragezyklus durch ein „In Car Speech Dialog Systems“ mit Fragen zum Innenraumklima („Ist die Innentemperatur angenehm?“) oder Unterhaltungssystembezogenen Anliegen („Möchten Sie die Staunachrichten hören“) denkbar. Bei detektierter sicherheitskritischer Schläfrigkeit kann dieser Zyklus entsprechend auf fünf Minuten Zyklen reduziert werden. Potenziell aufkommende Sprechsituationen sind in der nachfolgenden Darstellung aufgelistet.

Tabelle 11: Sprechsituationen im Fahrzeugkontext unterteilt nach Empfänger und Komplexität des Sprachinhalts

	Empfänger: System	Empfänger: Mensch
Monolog-Situation	-Einfache sprachgesteuerte Fahrerassistenzsystem Eingaben	- Selbstgespräche, Befluchen von Verkehrsteilnehmern
Dialog-Situation	- Komplexe sprachgesteuerte Fahrerassistenzsystem Eingaben („In Car Speech Dialog System“)	-Telefongespräche -Mitfahrer-Kommunikation

4.2 Vorverarbeitung

Messfehlerdetektion und Korrektur. Im Rahmen der Datenvorverarbeitung erfolgt zunächst eine Aufbereitung der Rohdaten. Dabei werden eventuelle Fehler und Lücken in den Rohdaten identifiziert (z.B. mittels Plausibilitätsprüfung). Diese können beispielsweise durch Messfehler in der Datenerfassung, durch den Ausfall von Detektionseinrichtungen oder durch Unterbrechungen des Datenübertragungsstroms zwischen der Erfassung und der Verarbeitung im System entstehen. Mit einem den möglichen Wertebereich der Absolutwerte und Veränderungswerte des Messobjektes berücksichtigenden Filter können unplausible, fehlerhafte Daten identifiziert werden. Die Messfehlerbehandlung wiederum nimmt die vollständige Elimination oder aber auch die Interpolation des Messsignals vor. Ferner kann über eine Filterung ganzer Frequenzbereiche das Signal-Noise-Ratio in relevantem Maße verbessert werden. Denkbar ist hier eine Filterung Nebengeräusch-spezifischer Quellen wie z.B. Türengeräusche im Untersuchungsraum oder Fahrtgeräusche im Auto durch einen auf diese Frequenzen eingerichteten Bandpass-Filter.

Segmentierung. Bevor die akustischen Kennzahlen bestimmt werden können, muss die Analysegranularität, also das Aggregationsniveau des Sprachsignals festgelegt werden. Welche sprachlichen Segmente wie Phone, Silben, Worte oder Intonationsphrasen (Äußerungen, Tunes) sind geeignet, um den Zielzustand maximal präzise zu erfassen? Das Segmentieren der „Units of Interest“, der relevanten sprachlichen Abschnitte kann manuell oder automatisch erfolgen. Bei der im Falle eines Echtzeitbetriebs erforderlichen automatischen Segmentierung wird die Untergliederung über die Detektion von Sprechpausen (Intonationsphrasen- Ebene) oder mittels Automatic Speech Recognition (ASR) Systeme (Phonem-, Silben-, Wort-Ebene) erzielt. Relevante sprachliche Einheiten können Phone, Silben (Konsonant-Vokal-Konsonant oder auch Vokal-Konsonant Einheiten), Wort oder Phrasale Einheiten sein. Bei den linguistischen Spracheinheiten wiederum lassen sich zusätzlich auch Positionen innerhalb der übergeordneten Spracheinheit differenzieren. So können sich z.B. Silben in unterschiedlicher Wortposition (Anfangs-, Mitten-, Endposition) oder Worte in unterschiedlicher Phrasaler Einheit Position befinden. Darüber hinaus kann auch das Merkmal der betont bzw. unbetonten Lage von Silben von Interesse sein. Auch verschiedene Satzarten wie Frage, Aussage, Befehlssätze können zusätzlich potenziell schläfrigkeitssensitive relevante Einheiten definieren.

Das Selektieren der relevanten sprachlichen Abschnitte kann manuell oder automatisch erfolgen. Bei der im Falle eines Echtzeitbetriebs erforderlichen automatischen Segmentierung wird die Untergliederung über die Detektion von Sprechpausen (Intonationsphrasen-Ebene) oder mit Hilfe von Automatic-Speech-Recognition-Systemen (Phonem-, Silben-, Wort-Ebene) erzielt. Eine vergleichende Darstellung manueller und automatisierter Segmentierung zeigt Abbildung 22.

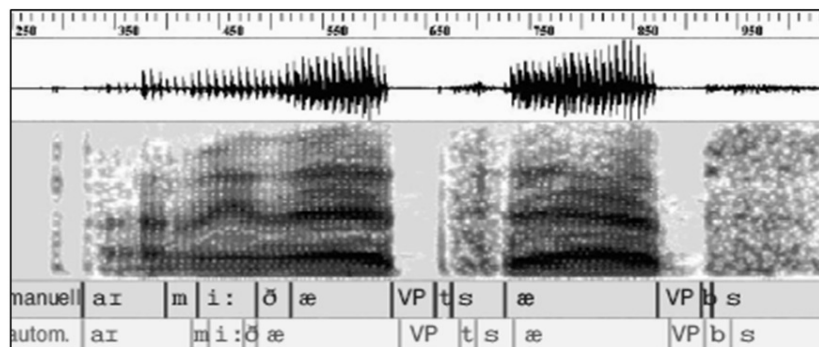


Abbildung 22: Manuelle und automatische Segmentierung des Satzes „I mean thats abs...“. Abgebildet sind die Waveform, das Spektrogramm und die dazugehörigen manuellen und automatischen Annotationszeilen

Zielsprecher Stimmaktivitätserkennung. Wenn spontane Sprachaufnahmen der Benutzerschnittstelle ohne manuelle Tasteninitiiierung im offenen Mikrofonbetrieb aufgezeichnet werden sollen, ist eine Stimmaktivitätserkennung erforderlich, um die Stimmaktivitätsphasen des Zielsprechers aus dem kontinuierlichen Signalstrom von Nebengeräuschen und Fremdsprecheranteilen zu identifizieren. Die Eliminierung oder Verminderung von Störungen und Nebengeräuschen (z.B. Motor-, Wind-, Radio- und Fahrtgeräusche einer Autobahnfahrt) stellt auch in dieser Analysephase eine wichtige Optimierung des Signal-Noise-Ratios und der Identifikationsleistung der Zielsprecher-Stimmaktivität dar.

4.3 Merkmalsberechnung

Der operative Teil der Merkmalsberechnung erfolgt über das (a) Framing und Windowing, (b) Berechnung von fensterbezogenen Einzelkennwerten, (c) Fusionierung der fensterbezogenen Einzelkennwerten zu geglätteten Merkmalskonturverläufen, (d) Berechnung von Konturverlauf beschreibenden Funktionalen, (e) Normalisierung und (f) automatischen Generierung von Merkmalen über Genetic Algorithms (Mierswa, 2004; Schuller, 2006).

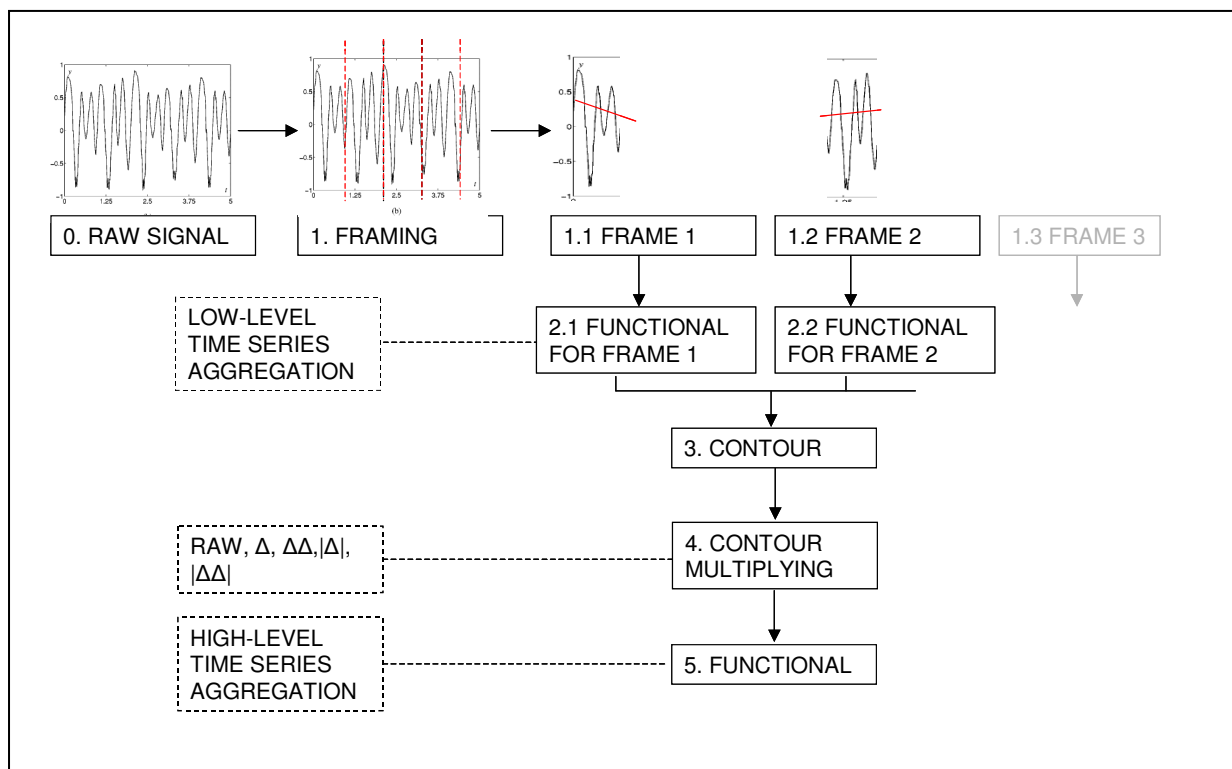


Abbildung 23: Ablauf des Merkmalsberechnungsprozesses über Framing des Rohsignals, Berechnung von Funktionalen für diese Frames, Aggregation dieser Funktionalen zu Konturen, Vervielfältigung der Konturen über erste und zweite Ableitungen, Berechnung von Funktionalen zur Aggregation dieser Verläufe

Framing und Windowing. Das Framing des Rohschallsignals, teilt das Signal in sogenannte *Signalrahmen* auf. Die gewählte Länge eines Fensters muss dabei eine Mindestdauer aufweisen, um eine ausreichende Wertauflösung transformierter Verläufe zu erlauben (ca. 20 ms.). Wählt man die Fensterweite groß, erzielt man zwar eine gute Genauigkeit der berechneten Fensterkennwerte, die zeitliche Auflösung kann jedoch so stark reduziert sein, dass wichtige Informationen der Signaldynamik verloren gehen. Um die Auswirkungen des abrupten Starts und Ende auf die Ergebnisse der Spektralanalysen zu reduzieren, wird das Signal vor der Anwendung der Spektralanalyse an den Rändern des Fensters abgeschwächt (*Fensterung, Windowing*), in dem das Signal über die gesamte Länge des Fensters mit einer Fensterfunktio-

on multipliziert wird. In der Mitte nimmt diese Funktion den Wert 1 an und gleitet ähnlich einer Normalverteilung zu beiden Seiten gegen Null. Die in der Sprachsignalverarbeitung verwendeten Fensterfunktionen sind die Bartlett-, Blackman-, Kaiser-, Welch-, Hanning- und Hamming-Funktion. Die am häufigsten eingesetzte Fensterfunktion ist die Hamming-Fensterfunktion (siehe Abbildung 24).

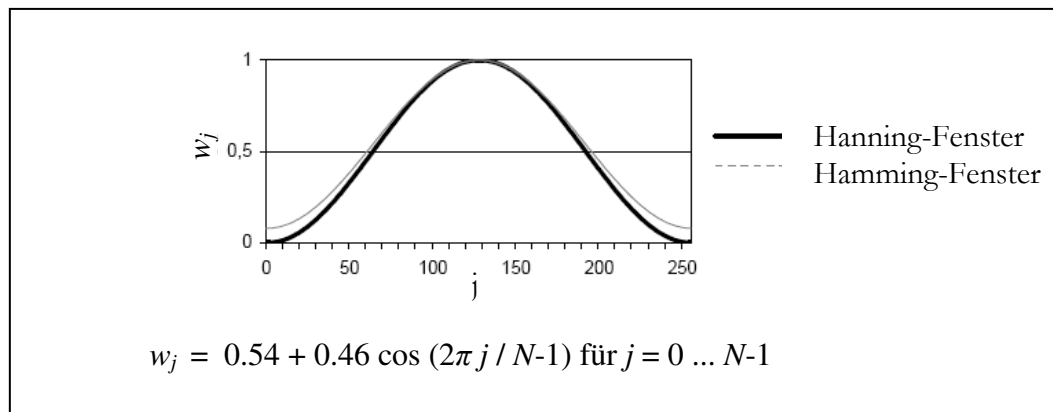


Abbildung 24: Fensterfunktionen Hamming und Hanning für die Fensterbreite $N=256$

Berechnung von fensterbezogenen Sprachkennwerten. Für jedes Signalfenster werden eine Reihe von Sprachsignalkennwerten mit der Sprachanalyse-Software Praat berechnet. Diese Kennwerte sind: mittlere quadrierte Amplitude (Intensität), Fundamental Frequenz (Tonhöhe, F_0), Lage der Energie-Maxima im spektralen Amplitudenspektrum (Formanten Frequenzen, F_1 - F_6), Breite der Energie-Maxima im spektralen Amplitudenspektrum (Formant Bandwidth, F_1 - F_6), Zentroid der spektralen Energieverteilung, 36 MFCC Mel Frequency Cepstral Coefficients (sowie 12 Δ und 12 $\Delta\Delta$ MFCC), kleine Schwankungen der Tonhöhe (Jitter), kleine Schwankungen der Intensität (Shimmer), relative Stärke einzelner Frequenzbänder (z.B. HF 500), Regressionssteigung der Energieverteilung im Frequenzband über 1 kHz (Spectral Slope), Anteil der periodischen Signalenergie (Harmonics-to-Noise Ratio). Im Folgenden werden die wichtigsten Kennwerte näher erläutert. Einen guten Überblick zu möglichen Low-level Zeitreihen Features (Low-level Descriptors, LLE) gibt Schuller et al. (2007) und für den Audio Processing Bereich Mörchen et al. (2006).

Tabelle 12: Überblick zu weitergehenden möglichen Low-level Zeitreihen Features (vgl. Mörchen et al., 2005)

Low-level Feature	Anzahl der Features	Referenz
Intensity	1	Schuller, 2006
Zerocrossing	1	Schuller, 2006
Fundamental Frequency	1	Schuller, 2006
Jitter	1	Schuller, 2006
Shimmer	1	Schuller, 2006
Harmonics-to-noise Ratio	1	Schuller, 2006
Spectral Centroid	1	Bindreiter, 2006; Li et al., 2001
Spectral Bandwidth	1	Li et al., 2001
Spectral Rolloff	1	Li et al., 2001
Spectral Crest Factor	1	Jayant & Noll, 1984
Spectral Flatness Measure	1	Jayant & Noll, 1984
Spectral Flux	1	Tzanetakis & Cook, 2002
Spectral Regression Slope	1	Mierswa & Morik, 2005
Spectral Regression Y Intercept	1	Mierswa & Morik, 2005
Spectral Regression Maximum Error	1	Mierswa & Morik, 2005
Spectral Regression Median Error	1	Mierswa & Morik, 2005
Frequency Band Mel Magnitude	36	Mierswa et al., 2005
Band Energy Ratio	10	Li et al., 2001
Formant Frequenzen (F1-F6)	6	Schuller, 2006
Formant Amplitude (F1-F6)	6	Schuller, 2006
Formant Bandwidth (F1-F6)	6	Schuller, 2006
Formant Differenzen (F1-F6)	21	Schuller, 2006
Linear Predictive Coefficient	16	Rigoll, 2004
Perceptual Linear Prediction	18	Zeppelzauer, 2005
Mel Frequency Cepstral Coefficient	12	Tukey, Bogert & Healy, 1963
Linear Frequency Cepstral Coefficient	12	Rigoll, 2004
Summe	159	

Fusionierung der fensterbezogenen Kennwerte zu Konturverläufen. Aus den pro Fenster berechneten Einzelkennwerten werden neue Konturverläufe gebildet. Exemplarisch dargestellt werden in der unten dargestellten Abbildung 25 die Konturverläufe für die Formantfrequenzen (F1-F4).

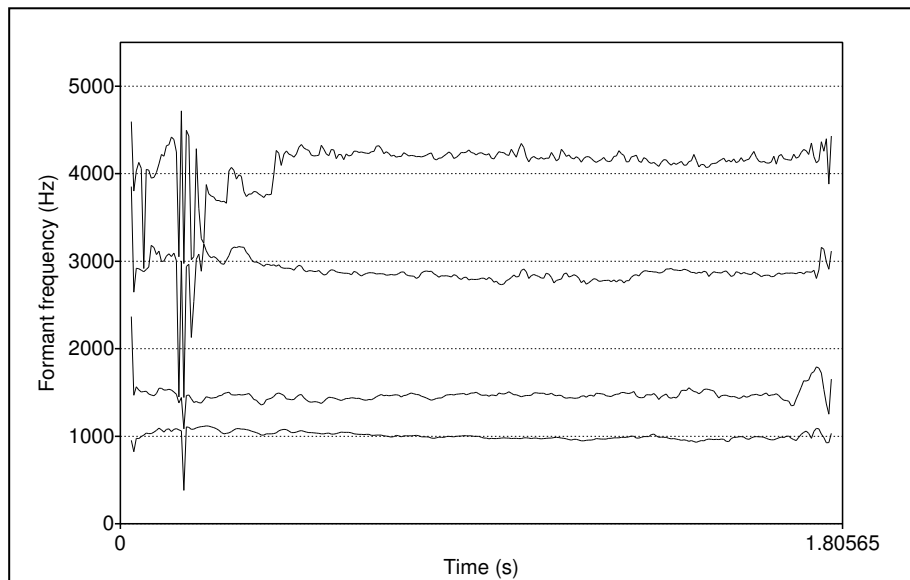


Abbildung 25: Beispielhafter Konturverlauf der Formantfrequenzen F1 bis F4

Konturglättung. Ein weiterer Arbeitsschritt im Rahmen der Merkmalsberechnung ist die Glättung der neu berechneten Konturverläufe, um Elemente im Zeitreihenverlauf zu eliminieren, die für die Klassifikation unwesentlich sind. Unter Glättung wird allgemein die Reduktion von irregulären Schwankungen einer Zeitreihe verstanden. Bei der Erfassung von Signalen und der Ableitung transformierter Verläufe sind durch Quantisierung und Fensterung veräuschte oder fehlerhafte Daten unvermeidbar, wodurch sich eine Abweichung einzelner gemessener oder berechneter Werte von ihren realen Werten ergeben kann. Diesem Aspekt kann durch eine Glättung von gebildeten Konturverläufen Rechnung getragen werden. Beim Einsatz von Glättungsverfahren sind jedoch die Folgen der Veränderung der Rohdaten abzuwägen, da beispielsweise einige mathematische Verfahren (z.B. Moving Average-Verfahren oder Verfahren der exponentiellen Glättung) wichtige Charakteristika (Extrema) im Zeitreihenverlauf glätten, und somit das Ergebnis verfälschen können (Wild, 1996). Eine Glättung des zeitlichen Verlaufs kann z.B. durch eine Überlappung der Signalrahmen gewährleistet werden.

Denkbare Erweiterungen der konturbildenden Methodik ist die Vervielfältigung der berechneten Low-level Feature Zeitreihen mittels erster und zweiter Ableitung (Delta, Delta Delta Differenzverläufe) sowie deren Absolutwerte (Betrag von Delta, Betrag von Delta Delta Differenzverläufen). Dieser Schritt beinhaltet prinzipiell großes Potenzial, die in den Datenverläufen liegende Information optimal auszuschöpfen.

Berechnung von Konturverlauf beschreibenden Funktionalen. Nach der Konturextraktion werden im nächsten Schritt die letztendlich zur Vorhersage des Schläfrigkeitstages verwendeten Kennwerte berechnet. Hierzu werden Funktionale aus dem (a) Zeit-Bereich, (b) des Frequenz-Bereichs und (c) des Phasen-Bereichs bestimmt. Im Zeitbereich sind diese deskriptiven Größen z.B. Maximum, Minimum, Range, lineare Momente erster bis vierter Ordnung (Mittelwert, Standardabweichung, Kurtosis, Schiefe), erste bis dritte Quartilposition, Regressionsgradensteigung und Regressionsfehler. Einen Überblick zur möglichen Vielfalt Konturverlaufs-beschreibender Funktionale liefert Tabelle 12.

Die einfachste statistische Aggregation sind die linearen Momente erster bis vierter Ordnung (Mittelwert, Standardabweichung, Kurtosis, Schiefe). Da diese Statistiken nicht robust gegenüber extremen Werten sind, können vor der Momentenberechnung auch noch die höchsten und niedrigsten 2.5% eliminiert werden. Um die temporäre Struktur noch besser zu erfassen, können diese Kennwerte auch auf die erste und zweite Ordnung Differenzen angewendet werden. Um die Korrelationsstruktur zu erfassen, können Autokorrelation Funktionen (ACF) und Partial Autokorrelations-Funktionen (PACF) mit einem Lag bis 20 Datenpunkte berechnet werden. Des Weiteren beschreiben regressionsstatistische Kennwerte den Verlauf der ACF. Über das spektrale Verhalten der Konturverläufe (Low-level Feature Zeitreihe) werden der Spectral Centroid, Bandwidth, Cepstrale Koeffizienten Parameter geschätzt. Wie in McKinney und Breebaart (2003) beschrieben, kann die Modulations-Energie in drei Frequenzbändern (1-2Hz, 3-15Hz und 20-43Hz) gemessen werden. Die absoluten Werte werden komplettiert durch die an der Gesamtmodulationsenergie normierten relativen Band-Stärken.

Tabelle 13: Überblick zu möglichen verlaufsbeschreibenden Kenngrößen (Funktionalen) (siehe auch Functionals, vgl. Mörchen et al., 2005)

High-level Feature	Anzahl der Features	Referenz
Lineare Momente (MW, SD, Schiefe, Kurtosis)	4	Mierswa, 2005
Robuste lineare Momente	4	Mierswa, 2005
Extrema Position und Wert	6	Schuller, 2006
Regression Slope	1	Schuller, 2006
Länge von Schwellenwertüberschreitungen	2	Zeppelzauer, 2005
Regression Y Intercept	1	Schuller, 2006
Regression Maximum Error	1	Schuller, 2006
Regression Median Error	1	Schuller, 2006
Spectral Centroid	5	Bindreiter, 2006; Li et al., 2001
Spectral Bandwidth	1	Li et al., 2001
Spectral Regression Slope	1	Mierswa & Morik, 2005
Spectral Regression Y Intercept	1	Mierswa & Morik, 2005
Spectral Regression Maximum Error	1	Mierswa & Morik, 2005
Spectral Regression Median Error	1	Mierswa & Morik, 2005
Cepstrum Coefficient	36	Rigoll, 2004
Autokorrelation Lag	10	Mierswa et al, 2005
Autokorrelation Slope	1	Mierswa et al, 2005
Autokorrelation Cut	1	Mierswa et al, 2005
Phase Space- Moments Distances	4x10	Mierswa et al, 2005
Phase Space- Moments Angels	4x10	Mierswa et al, 2005
Phase Space – Lyapunov	1	Mierswa et al, 2005
Summe Funktionale	159	
Summe Low-Level Features	159	
Summe High-Level Features	25281	

Non-lineare Analysen der Zeitreihe sind eine zur korrelativen und spektralen Analysen alternative Beschreibungsform der Zeitreihe (siehe Abbildung 26). Die Rekonstruktion des Phasenraums (Takens, 1981) kann verwendet werden, um die zeitliche Struktur der Zeitreihe abzubilden (Mierswa and Morik, 2005). Mittelwert und Standardabweichung der Distanzen und Winkel im Phasenraum mit variierenden Embedding Dimensionsgraden und Time Lags können beispielsweise verwendet werden (siehe Abbildung 27). Entgegen einer gebräuchlichen Schätzung des Time Lags über das Maximum der Autokorrelationsfunktion (Lindgren et al., 2004) lassen sich diese auch einfach von 1 bis 10 durchvariieren. Andere Maße, die zur Beschreibung des Phasenraumes dienen, wie der höchste Lyapunov Koeffizient, Fraktal Dimension, approximierte Entropie (Beckers, 2002) oder Delay Vector Variance (Golz et al. 2007) können ebenfalls in Erwähnung gezogen werden.

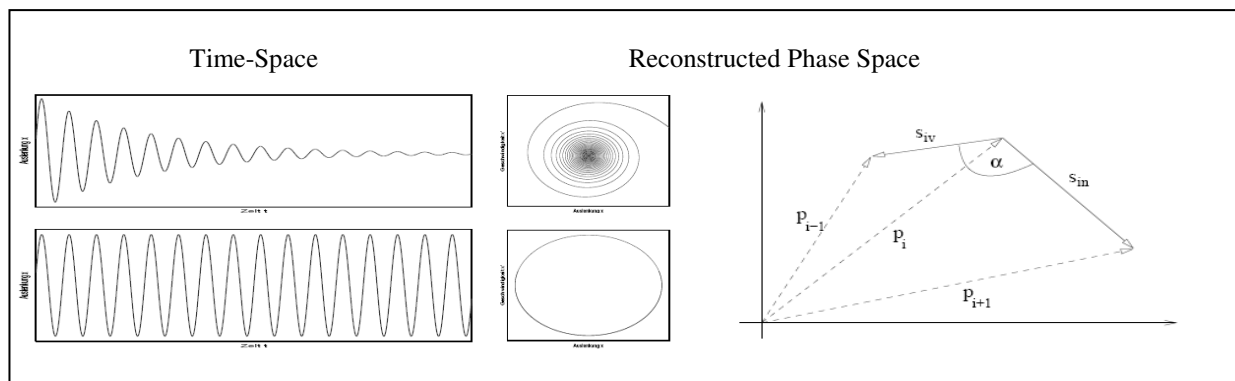


Abbildung 26: Transformationen einer Sinusschwingung und einer gedämpften Sinusschwingung in den Reconstructed Phase Space (links). Illustration dreier Punkte im Reconstructed Phase Space und der aus ihnen gebildeten Trajektorenschenkel (rechts)

Das zuvor beschriebene Prozedere der Merkmalsberechnung von Framing und Windowing, Berechnung von fensterbezogenen Einzelkennwerten, Fusionierung der fensterbezogenen Einzelkennwerten zu geglätteten Merkmalskonturverläufen und Berechnung von Konturverlauf beschreibenden Funktionalen (High-level Features) liefert für das hier aufgestellte Beispiel 137 fensterbezogenene Konturverläufe mal 157 Funktionalen gleich 21509 (!) High-level Features zur Beschreibung eines Sprachabschnitts. Die Nutzung der angesprochenen Vervielfältigung über Delta und Delta Delta Verläufe resultiert in einem Merkmalsset von über 100000 akustischen Sprachmerkmalen.

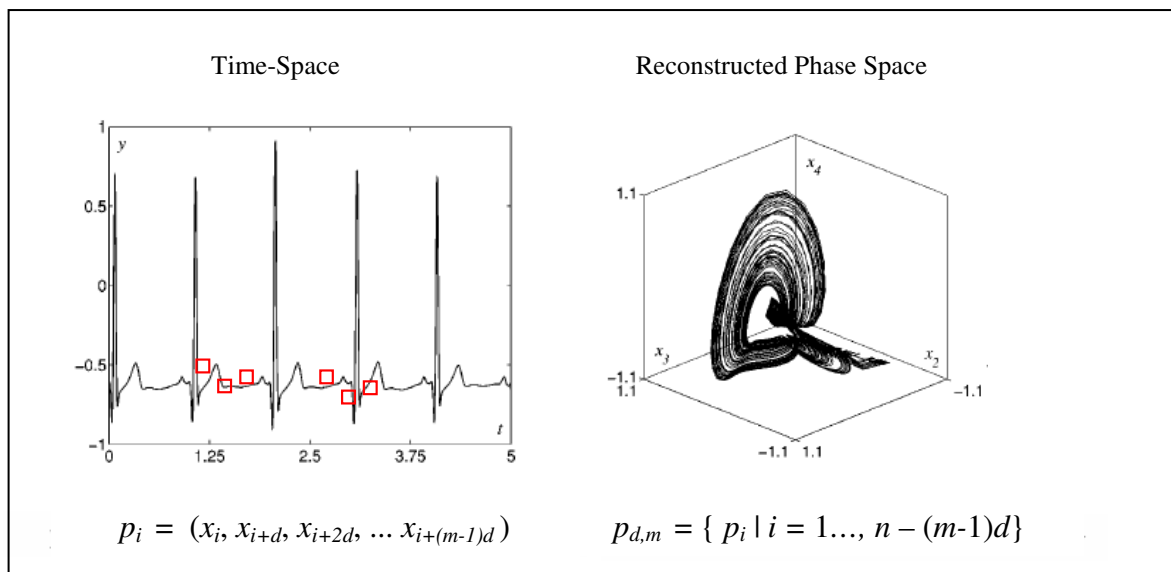


Abbildung 27: Transformation eines EKG-Signals im Zeitbereich in den dreidimensionalen Phasenraum (vgl. Feo, 2003)

Normalisierungen. Einige Kennwerte sind vor allem aussagekräftig, wenn sie am Referenzwert des Mittelwertes (aber auch Maximum, Minimum oder Range) der (a) aktuellen Aufnahme des Sprechers A, (b) des Referenzwertes der durchschnittlichen Aufnahme des Sprechers A oder aber (c) am Referenzwert aller aufgenommenen Sprecher korrigiert werden. Die häufigste Normalisierung ist die über alle Sprecher und Aufnahmen gemittelte Intensitätsnormalisierung. Sinnvoll ist diese Normalisierung, wenn aufgrund von Mikrofonabstandsvariationen verzerrte Intensitätsabsolutwerte berechnet werden könnten. Eine darüber hinaus häufige Normalisierung ist die sprecherspezifische z-Normalisierung. In ihr dient der Mittelwert aller Aufnahmen von Sprecher A als Referenzwert. Vorsicht ist jedoch geboten, wenn die Übertragbarkeit von normalisierten Kennzahlen-Ergebnissen dadurch gefährdet wird, dass ungleiche Müde-Wach-Sprachbeispielquotienten bei der Berechnung von sprecherspezifischen z-Normalisierungen verwendet werden. Sinnvoller kann es daher sein, nicht den Mittelwert aller Sprachproben (der vom Mischungsverhältnis abhängt), sondern vielmehr den Mittelwert der wachen Sprachproben als Referenzwert zu nutzen.

Automatische Merkmals-Generierung mittels Genetischer Programmierung. Evolutionäre Algorithmen versuchen, in einem Suchraum den optimalen Suchpunkt für die Lösung einer Aufgabe zu finden (Mierswa, 2004). Ähnlich dem biologischen Vorbild werden in einer Vielzahl von Generationen nach dem Prinzip der natürlichen Selektion, also dem „Survival of the Fittest“ über Mutation und Kreuzung selbständig neue Kennzahlen berechnet (siehe Abbildung 28). Diese neuen Kennzahlen werden im Trial-and-Error Verfahren über einfache arithmetische Operatoren (+, -, *, /, 1/x, sin, cos, tan, exp, log) aus den bestehenden berechnet. Verbessert sich in einer Generation durch die Hinzunahme dieser neuen Kennzahl die Gesamtvorhersageleistung, verbleibt die Kennzahl im Merkmalsvektor. Durch diesen Prozess gelingt es den Automaten, sich in einer iterativen Prozedur bis zum Erreichen eines Abbruchkriteriums, analog zum biologischen Vorbild der Evolution besser und besser an die Umwelt anzupassen, also zu lernen, zu wachsen und sich selbst zu reparieren.

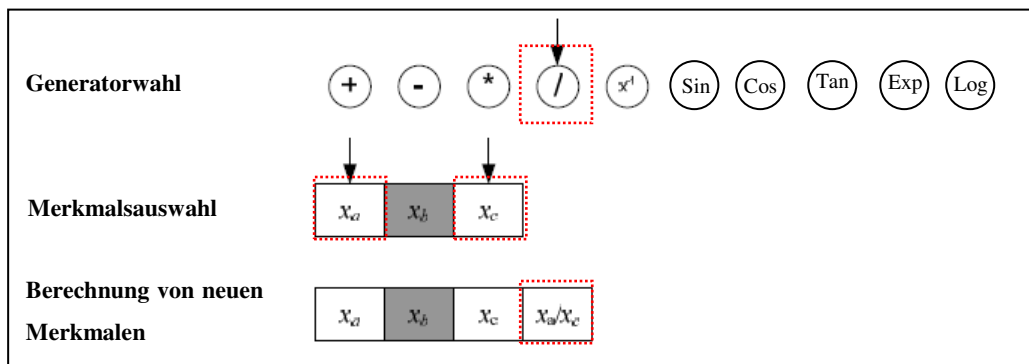


Abbildung 28: Vorgang der automatischen Merkmalsberechnung mit Genetic Algorithms

4.4 Dimensionalitätsreduktion

In der Mustererkennung spielt die Dimension des Eingangsvektors eine wichtige Rolle. Komplexe Eingangsmuster (Vektoren hoher Dimensionalität) enthalten häufig redundante und irrelevante Informationen, die eine einfache und schnelle Klassifikation erschweren (v. d. Ruhren, 2006). Die Dimensionalitätsreduktion dient neben der Gesamtvorhersageleistung, auch der Identifikation von relevanten Stimmkennwerten und der Reduktion von Rechenzeit. So kann die zusätzliche Aufnahme nur einer Zufallsvariablen die Gesamtvorhersageleistung bereits um 10% verschlechtern. Dieses Phänomen wird generell als „Curse of Dimensionality“ bezeichnet, d.h. die gleiche Anzahl an Fällen füllt mehr des gegebenen Raumes, wenn ihre Dimensionalität gering ist (Donoho, 2000). Für ein gegebenes Example Set bedeutet dies, dass es eine optimale Anzahl an Merkmalen gibt, darüber hinaus verschlechtert sich die Klassifikationsleistung wieder. Werden dementsprechend zu viele Merkmale selektiert, kommt es ebenso zu einem Informationsverlust und einer Verschlechterung der Klassifikationsleistung. Eingeteilt werden kann die Dimensionalitätsreduktion in (a) Feature Subset Selection und (b) Feature Transformation Methoden. Die Feature Subset Selection schließt wenig brauchbare Kennzahlen vollständig aus. Die Feature Transformation (z.B. Principle Component Analysis, Linear Discriminant Analysis) hingegen versucht unter Beibehaltung der ursprünglichen Kennzahlenvielfalt, diese lediglich in einen niedriger dimensional Raum zu projizieren.

Die Auswahl relevanter Merkmals subsets kann in supervised (überwacht) und unsupervised (unüberwachte) filter- bzw. wrapperbasierte Feature Subset Selection eingeteilt werden. Merkmale mit maximaler Relevanz und minimaler Redundanz werden aus dem Pool der potenziell relevanten Merkmale selektiert. Ziel dieser Reduktion ist zum einen, die Rechenzeit der Merkmalsberechnung und Klassifikation zu reduzieren, und zum anderen, die Gesamtzuordnungsleistung zu optimieren, indem der störende Einfluss irrelevanter Merkmale reduziert wird.

Die filterbasierten Verfahren selektieren unabhängig vom anvisierten Klassifikator mittels der isolierten Bewertung singulärer Attribute. Filterbasierte Methoden wenden eine Bewertungsgröße (z.B. Maximum Relevance, Minimum Redundancy) an jeder einzelnen Merkmals-Kriterium Kombination an und erstellen somit eine zur Selektion brauchbare Rangreihe der Merkmalsgüte (vgl. Schuller, 2006).

Verwendet werden hier Distanz-, Informations-, Assoziationsmaße. (z.B. wie Information Gain Ratios; Korrelationen und Auto Mutual Information). Aufgrund der geringen Rechenkapazität werden diese Maße vor allem zur Vorreduktion verwendet.

Anders als filterbasierte Selektionsansätze nutzen wrapperbasierte Ansätze bereits die simultane Information ganzer Merkmals subsets, indem sie schrittweise einen Klassifikationsalgorithmus anwenden. Zur Optimierung des Merkmalsraums finden in einer zweiten Stufe daher zumeist wrapperbasierte Verfahren Eingang. Wrapperbasierte Ansätze versuchen, über sequentiell gleitende Suchverfahren schrittweise einen optimalen Merkmalsverbund zu selektieren. Als Bewertungsfunktion sollte in diesem iterativen Verfahren jedoch auf einfache, wenig Rechenzeit verbrauchende Verfahren zurückgegriffen werden (z.B. lineare Regression). Aufgrund der unüberschaubaren Vielzahl von Subset Kombinationen ist das vollständige Absuchen des Merkmalsraums unmöglich. Die Suche verläuft daher schrittweise über Näherungsverfahren wie der Forward Selection, Backward Elimination, Sequential Floating Forward Selection oder Genetic Algorithms (siehe Abbildung 29).

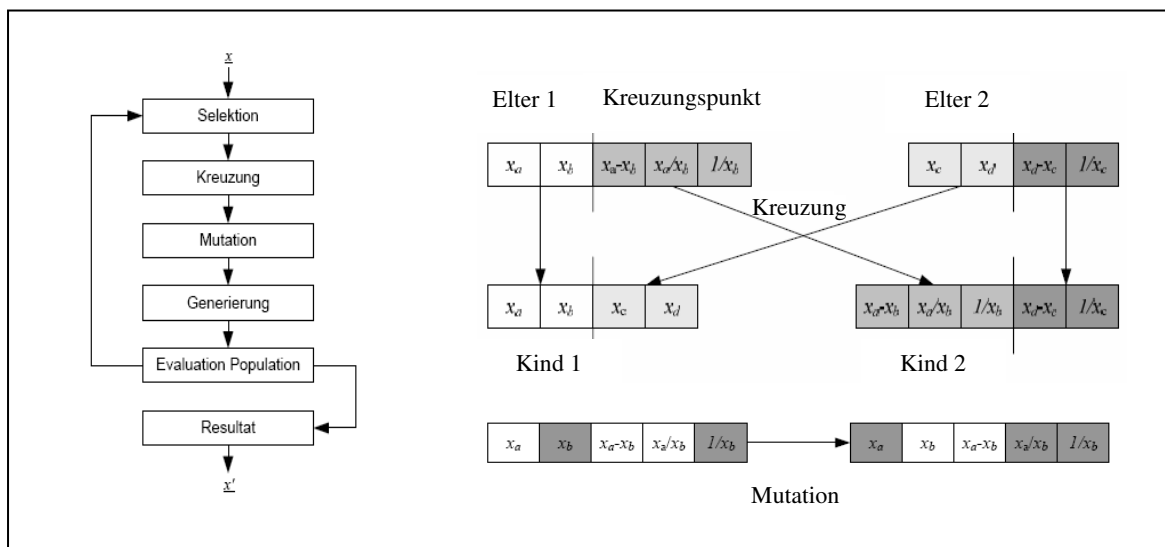


Abbildung 29: Ablaufdiagramm der Genetischen Programmierung in der Gesamtübersicht (links) und Generieren von neuen Feature Subsets über Kreuzung (Single-Point-Kreuzung mit variabler Länge; während der Mutation wechselt der Aktivierungsstatus eines Merkmals) und Mutation (rechts). Dunkelgrau eingefärbte Merkmale („Gene“) entsprechen deaktivierten Merkmalen

Eine der beliebtesten unsupervised Feature Transformation Methoden ist die Principal-Component-Analysis (PCA), die beispielsweise eine Reduktion der zu klassifizierenden

Merkmale auf 10 bis 100 Principle Components (Faktoren) erzielt (Chuang & Wu, 2004; Lee et al., 2002; Schauten et al., 2003; Ververidis et al., 2004). Die zur Datenkomprimierung und Dimensionsreduktion häufig angewandte PCA reduziert wie alle anderen Merkmalsextraktionsverfahren jedoch nur die Zahl der im Anschluss zu klassifizierenden Merkmale, nicht aber die Zahl der ursprünglich zu extrahierenden (vgl. Kraiss, 2003, Golz, 2007). Neben der Principle-Component-Analysis basierten Merkmalsextraktion sind auch andere unsupervised Methoden (unüberwachte Methoden) wie die Kernel PCA, die Independent Component Analysis (ICA), die No-Negative Matrix Factorization (NMF), die Locally Linear Embedding Methode (LLE). Diese Verfahren haben die Signal Repräsentation zum Ziel, d.h. das akkurate Mapping der Daten in einen niedrig-dimensionalen Raum. Supervised Methoden wie die Linear Discriminant Analysis hingegen nutzen die Information über die Klassenzugehörigkeit der Daten, um dieses für eine an der Klassifikationsleistung orientierte Dimensionsextraktion zu nutzen (maximiere die Between-class durch Within-class Varianz). Eine weitere wenig genutzte Möglichkeit der Dimensionalitätsreduktion liegt nicht in der weiteren Verwendung der transformierten Features (z.B. der stärksten 50 Principle Components oder Diskriminanz Faktoren), sondern in der Verwendung der um die Fehlervarianz der schwächsten Faktoren bereinigten Originalfeatures. Der Vorteil dieses Vorgehens ist seine resultierende bessere Interpretierbarkeit gegenüber dem Gebrauch von hochaggregierten und unanschaulichen Dimensionen und Faktoren.

Tabelle 14: Überblick zu Dimensionalitätsreduktions-Verfahren

Filter-basierte Feature Subset Selection	Beispiele
unsupervised	Densität, Smoothness, Entropie, Salienz
supervised	Korrelation, Information Gain Ratio, T-Test Wert (Schuller, 2006)
Wrapper-basierte Feature Subset Selection	
Suchstrategie	Komplett, heuristisch (Forward Selection, Backward Elimination, Sequential Forward Floating Search), zufällig (genetische Algorithmen)
Klassifikationsverfahren	Support Vector Machine, Multilayer Perceptron, k-Nearest- Neighbour, Decision Trees, Linear Discriminat Analysis
Feature Extraction (Feature Transformation)	
unsupervised	Principle Component Analysis (PCA), Kernel PCA, PCA Network, Nonlinear Autoassociative Network, Independent Component Analysis (ICA), Multidimensionale Skalierung (MDS), Sammon Map, Enhanced Lipschitz Embedding, SOM, Spectral Transformation, Wavelet Transformation
supervised	Linear Discriminant Analysis (nicht Fishers LDA) (Songcan & Xubing, 2004)

4.5 Klassifikation

Das Ziel des Klassifikations-Phasenabschnitts ist es, die Vorhersageleistung einzelner Stimmindikatoren durch simultane Bündelung ihres Informationsgehaltes mittels Fusionierungsalgorithmen zu erhöhen. Die Verfahren der automatischen Mustererkennung ordnen nach einer Lernphase unbekannte Eingabemuster einer bereits bekannten Klasse zu (z.B. über Lineare Diskriminanzanalysen). Die Muster setzen sich aus den selektierten Stimmerkmalen zusammen, die zuzuordnenden Klassen entsprechen den vorherzusagenden Zielbefindlichkeiten. Für diese Aufgabe stehen eine Reihe von Klassifikatoren mit unterschiedlichen Stärken- und Schwächenprofil zur Auswahl. Als Anforderungen an ein Klassifikationsverfahren gelten: optimale Erkennungsleistung, Umgang mit fehlenden Werten, Trainingsstabilität, Toleranz gegenüber Dimensionserhöhung, Echtzeitfähigkeit in der Erkennung, kurze Trainingszeit, geringer Bedarf an Lernbeispielen, Rechenleistung und Speicher.

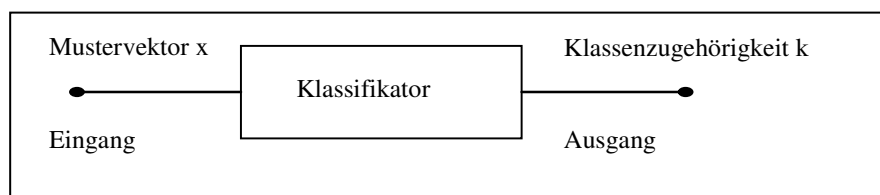


Abbildung 30: Klassifikationsprozess als Zuordnung eines Mustervektors x mittels eines Klassifikator Algorithmus zu einer Klasse k

Aufgabe der Klassifikation ist es, ein zunächst unbekanntes Eingabemuster einer bereits bekannten Klasse zuzuordnen (siehe Abbildung 30). In diesem Klassifikationsproblem wird mittels eines mathematischen Klassifikator-Modells der Merkmalsraum in disjunkte Bereiche eingeteilt und jeweils einer Klasse zugeordnet (Hastie, Tibshirani & Friedman, 2001; v. d. Ruhren, 2006). Am Eingang eines Klassifikators liegt das Merkmalsmuster, an seinem Ausgang wird die Klassenzugehörigkeit ausgegeben. Im vorliegenden Fall setzen sich die Merkmalsmuster aus den Kennwerten der selektierten Stimmerkmale zusammen. Die zuzuordnenden Klassen entsprechen den vorherzusagenden Schläfrigkeitsstadien.

Anstelle einer direkten Klassifikation der Waveform des Stimmsignals (die Klassifikation der Rohwaveform selbst wäre aufgrund der großen Anzahl an Abtastwerten unmöglich), werden die spezifischen Informationen des Wellenmusters zu Stimmkennzahlen verdichtet und in Form eines Merkmalsvektors zusammengefasst (Huber, 2002). Dieses durch einen Merk-

malsvektor beschriebenen Muster einer Sprachäußerung, entspricht jetzt einem Punkt im N -dimensionalen Merkmalsraum M . Muster lassen sich nur dann zuverlässig einzelnen Klassen zuordnen, wenn Muster gleicher Klassenzugehörigkeit im Merkmalsraum Ballungen und Cluster bilden (siehe Abbildung 31).

Allgemeine Problemstellung. Die einzelnen Klassen werden zum Klassifizierungsraum K zusammengefasst. Dieser hat die Dimension k , wenn k die Zahl der verschiedenen Klassen ist. Das Problem der Klassifizierung ergibt sich als Abbildung $M \rightarrow K$ des Merkmalsraums M auf den Klassifizierungsraum K (siehe Abbildung 31).

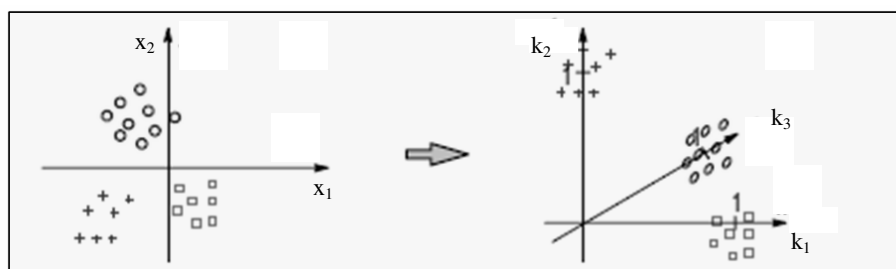


Abbildung 31: Ziel der Klassifikation als gesuchte Abbildung des Merkmalsraums in den Klassifizierungsraum (aus Hess, 2002a)

Bei einem neuen Klassifikationsproblem weiß der Klassifikator zunächst nicht, wo im Merkmalsraum die Ballungsmuster liegen. Er kennt also die Abbildung $M \rightarrow K$ nicht. In einer Trainingsphase, auch Lernphase genannt, muss der Klassifikator die Zuordnung $M \rightarrow K$ erst lernen. Hierzu benötigt er eine Lernstichprobe, also eine statistisch repräsentative Menge von Mustern, die im Merkmalsraum möglichst genauso verteilt sind wie die Muster, die der Klassifikator später klassifizieren, also ‘erkennen’ soll. Die Optimierung des mathematischen Modells erfolgt demnach an Hand einer Stichprobe von Eingabemustern mit bekannter Klassenzugehörigkeit (Trainingsmenge), indem der Fehler, der bei der Klassifikation der Trainingsmenge auftritt, minimiert wird (Walter, 2000).

Eine wichtige Eigenschaft von Klassifikatoren ist ihre Generalisierungsfähigkeit. Dies bedeutet, dass auch unbekannte Eingabemuster, die den Trainingsmustern zwar ähnlich, jedoch nicht in der Trainingsmenge enthalten sind, richtig klassifiziert werden können. Die Aufgabe des Trainingsprozesses besteht darin, eine Parametereinstellung für die Trennfunktionen zu finden, die eine möglichst gute Zuordnung von im Training nicht enthaltenen Merkmalsvektoren ermöglichen soll (Huber, 2002). Günstig sind zu diesem Zweck ein möglichst hoher Fallanzahl / Featureanzahl Quotient sowie die Berücksichtigung zunächst einfacher Klassifikatoren.

Zur Lösung von Klassifikationsproblemen steht eine Vielzahl von Klassifikationsverfahren zur Verfügung. Einige der am häufigsten zur akustischen Emotionserkennung eingesetzten Klassifikationsverfahren sind: Lineare Diskriminanzanalyse (Batliner et al., 2000; Kwon et al., 2003; McGilloway et al., 2000), Naive-Bayes (Schuller et al., 2005), Künstliche Neuronale Netze (Petrushin, 1999; Tato et al., 2002), Support-Vektor-Maschinen (Kwon et al., 2003; McGilloway et al., 2000; Schuller et al., 2004), K-Nearest-Neighbour Verfahren (Petrushin, 1999; Yacoub et al., 2003), Bayessche Netze (Ball & Breese, 1999), Dynamic-Time-Warping (Schuller, 2002d), Entscheidungsbäume (Litman et al., 2003; Yacoub et al., 2003), Gaußsche-Mixtur-Modelle (Breazeal & Aryananda, 2000) und Hidden-Markov-Modelle (Kwon et al., 2003; Nogueiras et al., 2001). Im Folgenden werden die drei Klassifikatoren, die in dieser Arbeit verwendet werden kurz vorgestellt: Künstliche Neuronale Netze, Support-Vektor-Machines und der K-Nearest-Neighbour Klassifikator. Detailliertere Angaben finden sich z.B. bei Bishop (2006), Duda, Hart und Stork (2001), Haykin (1999), Niemann (1990), Rojas (1993) oder Zell (1994).

Künstliche Neuronale Netze

Künstliche Neuronale Netze (NN) versuchen, die Intelligenz biologischer Systeme nachzubilden (McCulloch & Pitts, 1943; Rosenblatt, 1962; Rumelhart et al. 1986; Widrow & Hoff, 1960) und wurden in Analogie zum natürlichen neuronalen Netz entwickelt. Neuronale Netzwerke sind in einer Vielzahl von Anwendungsgebieten, wie der Biotechnologie, Mustererkennung und Sprachanalyse zu finden. Neuronale Netzwerke haben große Ähnlichkeit mit den Lernprozessen des menschlichen Gehirns. Im Gegensatz zu menschlichen NN sind künstliche NN nicht mit Millionen sondern nur mit einigen Hundert Neuronen zusammengesetzt. Ein einzelnes Neuron verarbeitet jeweils den Input vieler Eingänge und integriert diese

über eine Transferfunktion zu einem Ausgangssignal. Während der Trainingsphase wird der Fehler zwischen Vorhergesagten und tatsächlichen Kriteriumswerten minimiert, indem Gewichtungsfaktoren und Biasterme graduell adjustiert werden.

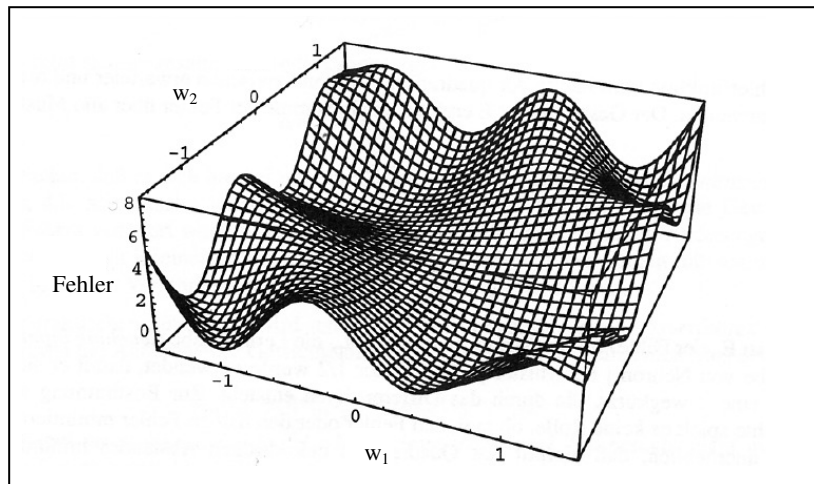


Abbildung 32: Fehlerfläche eines Neuronalen Netzwerks als Funktion der Gewichte w_1 und w_2 (nach Zell, 1994)

Die in Abbildung 32 für den zweidimensionalen Fall exemplarisch dargestellte Fehlerfunktion gibt den Fehler an, den das NN bei gegebenen Gewichten über alle Trainingsmuster summiert besitzt. Der Fehler des NN als Funktion der Gewichte w lässt sich somit als Fehlerfläche interpretieren. Ziel ist es die Gewichte so zu wählen, dass ein globales Minimum der Fehlerfunktion gefunden wird. Mittels des Gradientenabstiegsverfahrens wird versucht dieses Minimum und somit den globalen Netzwerkfehler zu finden. Erreicht wird dies, indem mit Hilfe der „Methode des steilsten Abstieges“ die Gewichte in Richtung des steilsten Gradienten der Fehlerfunktion verändert werden (Zell, 1994). Die Gewichte werden zunächst auf Zufallswerte initialisiert. Die Differenz von vorhergesagtem und tatsächlichem Wert für jedes Muster wird durch das nichtlineare Glied hindurch propagiert und sorgt für eine entsprechende Veränderung der Gewichte. Die Anpassung der Gewichte erfolgt schließlich über den Term:

$$\Delta w_{ij} = w_{ij}(t+1) - w_{ij}(t) = -\alpha (\partial E / \partial w_{ij}) \quad (4.1)$$

Die Lernrate η bestimmt die Geschwindigkeit der Konvergenz, aber auch die Sprunghaftigkeit des Netzes. Üblich sind Werte in einer Größenordnung von 0.03 bis 0.06. In Abhängigkeit von der Lernrate sind mehrere hundert Iterationen, des Backpropagation-Algorithmus notwendig, um einen vorgegebenen Wert der Fehlerfunktion E zu unterschreiten.

Funktionsweise. Künstliche neuronale Netze basieren auf einem vereinfachten Modell eines Neurons und werden auch als Neuronen-ähnliche Netze bezeichnet. Ihre zentralen Verarbeitungseinheiten sind Neuronen. Ein Neuron ist äquivalent zu einem linearen Klassifikator für ein Zwei-Klassenproblem und erhält über eine beliebige Anzahl von Eingängen Eingabedaten von vorgeschalteten Neuronen, die innerhalb des Neurons verrechnet werden. Die Verrechnung findet über eine gewichtete Summenbildung und eine anschließende Anwendung einer Aktivierungsfunktion (meist Sigmoidfunktion) statt (siehe Abbildung 33 links). Die Aktivierungsfunktion begrenzt das Ausgangssignal auf Werte zwischen 0 und 1. Im Falle der Zugehörigkeit eines Eingangsmusters zur Klasse k , sollte das Ausgangssignal y auch nahe 1, bei Nicht-Zugehörigkeit nahe 0 liegen (vgl. Hess, 2002a). Für den Fall eines Mehrklassenproblems ist für jede mögliche Klasse ein Ausgabeschichtneuron einzusetzen, das die jeweilige Stärke der Klassenzugehörigkeit angibt.

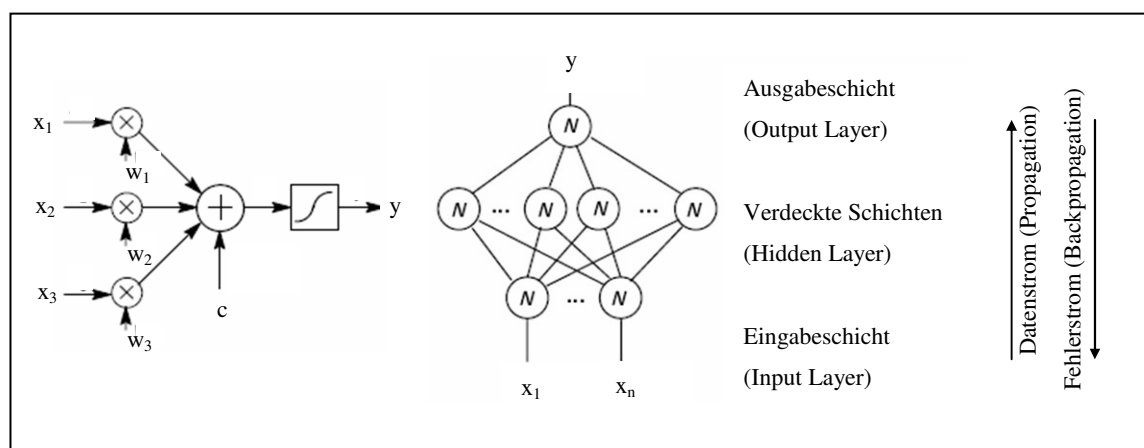


Abbildung 33: Schematische Darstellung eines künstlichen Neurons mit 3 Eingängen (x_1 , x_2 , x_3), denen ihnen multiplikativ zugeordneten Gewichten (w_1 , w_2 , w_3) sowie der anschließenden Summenbildung und Aktivierungsfunktionsanwendung (hier Sigmoid Funktion) (links). Architektur eines dreilagigen aus künstlichen Neuronen aufgebauten neuronalen Netzes (rechts)

Künstliche Neuronale Netze sind in *Schichten* (Layer) organisiert. Es werden, wie in Abbildung 33 (rechts) ersichtlich, Eingabeschicht, Ausgabeschicht und verdeckte Zwischenschichten (Hidden Layer) unterschieden (Bishop, 2007). Die Anzahl der Neuronen der Eingabeschicht entspricht der Anzahl der Eingangsgrößen (Anzahl der Merkmale). Jedem Neuron der Ausgangsschicht ist (Ausgangsneuron) eine Ergebnisklasse zugeordnet. Neben der Anzahl von Schichten ist ein weiterer wichtiger Aspekt der Topologie eines NN die Verknüpfungsrichtung der Schichten. Die Datenweitergabe erfolgt bei vorwärtsverknüpften Netzen (Feedforward Netze) ausschließlich von der niedrigeren Schicht zur nächst höheren Schicht, bei einem rückwärts verketteten Neuronalen Netzen ist es umgekehrt (Duda et al., 2001). Auch in beide Richtungen oder seitlich verknüpfte Netztopologien sind möglich (siehe Abbildung 34). Zu den am weitesten verbreiteten Netzwerkarchitekturen gehört das Multilayer Perceptron (MLP) und die Radial Base Function (RBF).

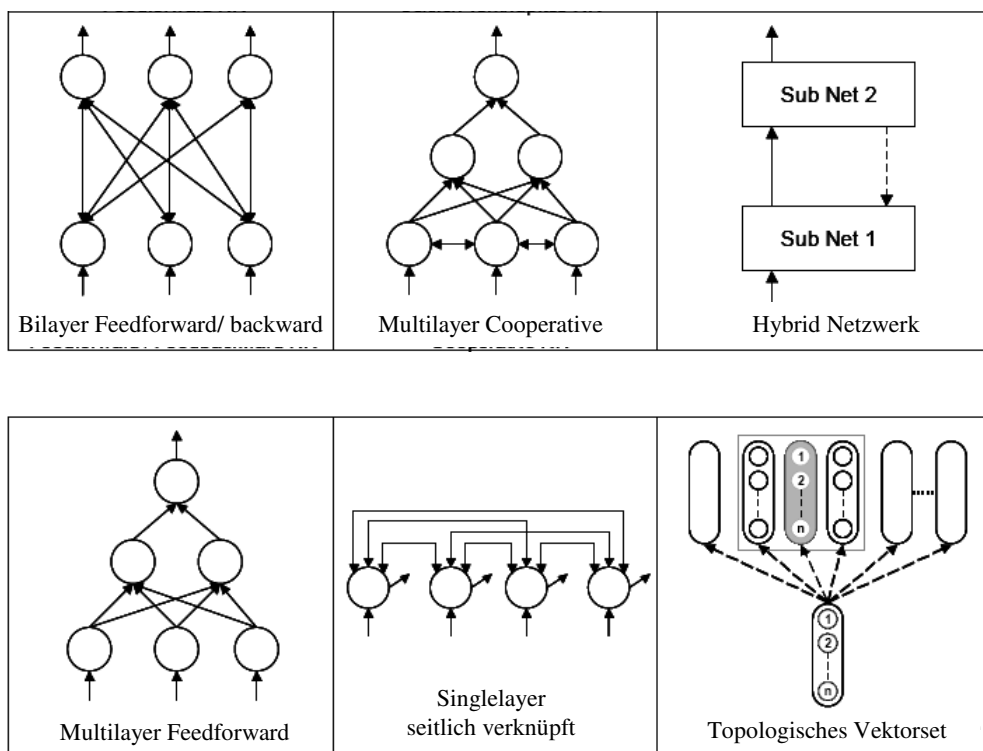


Abbildung 34: Basistopologien Neuronaler Netze (nach Maren, Harston und Pap, 1990)

Die Vorhersageleistung eines Neuronalen Netzes ist determiniert durch die Anzahl der verdeckten Schichten (siehe Abbildung 35). Zweilagige Netze (eine Eingangs- und eine Ausgangsschicht) sind lediglich zu linearer Trennung des Merkmalsraums fähig. Ein dreilagiges Netz (eine verdeckte Schicht) ermöglicht es, den Merkmalsraum durch einen Linienzug zu begrenzen und so konvexe Polygone zu klassifizieren (siehe auch Abbildung 36). Jede beliebige Art der Separation ermöglichen vierlagige Netze, die durch Überlagerung und Schnitt konvexer Polygone Mengen beliebiger Form repräsentieren (Zell, 1994). Für viele praktische Klassifikationsanwendungen leistet ein Netzwerk mit drei Schichten sehr gute Vorhersagen. Sowohl die Anzahl der Schichten, als auch die Anzahl der Neuronen in den jeweiligen Schichten bestimmen die Architektur des MLP. Die Anzahl der Eingangsneuronen entspricht der der Anzahl der Eingangsgrößen (z.B. Sprachfeatures) des zu modellierenden Prozesses, die Zahl der Ausgangsneuronen wird durch die Anzahl der zu separierenden Klassen bestimmt. Heuristische Regeln liefern hierzu jeweils Näherungen bezüglich der optimalen Anzahl der Neuronen in den verdeckten Schichten (vgl. Janetzke & Falk, 1994).

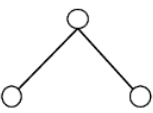
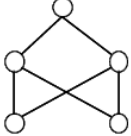
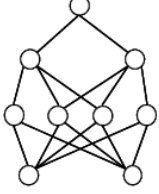
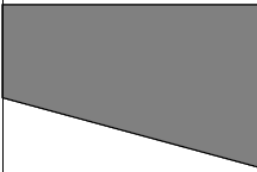
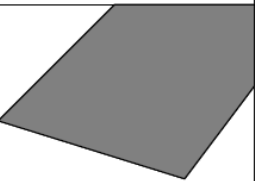
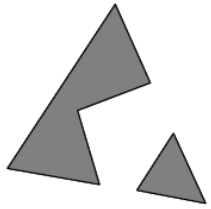
Architektur des NN	<p>2-lagiges NN</p> 	<p>3-lagiges NN (1 Hidden Layer)</p> 	<p>4-lagiges NN (2 Hidden Layer)</p> 
Trennung des Merkmalsraums	Lineare Trennfläche	konvexe Polygone	Mengen beliebiger Formen
			

Abbildung 35: Separationfähigkeit Neuronaler Netze in Abhängigkeit von der Anzahl verdeckter Schichten (Hidden Layer) (nach v. d. Ruhren, 2006)

Die Adaptation des NN an eine Klassifikationsaufgabe erfolgt während des Lern- oder Trainingsprozesses durch Modifikation der Verbindungsgewichte. Ziel ist es Klassifikationsfehler für alle Trainingsmuster zu minimieren (v. d. Ruhren, 2006). Ein Vorteil künstlicher neuronaler Netze liegt in ihrer Eigenschaft, selbstständig funktionale Zusammenhänge von Daten zu erlernen. Des Weiteren können sie auf Veränderungen in der Problemstellung reagieren, indem sie durch Nachlernen ihre Verbindungsgewichte anpassen. Sie sind generalisierungsfähig und liefern auch für nicht gelernte Eingaben korrekte Ausgaben. Auch besteht eine Fehlertoleranz gegenüber unvollständigen oder leicht verfälschten Eingaben. Darüber hinaus sind sie ein geeignetes Werkzeug zur Modellierung auch komplexer nicht-linearer Prozesse. Dazu sind mathematisch-statistische Verfahren, aufgrund der erforderlichen Annahmen über die Verteilungsdichtefunktion der Trainingsmenge und der a priori Wahrscheinlichkeiten der Klassen, häufig nicht geeignet (v. d. Ruhren, 2006). Aufgrund dieser vielfachen Stärken haben bereits eine Reihe von Studien (Andreeva et al. 2004; Fukumi, 2005; Wang et al., 2003) Neuronale Netzwerke zur fusionierten Vorhersage von Schläfrigkeit genutzt.

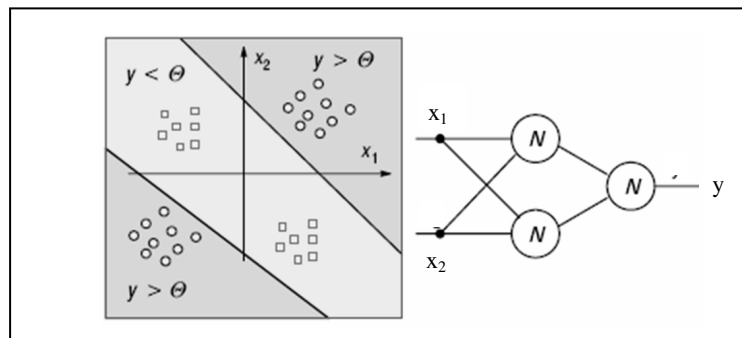


Abbildung 36: Schematische Darstellung konvexer Polygon Trennflächen dreilagiger neuronaler Netze (nach Hess, 2002a)

Die Schwäche künstlicher neuronaler Netze liegt u.a. in der schwierigen Nachvollziehbarkeit der zugrundeliegenden Modellstruktur und somit auch der schwierigen Interpretierbarkeit der Bedeutung einzelner Inputmerkmale. Die Ergebnisse lassen sich aufgrund der Struktur der inneren Schichten oft nur schwer nachvollziehen. Darüber hinaus ist ähnlich wie in anderen komplexen Klassifikationsverfahren die Frage nach der für ein gegebenes Sample und Problem adäquaten Netzarchitekturen nur mit groben Heuristiken zu beantworten. Detaillierte Beschreibungen zu dem umfangreichen Forschungsfeld Neuronale Netze finden sich z.B. bei Zell (1997) und Rojas (1993).

Support-Vektor-Maschinen

Support-Vektor-Maschinen (SVM) wurden von Vapnik (1995) zur Klassifizierung von Daten eingeführt und sind in der statistischen Lerntheorie begründet. Eine Support-Vektor-Maschine unterteilt eine Menge von Objekten zunächst in zwei Klassen. Hauptidee der SVM ist die Maximierung des Margins (Abstandrands) zwischen zwei Cluster. Dieses wird erreicht, indem eine Hyperebene so platziert wird, dass der Trainingsfehler mittels Lagrange Multiplier minimiert wird (Benett & Campbell, 2000). Das Verfahren ist so konzipiert, dass ein möglichst breiter Abstandrand um die Klassengrenze (Hypertrennebene) frei von Objekten bleibt. Aus diesem Grund wird die Support-Vektor-Maschine auch Breiter-Rand-Klassifikator (Large Margin Classifier) genannt. Dabei wird der Abstand der Vektoren, die der Hypertrennebene am nächsten liegen, maximiert. Der freie, breite Rand soll es ermöglichen, Objekte, die nicht Teil der Trainingsmenge sind, möglichst zuverlässig zu klassifizieren. Ziel des Trainings bzw. der Entwicklung einer SVM ist es, die Hypertrennebene mit dem breitesten freien Rand zwischen den Klassen zu finden (Duda et al., 2001).

Die Optimierung der Hypertrennebene wird über die Bestimmung der sogenannten Lagrange-Faktoren berechnet (Bishop, 2006). Beim Einsetzen der Hypertrennebene werden nicht alle Trainingsvektoren berücksichtigt. Die Hypertrennebene ist nur von den ihr am nächsten liegenden Vektoren abhängig – und auch nur diese werden benötigt, um die Ebene mathematisch exakt zu beschreiben. Diese nächstliegenden Vektoren werden nach ihrer Funktion Stützvektoren (Support Vectors) genannt. Die Supportvektoren sind transformierte Trainingsmuster und haben jeweils die gleiche Entfernung zur Hypertrennebene (Duda et al., 2001). Sie definieren die trennende Hyperebene optimal und sind die am schwierigsten zu klassifizierenden Muster (Duda et al., 2001). Die Hypertrennebene soll demnach folgende Bedingungen erfüllen (Barton, 2004): die Größe der Klassifikationsfehler in der Lernstichprobe sei minimal; der Mindestabstand zwischen den Klassen sei maximal. Die Hypertrennebene wird nur aus den sogenannten Supportvektoren berechnet, also den Mustern der Klassen, die relativ nah an der Klassengrenze liegen. Da die Hypertrennebene eine kanonische Form besitzen soll, erhält ihr Funktionswert für alle Supportvektoren die Werte -1 bzw. 1 .

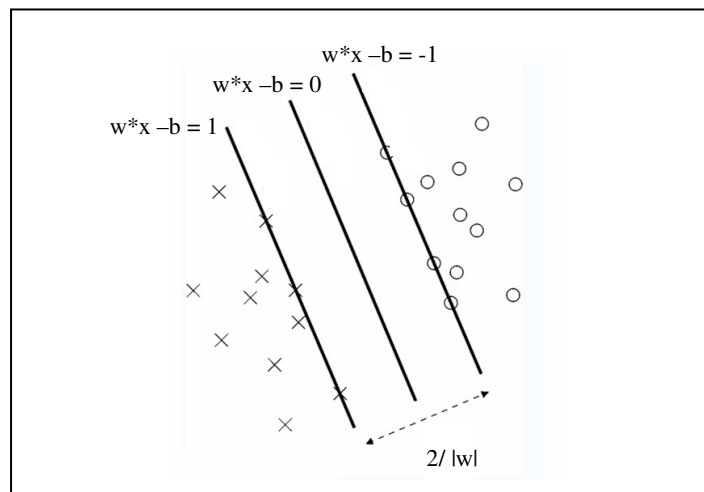


Abbildung 37: Maximum-Margin Hypertrennebenen einer Support-Vektor-Maschine mit Support-Vektoren.

Die Trennung mit einer Hypertrennebene ist jedoch nur dann möglich, wenn die Klassen linear trennbar sind (siehe Abbildung 37). Dies ist in realen Anwendungsfällen häufig nicht der Fall. Im Fall nicht-separierbarer Klassen verwenden SVM Kernel-Funktionen, um nicht lineare Klassengrenzen zusetzen. Es werden sogenannte Slack-Variablen in die Gleichung eingeführt, die die nicht erfüllten Nebenbedingungen der Gleichung korrigieren (Berton, 2004). Durch die nichtlineare Transformation der Merkmalsvektoren sollen die Klassen im höherdimensionalen Raum wieder linear trennbar sein. In seiner ursprünglichen Form können SVM lediglich 2-Klassen Probleme angehen. Über das Splitting eines Mehrklassen Problems in einzelne 2-Klassen Probleme kann die SVM jedoch auch für die Mehrklassen Klassifikation eingesetzt werden.

Tabelle 15: Gebräuchliche Kernel-Funktionen der SVM

Kernelfunktion	Gleichung
Kernel	$K(x_i, x_j)$
Polynomial Kernel	$K(x_i, x_j) = (x_i * x_j / a + b)^d$
Radial Kernel	$K(x_i, x_j) = e^{-\gamma \ x_i - x_j\ ^2}$
Sigmoid Kernel	$K(x_i, x_j) = \tanh(\beta x_i * x_j + b)$

Freie Parameter. Die SVM besitzen relativ wenige freie Parameter zur Modellspezifikation. Zu ihnen gehören die Kernelfunktionen der SVM. Die gebräuchlichsten sind hier in Tabelle 15 aufgelistet. Die wichtigen Feinjustierungsparameter der SVM sind der Gamma und der C Parameter. Mittels des Gamma Parameters wird die Komplexität der Merkmalsraums-Grenze modelliert. Ein größeres Gamma ermöglicht eine höhere Komplexität der Grenzziehung, vergrößert aber gleichzeitig die Gefahr des Overfittings, also der Generalisierbarkeit des Modells. Der Parameter C wiederum kontrolliert das Verhältnis von False Alarm vs. Misses Fehler. Eine größeres C produziert somit eine größere Anzahl von False Alarms. Je nach Problemstellung und somit produzierten Kosten von False Alarm und Misses Fehlern, kann die SVM eine problemadäquate Balance finden. Zur Illustration des Einflusses der Parameter C und Gamma auf den Klassifikationserfolg sind zwei Beispiele aus dem Bereich der Mikroschlafdetektion und der Fahrer Distraktionserkennung aufgeführt. Weitere detaillierte Beschreibungen der mathematischen Hintergründe der SVM finden sich bei Burges (1998), Campbell (2000) oder Christianini und Shawe-Taylor (2000).

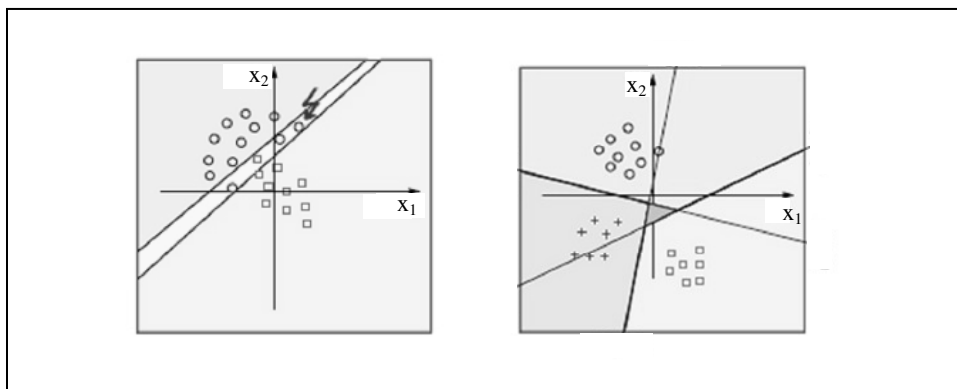


Abbildung 38: Schematische Darstellung linearer Trennbarkeit (links) und undefinierbarer Teilbereiche (rechts) (nach Hess, 2002a)

Ein bekanntes Problem künstlicher neuronaler Netze besteht in der möglichen Überadaptation der Klassifikatoren bei der Verwendung zu vieler Parameter oder einer zu kleinen Lernstichprobe. Während zu wenige Parameter zu einer schlechten Approximation führen, verursachen zu viele Parameter eine zu hohe Varianz der Lösungen. „Die neue Idee der Support-Vektor-Maschinen besteht in der Minimierung des strukturellen Risikos beim Klassifikatortraining, d.h. dass neben dem empirischen Risiko auch die Zuverlässigkeit des trainierten Klassifikators in Abhängigkeit von der Lernstichprobe berücksichtigt wird“ (Berton, 2004, S. 69). Die Zuverlässigkeit wird durch ein Konfidenzintervall beschrieben, das von der Größe der Lernstichprobe und der Dimension ihrer Vektoren abhängig ist. Während bei anderen Klassifika-

tionstechniken das Konfidenzintervall vorgegeben und danach das empirische Risiko minimiert wird, legt man bei der SVM erst das empirische Risiko fest, um dann durch die Wahl einer geeigneten Klassifikatorstruktur das Konfidenzintervall in Abhängigkeit von der Lernstichprobe zu minimieren (Berton, 2004). Auf diese Weise kann eine Überadaptation bzw. Überlernen eingeschränkt oder vermieden und somit eine gute Generalisierungsfähigkeit erreicht werden. Weitere Vorteile von SVM sind, dass sie anders als künstliche neuronale Netze, weitgehend unabhängig von der Komplexität und Anzahl betrachteter Merkmale sind. Es gibt nur wenige Modellparameter, die vom Anwender festgelegt werden können und müssen. Eine komplizierte Parametrisierung, wie sie z.B. bei künstlichen neuronalen Netzen erforderlich ist, entfällt somit. SVM sind linear im Merkmalsraum, also konzeptionell und rechnerisch einfach; sie erzeugen aber nichtlineare Trennung im Eingaberaum (Schölkopf et al., 1999). Mit SVM wird ein Mittelweg zwischen künstlichen neuronalen Netzen, die beliebig nicht-lineare Zusammenhänge aufdecken können, und Entscheidungsbaumverfahren, die intuitiv zu interpretieren, aber auf die Erkennung linearer Zusammenhänge beschränkt sind, beschränkt. Es gibt keine algorithmisch bedingten Zufälligkeiten (wie teilweise bei neuronalen Netzen), sondern stabile, reproduzierbare Ergebnisse, unabhängig vom verwendeten Optimierungsalgorithmus oder von Initialisierungswerten.

Nachteile der Support-Vektor Maschinen liegen darin, dass für neu hinzukommende Eingabedaten jedes Mal erneut ein Training erforderlich ist. Es besteht keine Möglichkeit, lediglich die vorhandenen Ergebnisse zu ergänzen. Des Weiteren steigt die Trainingszeit mit der Anzahl der Klassen an, da für jede Klasse eine neue SVM hinzugenommen werden muss. Bei der Auswertung in der späteren Klassifikation eines Datenpunktes steigt die Zeitkomplexität ebenfalls linear an. Für einen erfolgreichen Einsatz in der Praxis muss ein Klassifikator leicht zu handhaben und bedienerfreundlich gestaltet sein, so dass keine großen Zeitverluste in der Anwendung entstehen. Das gilt insbesondere für die Wahl möglicher Parameter, die vom Benutzer festzulegen sind. Die Einstellungen einer SVM erfordern vom Benutzer jedoch ein zum Teil tief gehendes Verständnis ihrer Arbeitsweise. Eine ausführliche Darstellung ist z.B. bei Vapnik & Cortes (1995) oder Burges (1998) zu finden.

K-Nearest-Neighbour Klassifikator

Nearest-Neighbour-Methoden (Nächster-Nachbar Methode) gehören zu den distanzbasierten nonparametrischen Klassifikationsverfahren (Bishop, 2006). Es werden keine Annahmen über Wahrscheinlichkeitsverteilungen oder parametrische Trennfunktionen getroffen. Stattdessen wird für jede Klassifikation die gesamte Trainingsstichprobe gespeichert und berücksichtigt (Huber, 2002). Ein typischer Vertreter dieser Klasse ist der distanzbasierte k-Nearest-Neighbour-Klassifikator ("Minimum-Distance-Klassifikator"). Der k-Nearest-Neighbour-Klassifikator bestimmt die k nächsten Nachbarn im Merkmalsraum und entscheidet sich dann für diejenige Klasse, zu der die Mehrzahl dieser k nächsten Nachbarn gehören (Huber, 2002). Ein Muster wird demnach derjenigen Klasse zugeordnet, zu deren Repräsentanten es die kürzeste euklidische Entfernung hat (siehe Abbildung 39).

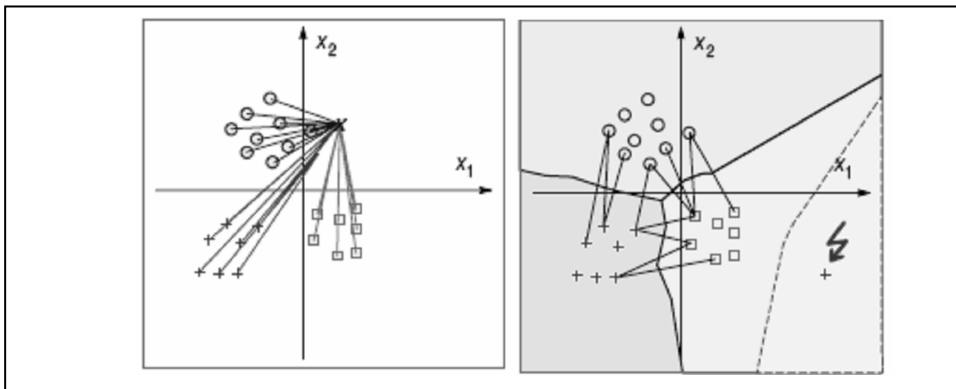


Abbildung 39: Nächster-Nachbar-Klassifikator (links) und die Auswirkung eines Ausreißers (rechts) (aus Hess, 2002a)

Ein Punkt im Eigenschaftsraum wird einer Klasse durch Mehrheitsentscheid zugewiesen, d.h. er wird der Klasse zugeordnet, deren Trainingsdaten den Punkt am häufigsten umgeben. Die zugrunde liegende Annahme ist, dass Muster aus derselben Klasse ähnlich, und Muster unterschiedlicher Klassen unähnlich sind. Unähnlichkeit ist als Abstand der Muster im Merkmalsraum definiert. Einzige Voraussetzung ist die Anwendbarkeit der Distanzfunktion für zwei beliebige Muster (Duda et al., 2001). Die Distanzfunktion, das Herzstück eines k-Nearest-Neighbour Klassifikators, ist im Allgemeinen als der euklidische Abstand definiert. Denkbar sind aber auch andere Metriken wie die City-Block Metrik.

$$d_{\text{euklid}} = \left[\sum_{n=1}^N (x_n - y_n)^2 \right]^{0,5} \quad (4.3)$$

$$d_{\text{cityblock}} = \sum_{n=1}^N |x_n - y_n| \quad (4.4)$$

Eine Distanzfunktion ist als Metrik im mathematischen Sinne definiert. Nachteil ist, dass die gesamte Stichprobe gespeichert und bei der Klassifikation durchsucht wird, was in den meisten Fällen einen hohen Speicher- und Zeitaufwand mit sich bringt (Huber, 2002). Vorteile des k -NN Klassifikators sind, dass er einfach zu implementieren ist und eine sehr gute und zuverlässige Klassifikation erlaubt, wenn die Anzahl der Nearest Neighbors optimiert wird.

Ensemble Klassifikation. Verbessert werden kann die Klassifikationsleistung (ca. + 5% bis 10% Zuordnungsgenauigkeit) durch eine Kombination der Klassifikationsverfahren in multiplen Instanzen (Ensemble-Klassifikation). Die Klassifikationsergebnisse der Level-0-Klassifikatoren werden in einem nachfolgenden Schritt mittels einer finalen Instanz, dem übergeordneten Level-1-Klassifikator verarbeitet (siehe Abbildung 40; Ting & Witten, 1999). Im einfachsten Fall der Kombination der Level-0 Ergebnisse, dem Majority Voting werden die endgültigen Klassenzuordnungsentscheidungen über einen Mehrheitsentscheid gefällt (Voting). Darüber hinaus lassen sich die Ensemble Klassifikatoren z.B. mittels Stacking (Wolpert, 1992), und Boosting kombinieren (Petrushin, 1999; Schuller, 2006).

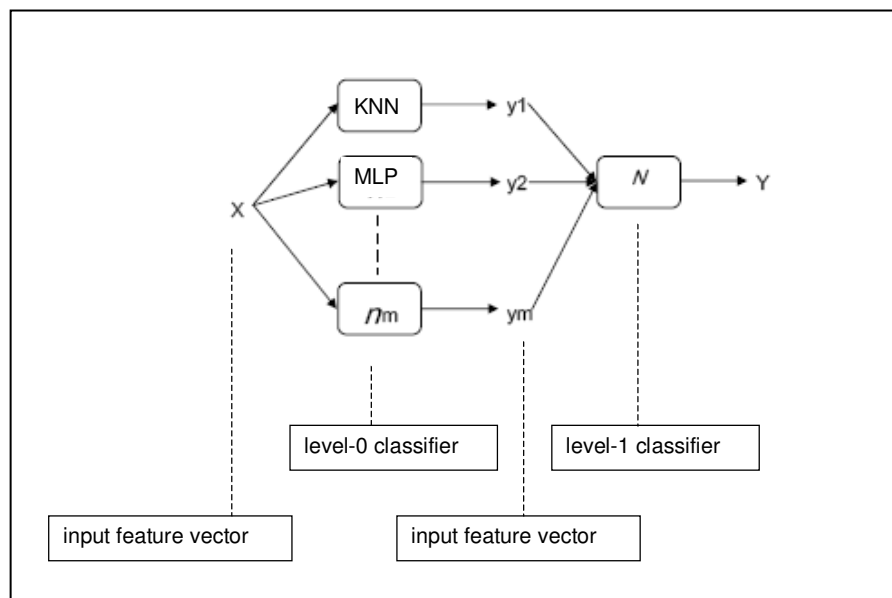


Abbildung 40: Ablauf der Ensemble-Klassifikation. Vorhersageergebnisse der Level-0 Klassifikation werden als neuer Input Merkmalsvektor im Level-1 Klassifikator verarbeitet.

Evaluierungsstrategien. Die in der psychologischen Forschung überwiegend angewandte Validierungsstrategie bestimmt an einem ungesplitteten Datensatz sowohl die Modellparameter (z.B. b-Regressionengewichte) als auch die Modellanpassungsgüte (z.B. R^2). Dieses Vorgehen führt (trotz Korrekturversuchen wie dem „korrigierten R^2 “) zu Überschätzungen der tatsächlichen Vorhersagekraft der Modelle. Um eine realistische Bestimmung der Erkennungsleistung abzuleiten, muss daher in einem Trainingsdatensatz das Modell trainiert werden und in einem davon disjunkten Datensatz die Anpassungsgüte getestet werden. Da dieses Prozedere die zum Training des Modells zur Verfügung stehende Datenmenge reduziert, kommen stichprobeneffiziente Vorgehen wie die Leave-One-Speaker-Out (LOSO) Strategie oder die j-fach stratifizierte Kreuzvalidierung (SCV) zum Einsatz (Witten & Frank, 2000). Dabei wird die gesamte Datenmenge zuerst in j getrennte Teilmengen partitioniert. Anschließend werden in j Testdurchläufen jeweils die j-te Teilmenge als Testmenge und die j-1 verbleibenden Teilmengen als Trainingsmengen verwendet. Als Gesamterkennungsleistung wird der Mittelwert aus den Einzelerkennungsleistungen der j Durchläufe berechnet. Im einfachsten Fall wird der Datensatz in zwei 50% umfassende Datenpartitionen gesplittet und die Teilmenge 1 einmal als Trainingsmenge und einmal als Testmenge verwendet. Andere gebräuchliche Partitionierungen nutzen Trainingsdatenanteile von 66,7%, 80%, 90% oder 95%. Für die Bewertung der erzielten Erkennungsleistungen stehen Kenngrößen wie Sensitivität (Detektionsrate; Anteil der z.B. als müde klassifizierten, müden Sprecher), Spezifität (1-False Alarm Rate; Anteil der als müde klassifizierten, nicht-müden Sprecher) und Klassifikationsgenauigkeit (Accuracy Rate; richtig zugeordnete Sprechsample durch Gesamtanzahl aller Sprechsample) zur Verfügung.

TEIL II

EMPIRISCHE VALIDIERUNGSTUDIE DES AKUSTISCHEN MESS- PROZESSES

1 METHODE

1.1 Stichprobe

Nachdem im ersten Teil der vorliegenden Arbeit die theoretischen und statistisch-methodischen Grundlagen der stimmbasierten akustischen Schläfrigkeitsanalyse gelegt worden sind, werden in diesem zweiten Teil Methodik und Ergebnisse einer empirischen Validierungsstudie der stimmbasierten akustischen Schläfrigkeitsanalyse vorgestellt. Ziel der Studie ist der experimentelle Validitätsnachweis der akustischen Schläfrigkeitsanalyse.

An der Untersuchung nahmen 24 Studentinnen und Studenten der Bergischen Universität Wuppertal teil: 15 Frauen und 9 Männer. Die Teilnahme an der Studie wurde mit den für den weiteren Studienverlauf wichtigen Versuchspersonenstunden entlohnt. Das Durchschnittsalter der Frauen betrug 26.85 Jahre ($SD= 3.82$), das der Männer 28.8 Jahre ($SD = 7.20$). Das Durchschnittsalter der gesamten Stichprobe betrug 27.07 Jahre ($SD = 5.38$).

Probandenexklusionskriterien. Exklusionskriterien zur Teilnahme an der Studie waren ein auffälliges Schlafmuster. Dieses beinhaltete keine regelmäßigen Bettgehzeiten zwischen 22.00 und 24.00 Uhr oder andere Schlafstörungen. Erhoben wurden die Schlafmuster der Probanden über ein ein wöchiges Schlaftagebuch. Ferner führte ein Pittsburgh Sleep Quality Index (PSQI Global Score) von größer acht zum Ausschluss. Des Weiteren fungierte eine fehlende Alkohol-, Nikotin- und Koffein-Karenz am Untersuchungstag als Exklusionskriterium.

1.2 Untersuchungsprozedere

Acht Tage vor dem geplanten Untersuchungstermin wurde in einem ersten Screening-Gespräch mit den Probanden zunächst anamnestisch abgeklärt, dass bei ihnen keine Wach- oder Schlafstörung vorlag, sie keine extremen Abendtypen waren und insgesamt einen regelmäßigen Schlaf-Wach-Rhythmus einhielten. Der eigentliche Untersuchungstag begann für die Probanden um 19.30 Uhr mit einer Einführung des Versuchsleiters über den Versuchsablauf und den Gebrauch der Instrumente im arbeitspsychologischen Labor der Bergischen Universität Wuppertal statt. Alle Probanden wurden explizit instruiert, auf keinen Fall einschlafen zu dürfen und bis zum Ende der Untersuchung wach zu bleiben. In der ersten halben Stunde wurde des Weiteren eine Fragebogenbatterie erhoben. Die Probanden machten sich ferner mit den späteren Sprachaufgaben vertraut. Die Versuchsteilnehmer der Schlafdeprivationsstudie wurden innerhalb eines within-subject Design untersucht (siehe Abbildung 41).

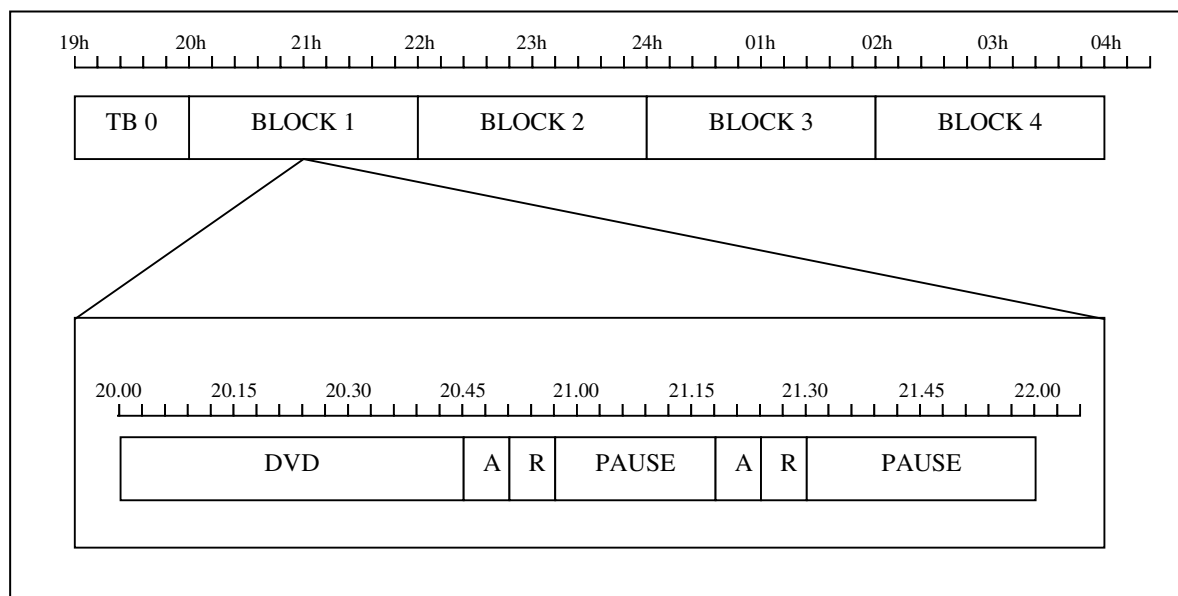


Abbildung 41: Zeitlicher Verlauf der sich jeweils von 19.30 Uhr bis 04.00 Uhr erstreckenden Schlafdeprivationsstudie. TB 0 = Trainingsblock 0; Block 1-4 = 120 minütige Messblöcke 1-4; DVD = DVD-Film schauen; A = Acoustic Sleepiness Analysis; R = Schläfrigkeitsselbst- und -fremdratings; PAUSE = freie Small Talk Pause

Die verwerteten Messungen fanden zwischen 20.00 Uhr und 04.00 Uhr statt. Der Untersuchungsablauf setzte sich aus vier Untersuchungsblöcken je zwei Stunden zusammen. In jedem Messblock wurden alle 40 Minuten einerseits die selbst- und versuchsleiterberichteten Schläfrigkeitssurteile und andererseits die Sprachaufnahmen aufgezeichnet. Die zwischen den Messungen liegende Zeit wurde vor allem mit dem Schauen von DVD-Filmen sowie mit Versuchsleiter Small Talk verbracht. Während der Untersuchungsnacht durften die Probanden den ca. 20m² großen, mit lichtundurchlässigen Jalousien verdunkelten Raum nur kurz (z. B. für einen Toilettengang) verlassen. Snacks und koffeinfreie Getränke wurden jede Stunde angeboten. Die Helligkeit im Raum entsprach der an Büroarbeitsplätzen üblichen über Leuchtstoffröhren produzierten Beleuchtungsstärke. Am Ende der Nachtuntersuchung stand es den Probanden frei, entweder in einem Laborraum die restliche Nacht zu verbringen, ein Taxi nach Hause zu nehmen oder eigenständig nach Hause zu gehen.

1.3 Instrumente

Vor Untersuchungsbeginn wurden bei den Versuchspersonen ein unregelmäßiges Schlafmuster und mangelnde Schlafqualität mit dem Pittsburgh Schlafqualitätsindex (PSQI) von Buysse, Reynolds, Monk, Berman und Kupfer (1989) ausgeschlossen. Mit dem PSQI wurden retrospektiv für einen Zeitraum von vier Wochen die Häufigkeit schlafstörender Ereignisse, die gewöhnlichen Schlafzeiten, Einschlafzeit und Schlafdauer, die Einnahme von Schlafmedikationen sowie die Tagesmüdigkeit erfragt. Zum gleichen Zweck wurde eine Woche vor Beginn der Untersuchung ein Schlaftagebuch eingesetzt, in dem die Zeit des Zubettgehens, die Einschlafzeit, die Aufwachzeit, der Zeitpunkt des Aufstehens sowie tagsüber stattgefundenen Schlafepisoden erfasst wurden. Auf die detaillierten Ergebnisse des PSQI und des Schlaftagebuchs wird im nachfolgenden nicht weiter eingegangen, da die Ergebnisse lediglich für die Erfassung der Exklusionskriterien benötigt wurden.

Zur letztendlichen Einschätzung der tatsächlichen Schläfrigkeit diente die derzeit am besten validierte Selbsteinschätzungsskala, die Karolinska Sleepiness Scale (KSS). Das Verfahren ist in der Schlaf- und Schläfrigkeitsforschung etabliert und gehört zu den am häufigsten eingesetzten subjektiven Verfahren zur Messung von Schläfrigkeit. Die KSS ist ein zehnstufiges Ratingverfahren (siehe Tabelle 15) und dient zur Einschätzung des jeweiligen Grades der Wachheit (vgl. Abschnitt 1.4, Teil I). Die KSS-Messungen werden entsprechend dem Messprotokoll in regelmäßigen Zeitabständen wiederholt und geben so jeweils Momentaufnahmen des gegenwärtigen Schläfrigkeitszustandes wieder. Die Häufigkeitsverteilung der selbstberichteten KSS-Rohwerte ($M = 5.06$; $SD = 2.02$) zeigt sich in der vorliegenden Studie wie folgt: 1- 2.9%; 2- 4.8%; 3 – 20.0%; 4- 15.2%; 5 – 15.2%; 6- 16.2%; 7– 10.5%; 8- 11.4%; 9 – 3.8%; 10 – 0%). Um eine weitergehende Objektivierung der Schläfrigkeitsmaße zu gewährleisten, wurden im folgenden lediglich die Messzeitpunkte weitergehend analysiert, an denen sich eine Übereinstimmung zwischen Fremd- (Versuchsleiter-) und Selbstrating der Schläfrigkeit ergab. Beurteilt wurde die Abweichung anhand der auch in der späteren Auswertung zum Einsatz kommenden, dichotomisierten KSS-Werte ($KSS > 6$).

Tabelle 15: Karolinska Sleepiness Scale und ihre prozentuelle Häufigkeitsverteilung (erweitert um die Stufe 10, siehe Marsalek et al., 2006)

Grad der Schläfrigkeit	Rating Skala	Häufigkeit
äußerst wach	1	2.9%
sehr wach	2	4.8%
normal wach	3	20.0%
ziemlich wach	4	15.2%
weder wach noch schläfrig	5	15.2%
etwas schläfrig	6	16.2%
schläfrig, ohne Mühe wach zu bleiben	7	10.5%
schläfrig, etwas Mühe wach zu bleiben	8	11.4%
schläfrig, große Mühe wach zu bleiben	9	3.8%
äußerst schläfrig, kann nicht wach bleiben	10	0%

1.4 Acoustic Sleepiness Analysis

Die *Acoustic Sleepiness Analysis* (ASA) besteht aus den fünf Phasen der Sprachaufnahme, Vorverarbeitung, Merkmalsberechnung, Dimensionalitätsreduktion und Klassifikation. Die einzelnen Schritte wurden bereits in Kapitel 4 (Teil I), ‘Akustischer Messprozess als Machine-Learning-Paradigma’, auf allgemeiner Ebene beschrieben. In den folgenden Abschnitten wird nun die konkrete Ausgestaltung der Acoustic Sleepiness Analysis ausgeführt.

Sprachaufnahme und Sprachmaterial. Das Sprachmaterial wurde in ein handelsübliches Ansteck-Mikrofon (Sony ECM/C115) gesprochen und mit einem DAT-Rekorder digital aufgezeichnet (22.05 kHz; 16 Bit; mono). Es wurden verschiedene Sprachmaterialien eingesetzt. Zum Einsatz kam die Phonation eines 3 Sekunden lang gehaltenen [a:], eine Probanden-Intro Phrase und eine Bild Beschreibung. Im gelesenen Sprachmodus wurden 9 Fahrerassistenzsysteme verwendet, wie sie im Kontext von „In-Car-Speech-Dialog Systemen“ denkbar sind.

Sprachmaterial: “Isolierte Vokalphonation [a:]“

Entspannt gesprochenes, ca. 3 Sekunden lang gezogenes “[a:]“.

Sprachmaterial: “Intro“

Zu Beginn jeder Sprachaufnahmen nannten die Versuchspersonen die Uhrzeit, ihre Versuchspersonennummer und ihren Schläfrigkeitwert: z.B. “19.50 Uhr, Vp 10, Sleepiness Scale Wert 1“.

Sprachmaterial: “freie Bildbeschreibung“

Spontansprache. Die Versuchspersonen suchten sich zu Beginn der Untersuchung aus einer Zeitschrift ein Bild aus, das in der Untersuchung wiederholt mit eigenen Worten beschrieben wurde.

Sprachmaterial: “Fahrerassistenzsätze“

Neun Fahrerassistenzsätze wurden gelesen. Folgende Untermenge ging in die weitere Analyse ein: “Ich suche die Friesenstraße“ (Satz 1), “Das ist falsch, es muss ein Irrtum sein“ (Satz 3), “Suchen Sie die nächste Raststätte heraus“ (Satz 4), “Am Nachmittag möchte ich nach Hamburg“ (Satz 6) und “Rufen Sie den Notdienst an“ (Satz 9).

Aufgrund der großen Menge segmentierter Daten konnte in der vorliegenden Arbeit nur ausgewähltes Sprachmaterial ausgewertet werden. In die Datenverarbeitung gingen daher nur folgende Sprachsegmente ein: extrahierter Vokal [a:] aus dem Fahrerassistenzsatz 6 („Nachmittag), extrahierter Vokal [o:] aus dem Fahrerassistenzsatz 9 („Notdienst“), isolierte Vokalphonation [a:], Fahrerassistenzsatz 3, Fahrerassistenzsatz 9, Fahrerassistenzsatz 4 – Wort 1 („suchen“), Intro und die freie Bildbeschreibung. Die Auswahl tatsächlich in die Analyse Eingang findenden Sprachaufnahmen orientierte sich an den selbstberichteten KSS-Werten. Es wurden Messungen ausgewählt, die jeweils die niedrigsten respektive höchsten KSS-Werten aufwiesen.

So resultierten für die Sprachmaterialien folgende Sprachsample Verteilungen: Extrahierter Vokal [a:] [51 wach / 21 müde]; Extrahierter Vokal [o:] [51 wach / 21 müde]; Isolierte Vokalphonation [a:] [67 wach / 25 müde]; Fahrerassistenzsatz 4 - Wort 1 [49 wach / 19 müde]; Fahrerassistenzsatz 3 [51 wach / 21 müde]; Fahrerassistenzsatz 9 [51 wach / 21 müde]; Intro-Phrase [103 wach / 33 müde]; freie Bildbeschreibung [63 wach / 21 müde].

Vorverarbeitung

Die Analyse der Rohdaten wurde mit der Sprachanalysesoftware Praat (Boersma & Weenik, 1996) durchgeführt. Zunächst wurden die vier Messzeitpunkte pro Versuchsperson nach visuellen und auditiven Kriterien segmentiert. Ferner führte Praat für die Berechnung der fensterbezogenen Kennwerte ein Framing (20 ms, mit 10 ms Überlappung der Signalrahmen) und eine Hamming Fensterung durch.

Merkmalsberechnung

Die Merkmalsextraktion wurde mit Praat durchgeführt. Die Pitchanalyse wurde mit folgenden Parametereinstellungen erstellt: *Sprungweite der Analyseframes (Time step; 10ms)*, *Framelänge (Länge des Analyseframes; 40ms)*, *Pitch floor* (unterste berücksichtigte Frequenz; 75 Hz), *Pitch ceiling* (oberste berücksichtigte Frequenz; 600 Hz), *Silence threshold* (undefinierte F0, wenn die Amplitude des Frames kleiner als 0.03 mal der globalen Maximumamplitude ist; 0.03), *Voicing threshold* (undefinierte F0, wenn die berechnete Autokorrelation kleiner als 0.45 mal der maximalen Autokorrelation ist. Um mehr unvoiced Entscheidungen für eine Frame zu erzielen, muss die Schwelle erhöht werden; 0.45), *Octave cost* (Maß der

Bevorzugung hochfrequenter F0 Kandidaten; 0.01 pro Oktave), *Octave-jump cost* (Maß der Nicht-Bevorzugung von großen F0 Sprüngen; 0.35) und *Voiced- unvoiced cost* (Maß der Nicht-Bevorzugung von voiced-unvoiced Übergängen; Um die Zahl der Übergänge zu reduzieren, muss der Wert erhöht werden; 0.14).

Die Formanten (1-5) wurden unter Verwendung eines Preemphasis Filters mit einer Frequency Response von 20 ms, einem Hamming Window von 10 ms Schrittweite berechnet. Insgesamt wurden bei der Merkmalsextraktion jeweils 169 Rohkennwerte pro Sprach-Sample extrahiert und anschließend z-normalisiert. So entstanden 338 Kennwerte (169 Rohdaten, 169 z-normalisierte Kennwerte), die in die Merkmalsselektion eingingen. In der vorliegenden Untersuchung wurden folgende Kennwerte extrahiert:

Es wurden 26 prosodiebezogene Kennwerte berechnet. Die Verläufe von Grundfrequenz (F_0) und Intensität wurden extrahiert und folgende Funktionale abgeleitet: Mittelwert, 2.-4. Quartil, Standardabweichung; Maximum, Minimum, Bandbreite sowie Position und Werte der Maxima und Minima. Des Weiteren wurden die Verläufe von Jitter und Shimmer extrahiert und Mittelwerte berechnet.

Des Weiteren wurden 107 spektrale Kennwerte bestimmt, darunter Frequenz und Bandbreite der Formanten 1-5. Darüber hinaus wurden 4 Hammarberg Indices, das LTAS (Long Term Average Spectrum) für 6 Frequenzbänder (125-200Hz, 200-300Hz, 500-600Hz, 1000-1600Hz, 5000-8000Hz), das Verhältnis von niedriger Frequenzenergie unter 500Hz und unter 1000Hz, die Steigung der Regressionsgeraden (Slope) der spektralen Energie über 1000Hz, die Harmonics-to-Noise Ratio (HNR) sowie spektrale Tilt-Features ('Open Quotient', 'Glottal Opening', 'Skewness of Glottal Puls', 'Rate of Glottal Closure') (vgl. Stevens & Hanson, 1994) berechnet.

Ferner wurden 36 MFCC Kennwerte berechnet, die sich aus 12 MFCC, 12 Δ MFCC, 12 $\Delta\Delta$ MFCC zusammensetzen. Um diese Koeffizienten zu berechnen, wurden der fensterweise berechnete Mel Frequency Cepstral Coefficient Verlauf sowie sein erste und zweite Differenzfunktion (Ableitung) jeweils gemittelt.

Einen Überblick über alle berechneten Rohfeatures liefert folgende Auflistung:

- **Intensität**

- *int_min*: Minimumwert des Verlaufs der Intensität
- *int_timemin*: zeitliche Position des Minimums des Verlaufs der Intensität
- *int_max*: Maximumwert des Verlaufs der Intensität
- *int_timemax*: zeitliche Position des Maximums des Verlaufs der Intensität
- *int_mean*: Mittelwert des Verlaufs der Intensität
- *int_sd*: Standardabweichung des Verlaufs der Intensität
- *int_25quant*: 25stes Perzentil des Intensitätsverlaufs
- *int_75quant*: 75stes Perzentil des Intensitätsverlaufs
- *int_mintime_rel*: relative zeitliche Position des Minimums des Verlaufs der Intensität
- *int_maxtime_rel*: relative zeitliche Position des Maximums des Verlaufs der Intensität
- *int_range*: Range des Verlaufs der Intensität
- *shimmer*: Mittelwert des Lautstärkeperturbationsverlaufs

- **Fundamentalfrequenz**

- *f0_min*: Minimumwert des Verlaufs der Fundamentalfrequenz
- *f0_timemin*: zeitliche Position des Minimums des Verlaufs der Fundamentalfrequenz
- *f0_mintime_rel*: relative zeitliche Position des Minimums des Verlaufs der Fundamentalfrequenz
- *f0_max*: Maximumwert des Verlaufs der Fundamentalfrequenz
- *f0_timemax*: zeitliche Position des Maximums des Verlaufs der Fundamentalfrequenz
- *f0_maxtime_rel*: relative zeitliche Position des Maximums des Verlaufs der Fundamentalfrequenz
- *f0_range*: Range des Verlaufs der Fundamentalfrequenz
- *f0_mean*: Mittelwert des Verlaufs der Fundamentalfrequenz
- *f0_25quant*: 25stes Perzentil des Fundamentalfrequenzverlaufs
- *f0_75quant*: 75stes Perzentil des Fundamentalfrequenzverlaufs
- *f0_sd*: Standardabweichung des Verlaufs der Fundamentalfrequenz
- *f0_slope*: Regressionssteigung des Fundamentalfrequenzverlaufs
- *f0_slope_octjump*: Regressionssteigung des Fundamentalfrequenzverlaufs (Octave Jump)
- *jitter*: Mittelwert des Tonhöhenperturbationsverlaufs

- **Formanten**
 - *f1_min*: Minimumwert des Verlaufs der ersten Formantenfrequenz
 - *f1_mintime*: zeitliche Position des Minimums des Verlaufs der ersten Formantenfrequenz
 - *f1_mintime_rel*: relative zeitliche Position des Minimums des Verlaufs der ersten Formantenfrequenz
 - *f1_max*: Maximumwert des Verlaufs der ersten Formantenfrequenz
 - *f1_maxtime*: zeitliche Position des Maximums des Verlaufs der ersten Formantenfrequenz
 - *f1_maxtime_rel*: relative zeitliche Position des Maximums des Verlaufs der ersten Formantenfrequenz
 - *f1_range*: Range des Verlaufs der ersten Formantenfrequenz
 - *f1_mean*: Mittelwert des Verlaufs der ersten Formantenfrequenz
 - *f1_sd*: Standardabweichung des Verlaufs der ersten Formantenfrequenz
 - *f1_bandw_q25*: 25stes Perzentil des Bandbreitenverlaufs der ersten Formantenfrequenz
 - *f1_bandw_q50*: 50stes Perzentil des Bandbreitenverlaufs der ersten Formantenfrequenz
 - *f1_bandw_q75*: 75stes Perzentil des Bandbreitenverlaufs der ersten Formantenfrequenz
 - *f2_min*: Minimumwert des Verlaufs der zweiten Formantenfrequenz
 - *f2_mintime*: zeitliche Position des Minimums des Verlaufs der zweiten Formantenfrequenz
 - *f2_mintime_rel*: relative zeitliche Position des Minimums des Verlaufs der zweiten Formantenfrequenz
 - *f2_max*: Maximumwert des Verlaufs der zweiten Formantenfrequenz
 - *f2_maxtime*: zeitliche Position des Maximums des Verlaufs der zweiten Formantenfrequenz
 - *f2_maxtime_rel*: relative zeitliche Position des Maximums des Verlaufs der zweiten Formantenfrequenz
 - *f2_range*: Range des Verlaufs der zweiten Formantenfrequenz
 - *f2_mean*: Mittelwert des Verlaufs der zweiten Formantenfrequenz
 - *f2_sd*: Standardabweichung des Verlaufs der zweiten Formantenfrequenz

- *f2_bandw_q25*: 25stes Perzentil des Bandbreitenverlaufs der zweiten Formantenfrequenz
- *f2_bandw_q50*: 50stes Perzentil des Bandbreitenverlaufs der zweiten Formantenfrequenz
- *f2_bandw_q75*: 75stes Perzentil des Bandbreitenverlaufs der zweiten Formantenfrequenz
- *f3_min*: Minimumwert des Verlaufs der dritten Formantenfrequenz
- *f3_mintime*: zeitliche Position des Minimums des Verlaufs der dritten Formantenfrequenz
- *f3_mintime_rel*: relative zeitliche Position des Minimums des Verlaufs der dritten Formantenfrequenz
- *f3_max*: Maximumwert des Verlaufs der dritten Formantenfrequenz
- *f3_maxtime*: zeitliche Position des Maximums des Verlaufs der dritten Formantenfrequenz
- *f3_maxtime_rel*: relative zeitliche Position des Maximums des Verlaufs der dritten Formantenfrequenz
- *f3_range*: Range des Verlaufs der dritten Formantenfrequenz
- *f3_mean*: Mittelwert des Verlaufs der dritten Formantenfrequenz
- *f3_sd*: Standardabweichung des Verlaufs der dritten Formantenfrequenz
- *f3_bandw_q25*: 25stes Perzentil des Bandbreitenverlaufs der dritten Formantenfrequenz
- *f3_bandw_q50*: 50stes Perzentil des Bandbreitenverlaufs der dritten Formantenfrequenz
- *f3_bandw_q75*: 75stes Perzentil des Bandbreitenverlaufs der dritten Formantenfrequenz
- *f4_min*: Minimumwert des Verlaufs der vierten Formantenfrequenz
- *f4_mintime*: zeitliche Position des Minimums des Verlaufs der vierten Formantenfrequenz
- *f4_mintime_rel*: relative zeitliche Position des Minimums des Verlaufs der vierten Formantenfrequenz
- *f4_max*: Maximumwert des Verlaufs der vierten Formantenfrequenz
- *f4_maxtime*: zeitliche Position des Maximums des Verlaufs der vierten Formantenfrequenz

- *f4_maxtime_rel*: relative zeitliche Position des Maximums des Verlaufs der vierten Formantenfrequenz
- *f4_range*: Range des Verlaufs der vierten Formantenfrequenz
- *f4_mean*: Mittelwert des Verlaufs der vierten Formantenfrequenz
- *f4_sd*: Standardabweichung des Verlaufs der vierten Formantenfrequenz
- *f4_bandw_q25*: 25stes Perzentil des Bandbreitenverlaufs der vierten Formantenfrequenz
- *f4_bandw_q50*: 50stes Perzentil des Bandbreitenverlaufs der vierten Formantenfrequenz
- *f4_bandw_q75*: 75stes Perzentil des Bandbreitenverlaufs der vierten Formantenfrequenz
- *f5_min*: Minimumwert des Verlaufs der fünften Formantenfrequenz
- *f5_mintime*: zeitliche Position des Minimums des Verlaufs der fünften Formantenfrequenz
- *f5_mintime_rel*: relative zeitliche Position des Minimums des Verlaufs der fünften Formantenfrequenz
- *f5_max*: Maximumwert des Verlaufs der fünften Formantenfrequenz
- *f5_maxtime*: zeitliche Position des Maximums des Verlaufs der fünften Formantenfrequenz
- *f5_maxtime_rel*: relative zeitliche Position des Maximums des Verlaufs der fünften Formantenfrequenz
- *f5_range*: Range des Verlaufs der fünften Formantenfrequenz
- *f5_mean*: Mittelwert des Verlaufs der fünften Formantenfrequenz
- *f5_sd*: Standardabweichung des Verlaufs der fünften Formantenfrequenz
- *f5_bandw_q25*: 25stes Perzentil des Bandbreitenverlaufs der fünften Formantenfrequenz
- *f5_bandw_q50*: 50stes Perzentil des Bandbreitenverlaufs der fünften Formantenfrequenz
- *f5_bandw_q75*: 75stes Perzentil des Bandbreitenverlaufs der fünften Formantenfrequenz
- *df1f2*: Differenz der Mittelwerte der Verläufe der ersten und zweiten Formantfrequenz
- *df1f3*: Differenz der Mittelwerte der Verläufe der ersten und fünften Formantfrequenz
- *df1f4*: Differenz der Mittelwerte der Verläufe der ersten und fünften Formantfrequenz
- *df1f5*: Differenz der Mittelwerte der Verläufe der ersten und fünften Formantfrequenz

- $df2f3$: Differenz der Mittelwerte der Verläufe der zweiten und dritten Formantfrequenz
- $df2f4$: Differenz der Mittelwerte der Verläufe der zweiten und vierten Formantfrequenz
- $df2f5$: Differenz der Mittelwerte der Verläufe der zweiten und fünften Formantfrequenz
- $df3f4$: Differenz der Mittelwerte der Verläufe der dritten und vierten Formantfrequenz
- $df3f5$: Differenz der Mittelwerte der Verläufe der dritten und fünften Formantfrequenz
- $df4f5$: Differenz der Mittelwerte der Verläufe der vierten und fünften Formantfrequenz

- **Mel Frequency Cepstrum Coefficients (MFCC)**
 - $MFCC1$: Mittelwert des Verlaufs des MFCC1
 - $MFCC2$: Mittelwert des Verlaufs des MFCC2
 - $MFCC3$: Mittelwert des Verlaufs des MFCC3
 - $MFCC4$: Mittelwert des Verlaufs des MFCC4
 - $MFCC5$: Mittelwert des Verlaufs des MFCC5
 - $MFCC6$: Mittelwert des Verlaufs des MFCC6
 - $MFCC7$: Mittelwert des Verlaufs des MFCC7
 - $MFCC8$: Mittelwert des Verlaufs des MFCC8
 - $MFCC9$: Mittelwert des Verlaufs des MFCC9
 - $MFCC10$: Mittelwert des Verlaufs des MFCC10
 - $MFCC11$: Mittelwert des Verlaufs des MFCC11
 - $MFCC12$: Mittelwert des Verlaufs des MFCC12

 - $\Delta MFCC1$: Mittelwert der ersten Ableitung des Verlaufs des MFCC1
 - $\Delta MFCC2$: Mittelwert der ersten Ableitung des Verlaufs des MFCC2
 - $\Delta MFCC3$: Mittelwert der ersten Ableitung des Verlaufs des MFCC3
 - $\Delta MFCC4$: Mittelwert der ersten Ableitung des Verlaufs des MFCC4
 - $\Delta MFCC5$: Mittelwert der ersten Ableitung des Verlaufs des MFCC5
 - $\Delta MFCC6$: Mittelwert der ersten Ableitung des Verlaufs des MFCC6
 - $\Delta MFCC7$: Mittelwert der ersten Ableitung des Verlaufs des MFCC7
 - $\Delta MFCC8$: Mittelwert der ersten Ableitung des Verlaufs des MFCC8
 - $\Delta MFCC9$: Mittelwert der ersten Ableitung des Verlaufs des MFCC9
 - $\Delta MFCC10$: Mittelwert der ersten Ableitung des Verlaufs des MFCC10

- *Delta MFCC11*: Mittelwert der ersten Ableitung des Verlaufs des MFCC11
- *Delta MFCC12*: Mittelwert der ersten Ableitung des Verlaufs des MFCC12
- *Delta Delta MFCC1*: Mittelwert der zweiten Ableitung des Verlaufs des MFCC1
- *Delta Delta MFCC2*: Mittelwert der zweiten Ableitung des Verlaufs des MFCC2
- *Delta Delta MFCC3*: Mittelwert der zweiten Ableitung des Verlaufs des MFCC3
- *Delta Delta MFCC4*: Mittelwert der zweiten Ableitung des Verlaufs des MFCC4
- *Delta Delta MFCC5*: Mittelwert der zweiten Ableitung des Verlaufs des MFCC5
- *Delta Delta MFCC6*: Mittelwert der zweiten Ableitung des Verlaufs des MFCC6
- *Delta Delta MFCC7*: Mittelwert der zweiten Ableitung des Verlaufs des MFCC7
- *Delta Delta MFCC8*: Mittelwert der zweiten Ableitung des Verlaufs des MFCC8
- *Delta Delta MFCC9*: Mittelwert der zweiten Ableitung des Verlaufs des MFCC9
- *Delta Delta MFCC10*: Mittelwert der zweiten Ableitung des Verlaufs des MFCC10
- *Delta Delta MFCC11*: Mittelwert der zweiten Ableitung des Verlaufs des MFCC11
- *Delta Delta MFCC12*: Mittelwert der zweiten Ableitung des Verlaufs des MFCC12

- **Spektrum**
 - *openQuot_OQG*: Open Quotient Gradient (Differenz des Amplitudenspektrumswerts der Fundamentalfrequenz und des Amplitudenspektrumswerts der doppelten Fundamentalfrequenz geteilt durch Frequenzabstand von F_0 und $2 \cdot F_0$)
 - *glottOpen_GOG*: Glottal Opening Gradient (Differenz des Amplitudenspektrumswerts der Fundamentalfrequenz und des Amplitudenspektrumswerts der ersten Formantenfrequenz geteilt Frequenzabstand von F_0 und F_1)
 - *skewness_SKG*: Spectral Skewness Gradient: (Differenz des Amplitudenspektrumswerts der Fundamentalfrequenz und des Amplitudenspektrumswerts der zweiten Formantenfrequenz geteilt Frequenzabstand von F_0 und F_2)
 - *rateClos_RCG*: Rate Closure Gradient (Differenz des Amplitudenspektrumswerts der Fundamentalfrequenz und des Amplitudenspektrumswerts der dritten Formantenfrequenz geteilt Frequenzabstand von F_0 und F_3)
 - *imcomplclos_IC*: Incompleted Closure Gradient (Bandwidth des ersten Formanten durch Frequenzposition von F_1)
 - *HNR*: Mittelwert des Harmonics-to-Noise Ratio Verlaufs
 - *energmax0_2000*: Maximum des Amplitudenspektrumswerts im Frequenzband von 0 bis 2000Hz

- *energmax2000_5000*: Maximum des Amplitudenspektrumwerts im Frequenzband von 2000 bis 5000Hz
- *hammarberg_s0*: Differenz des Maximum des Amplitudenspektrumwerts im Frequenzband von 0 bis 2000Hz und 2000 bis 5000Hz
- *hammarberg_s1*: Differenz des Amplitudenspektrumwerts im Frequenzband von 400 bis 600Hz und des Amplitudenspektrumwerts von F0
- *hammarberg_s2*: Differenz des Amplitudenspektrumwerts im Frequenzband von 400 bis 600Hz und des Amplitudenspektrumwerts von 1600Hz
- *hammarberg_s3*: Differenz des Amplitudenspektrumwerts im Frequenzband von 400 bis 600Hz und des Amplitudenspektrumwerts von 5000Hz
- *hammarberg_s4*: Differenz des Amplitudenspektrumwerts im Frequenzband von 400 bis 600Hz und des Amplitudenspektrumwerts im Frequenzband von über 5000Hz
- *spec_cm*: Central Moment des Amplitudenspektrums
- *spec_kurt*: Kurtosis des Amplitudenspektrums
- *spec_skew*: Schiefe des Amplitudenspektrums
- *spec_sd*: Standardabweichung des Amplitudenspektrums
- *spec_cog*: Schwerpunkt des Amplitudenspektrums
- *baen0_1000*: Spektrale Energiedichte im Frequenzband von 0 bis 1000Hz
- *baen1000_8000*: Spektrale Energiedichte im Frequenzband von 1000 bis 8000Hz
- *baen_hf_lf1000*: Spektrale Energiedichtequotient kleiner 1000Hz versus Energiedichte größer 1000Hz
- *baen125_200*: Spektrale Energiedichte im Frequenzband von 125 bis 200Hz
- *baen200_300*: Spektrale Energiedichte im Frequenzband von 200 bis 300Hz
- *baen500_600*: Spektrale Energiedichte im Frequenzband von 500 bis 600Hz
- *baen1000_1600*: Spektrale Energiedichte im Frequenzband von 1000 bis 1600Hz
- *baen5000_8000*: Spektrale Energiedichte im Frequenzband von 5000 bis 8000Hz
- *bade200_1000*: Spektrale Energiedichte im Frequenzband von 200 bis 1000Hz
- *baendi0_500_500_4000*: Spektrale Energiedichtedifferenz im Frequenzband 0 bis 500Hz und 500 bis 4000Hz
- *badeddiff0_500_500_1000*: Spektrale Energiedichtedifferenz im Frequenzband (0 bis 500Hz) und 500 bis 1000Hz
- *ltas_slope0_1000_1000_4000*: Long term Average Spectrum Slope des Frequenzbands 0 bis 1000Hz versus 1000 bis 4000Hz
- *ltas_min*: Minimum des Long Term Average Spectrum

- *ltas_minfreq*: Position der Minimumfrequenz des Long Term Average Spectrum
- *ltas_max*: Maximum des Long Term Average Spectrum
- *ltas_maxfreq*: Position der Maximumfrequenz des Long Term Average Spectrum
- *ltas_sd*: Standardabweichung des Long Term Average Spectrum
- *ltas_locpeak*: Lokale Peakhöhe des Long Term Average Spectrum (1700 bis 4200Hz versus 2400 bis 3200Hz)
- *ltas_trend_max*: Trendlinie des Long Term Average Spectrum zwischen 1000 und 5000Hz
- *pe1000_baenqout*: relative spektrale Energiequotient der niedrigen (>1000Hz) versus hohen (<1000Hz) Frequenzbereichen

Dimensionalitätsreduktion

Aus den 338 Kennwerten wurden in einer supervised filterbasierten Feature Subset Selection mittels eines Korrelationsfilters ($r > .15$) Kennwerte mit der höchsten Relevanz (Maximum Relevance Kriterium) herausselektiert. Im Rahmen der Dimensionalitätsreduktion wurden des Weiteren Kennwerte mit einer unsupervised Feature Transformation Methode (Principal Component Analysis; PCA) auf 100 reduziert. Dabei wurde die Dimensionalität der Kennwertematrix reduziert, um eine optimale Klassifikationsleistung der Mustererkennungsverfahren (Klassifikatoren) zu ermöglichen. Die ersten 100 Faktoren wurden extrahiert und über eine Rücktransformation in den Merkmalsraum die Höhe der Item-Dimensionen auf 100 reduziert. Die Anzahl der Kennwerte wurde in diesen Schritten von 340 auf ungefähr 50 reduziert. In Rahmen der Hauptselektion wurde wrapper-basiert über evolutionäre Suchverfahren (Genetic Algorithm) mit verschiedenen Learning Algorithmen schrittweise ein optimaler Merkmalsverbund selektiert. Für jedes Sprachmaterial wurde das Sub-Feature-Set mit der höchsten Vorhersageleistung mittels iterativer Anwendung der Klassifikationsverfahren k-Nearest-Neighbour ($k = 1$), Multilayer-Perceptron mit Feedforward-Netz und Backpropagation-Learning-Algorithm (Input Layer Type: Linear; 2 Hidden Layer mit je 10 Neuronen; Hidden Layer Type: Sigmoid; Output Layer Type: Sigmoid; Training Cycles: 200; Learning Rate: 0.3; Momentum = 0.2; Error Epsilon = 0.05) und Support-Vector-Machine mit polynomialer Kernelfunktion (Gamma = $1/340$; Epsilon = 0.001; Degree of Polynom Function = 3; Cost = 0) ermittelt.

Klassifikation und Validierungsstrategie

Auf das gefundene Feature-Set wurden die Klassifikationsverfahren k-Nearest-Neighbour, Multilayer-Perceptron und Support-Vektor-Maschine angewendet, um die Gesamtvorhersageleistung der „Acoustic Sleepiness Analysis“ zu ermitteln und die prädiktive Validität des Verfahrens zu bestimmen. Die Daten wurden in Trainings- und Test-Set getrennt (90% Trainingsdaten, 10% Testdaten). Das Test-Set umfasste ausschließlich Daten, die im Trainings-Set nicht enthalten waren (ungesehene Daten). Alle Klassifikatoren wurden zehnmal trainiert. Dazu wurden die Testdaten in 10 disjunkte Sets gleicher Größe unterteilt. Jedes dieser 10 Sets war in einem der 10 Trainingsdurchläufe nicht Teil der Trainingsdaten und wurde stattdessen als Test-Set verwendet. In der Untersuchung wurden die Klassifikationsfehler somit mittels einer 10-fachen Kreuzvalidierung berechnet. Die 10-fache Kreuzvalidierung reduzierte den Rechenaufwand bei gleichzeitiger Sicherung der strikten Sprecherunabhängigkeit. Als finaler Klassifikationsfehler wurde der Mittelwert aus den Klassifikationsfehlern der 10 Test-Sets verwendet. Für eine weitere Verbesserung der Klassifikationsmodelle wurden die Daten in männliche und weibliche Sprecher getrennt, um geschlechtsspezifische Modelle berechnen zu können. Die Klassifikation wurde mit dem Multilayer-Perceptron mit Feedforward-Netz und Backpropagation-Learning-Algorithm (2 Hidden-Layers mit jeweils 10 Knoten), einer Support Vector Machine mit polynomialer Kernelfunktion und dem k-Nearest-Neighbour Verfahren mit einem Nachbarn durchgeführt. Zur signifikanzstatistischen Abschätzung der gefundenen Klassifikatorprädiktion wurde aufgrund der geringen Fallanzahl, anstelle eines Chi-Quadrat Tests, der exakte Fisher-Test durchgeführt. Weitere Validierungsindizes ergeben sich aus den korrelativen Zusammenhängen von Selbstreportgrößen und einzelnen Sprachkennwerten. Tabelle 16 fasst den gesamten akustischen Messprozess zusammen.

Tabelle 16: Gesamtüberblick über den ASA Phasenablauf und seine wichtigsten Spezifikationen

ASA Machine-Learning Phase	Spezifikationen
Aufnahme	
Sprachsegment: Vokal	Extrahiertes [a:], extrahiertes [o:], isoliertes [a:]
Sprachsegment: Wort	Extrahiertes Wort "suchen" (aus Fahrerassistenzsatz 4)
Sprachsegment: Phrasale Einheit	Fahrerassistenzsatz 3, Fahrerassistenzsatz 9, Probanden-Introphrase
Sprachsegment: Multiple Phrasale Einheit	Freie Bildbeschreibung
Vorverarbeitung	
Filterung	Filterung von unplausiblen Wertebereichen
Segmentierung	Manuelle Identifikation von Sprachsegmenten
Merkmalsberechnung	
Prosodie	F0-, Intensität-, Formantposition und Formantbandbreiten-(F1-F5) Deskriptoren
Spectral I	Bandenergy, LTAS Verteilung, Hammarberg etc.
Spectral II	Mel Frequency Cepstral Coefficients (Roh, Δ , $\Delta\Delta$)
Dimensionalitätsreduktion	
Filterbasierte Subset Selection	Korrelationsfilter
Unsupervised Feature Transformation	Principle Component Analysis (100 Principle Components)
Wrapper-basierte Subset Selection	1-Nearest Neighbour, Multi-Layer Perceptron, Support Vector Machine
Klassifikation	
1-Nearest Neighbour	1 bis 5 Nearest Neighbours
Multi Layer Perceptron	Feedforward, Backpropagation-Learning-Algorithm, Input Layer Type: Linear; 2 Hidden Layer, 10 Neurone; Hidden Layer Type: Sigmoid; Output Layer Type: Sigmoid; Training Cycles: 200; Learning Rate: 0.3; Momentum = 0.2; Error Epsilon = 0.05
Support Vector Machine	Polynomial Kernelfunktion, Gamma = 1/340; Epsilon = 0.001; Degree of Polynom Function = 3; Cost = 0
Evaluation	Trainings- und Testdaten Split über 10-fache Kreuzvalidierung

2 ERGEBNISSE: SPRACHSEGMENT VOKAL

Nachdem im vorangegangenen Kapitel die Experimentalmethodik der Untersuchung beschrieben worden ist, beginnt dieses Kapitel mit einem Kurzüberblick zu den Hauptergebnissen des Kapitels 2. Im Anschluss daran werden korrelativen Zusammenhänge zwischen sprachlichen Einzelmerkmalen dreier Vokale und Schläfrigkeitratings vorgestellt. Das nachfolgende Unterkapitel stellt die für die Validitätseinschätzung zentralen, multivariaten Machine-Learning Klassifikationsergebnisse dar.

Ergebniszusammenfassung Sprachsegment Vokal

Sprachsegment extrahierter Vokal [a:]. Die linearen Zusammenhänge von 42 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeitsszuständen erreichen eine mindestens kleine Effektstärke von $r > .2^1$. Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt eine Vorhersagegenauigkeit von 80.7% (Sensitivität = 80.4%, Spezifität = 80.9%).

Sprachsegment extrahierter Vokal [o:]. Die linearen Zusammenhänge von 45 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeitsszuständen erreichen eine mindestens kleine Effektstärke von $r > .2$. Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt eine Vorhersagegenauigkeit von 79.6% (Sensitivität = 60.7%, Spezifität = 83.3%).

Sprachsegment isolierte Vokalphonation [a:]. Die linearen Zusammenhänge von 19 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeitsszuständen erreichen eine mindestens kleine Effektstärke von $r > .2$. Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt eine Vorhersagegenauigkeit von 83.3% (Sensitivität = 53.6%, Spezifität = 94.1%).

¹ Basierend auf Effektstärkenüberlegungen wird hier als Kriterium der Erwähnenswürdigkeit von korrelativen Zusammenhängen das Relevanzkriterium gewählt. Zusammenhänge mit einer zumindest kleinen Effektstärke ($r > .2$) besitzen unabhängig von Signifikanzaussagen einen potenziellen Wert für weitergehende Untersuchungen und werden daher hier als inhaltlich bedeutsam eingestuft.

2.1 Ergebnisse: Einzelkennzahlen

In diesem Kapitel werden Zusammenhänge zwischen sprachlichen Einzelkennzahlen dreier Vokalsegmente und dem selbstberichteten Schläfrigkeitsniveau (KSS) präsentiert. Die Ergebnisse des aus dem Fahrerassistenzsatz 6 („Am Nachmittag möchte ich nach Hamburg“) extrahierten Vokals [a:], die Ergebnisse des aus dem Fahrerassistenzsatz 9 („Rufen Sie den Notdienst an“) extrahierten Vokals [o:], sowie die isolierte 2 Sekunden Vokalphonation des gesprochenen [a:] dienen hier als Sprachmaterial. Zur weiteren Illustration der gefundenen korrelativen Analyseergebnisse werden jeweils für einzelne wache und schläfrige Stimmaufnahmen Kennwertverläufe dargestellt.

Sprachmaterial „extrahierter Vokal [a:]“: korrelative Einzelergebnisse

Korrelative Zusammenhänge zwischen Stimmkennzahlen des extrahierten Vokals [a:] und Schläfrigkeitsratings. 42 Einzelkorrelationen zwischen Stimmkennzahlen des extrahierten Vokals [a:] und selbstberichteten Schläfrigkeitsratings erreichen eine mindestens kleine Effektstärke von $r > .2$. Die fünf betragsmäßig höchsten linearen Zusammenhänge werden für die folgenden Merkmale erzielt:

f4_maxtime_rel: -0.36 (relative zeitliche Position des Maximums des vierten Formanten); delta-delta-MFCC6_z: 0.30 (z-transformierter Mittelwert der zweiten Ableitung des Verlaufs des Mel-Frequency Cepstrum Coefficient 6); ltas_maxfreq_z: -0.33 (z-transformierte Position der Maximumfrequenz des Long Term Average Spectrum); spec_cog_z: -0.33 (z-transformierter Schwerpunkt des Amplitudenspektrums); delta-delta-MFCC11_z: 0.29 (z-transformierter Mittelwert der zweiten Ableitung des Verlaufs des Mel-Frequency Cepstrum Coefficient 11).

Tabelle 17: Korrelative Zusammenhänge ($r > .20$) zwischen Sprachmerkmalen des extrahierten Vokals [a:] und KSS-Werten.

f2_range	f2_sd	f3_min	f4_maxtime_rel	f5_mintime_rel	f5_max	df1f2
-.20	-.26	.27	-.36	-.21	.20	-.27
df1f3	delta-delta-MFCC7	baen_5000_8000	df1f5	delta-delta-MFCC3		
-.25	-.24	.20	-.20	-.21		
f1_mintime_rel_z	f2_range_z	f2_sd_z	f3_min_z	f4_maxtime_rel_z	f4_range_z	f4_sd_z
-.21	-.26	-.23	.21	-.27	-.26	-.26
f5_max_z	f5_mean_z	f5_min_z	df1f3_z	df1f5_z	MFCC3_z	delta-delta-MFCC6_z
.23	.24	.21	-.23	-.23	.27	.31
ltas_maxfreq_z	ltas_trend_max_z	energmax2k_5k_z	Hammarberg_z	delta-delta-MFCC10_z	delta-delta-MFCC11_z	delta-delta-MFCC12_z
-.33	-.24	-.24	.23	-.26	.29	.25
int_sd_z	int_25quant_z	int_75quant_z	int_max_z	int_mean_z	int_min	baen_hf_lf1000_z
.24	-.25	-.24	-.24	-.25	-.25	-.25
ltas_slope 1k_4k_z	spec_cog_z					
-.28	-.33					

Im Weiteren werden die aus den korrelativen Analysen gewonnenen Hinweise auf schläfrigkeitssensitive sprachliche Einzelmerkmale pointiert dargestellt und vertiefend verfolgt. Die aus den frame-level (Low-level Descriptors, LLD) gewonnenen und zu Basiskonturen verkettenen Merkmalsverläufe von wachen und schläfrigen Sprachaufnahmen der korrelativ bedeutungsvollsten Features werden zu diesem Zweck kontrastierend gegenübergestellt. Neben der Illustration der gefundenen korrelativen Analyseergebnisse dient diese Betrachtung auch zur Generierung von weiteren die temporale Verlaufsinformation verdichtenden Kennzahlen (Functionals).

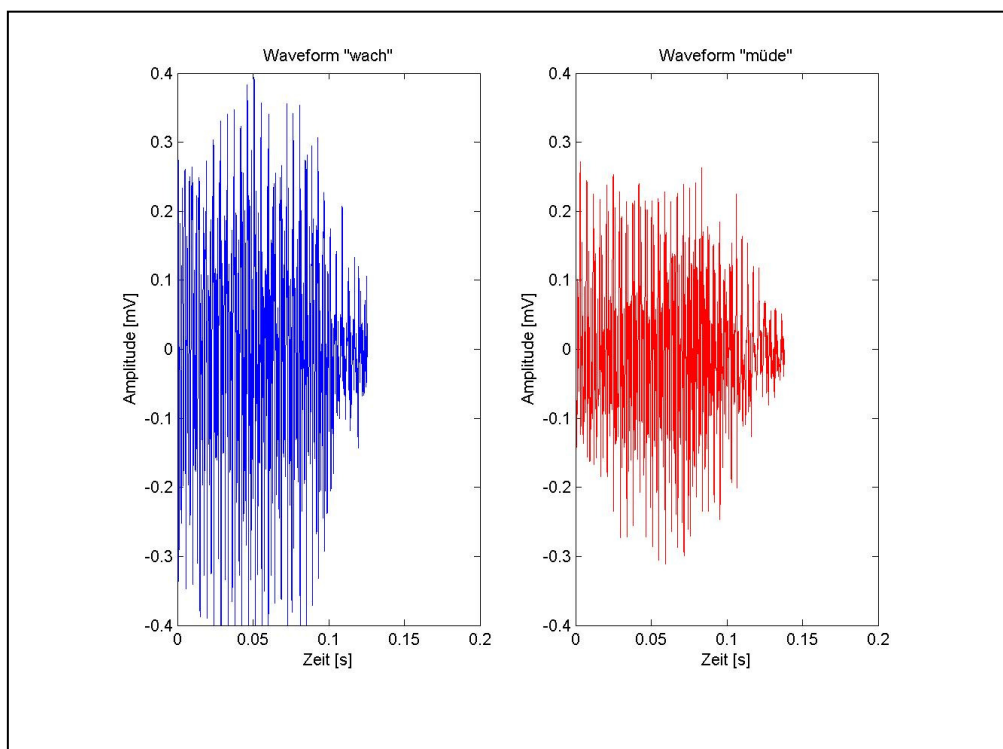


Abbildung 42: Rohwveform des extrahierten Vokals [a:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 11)

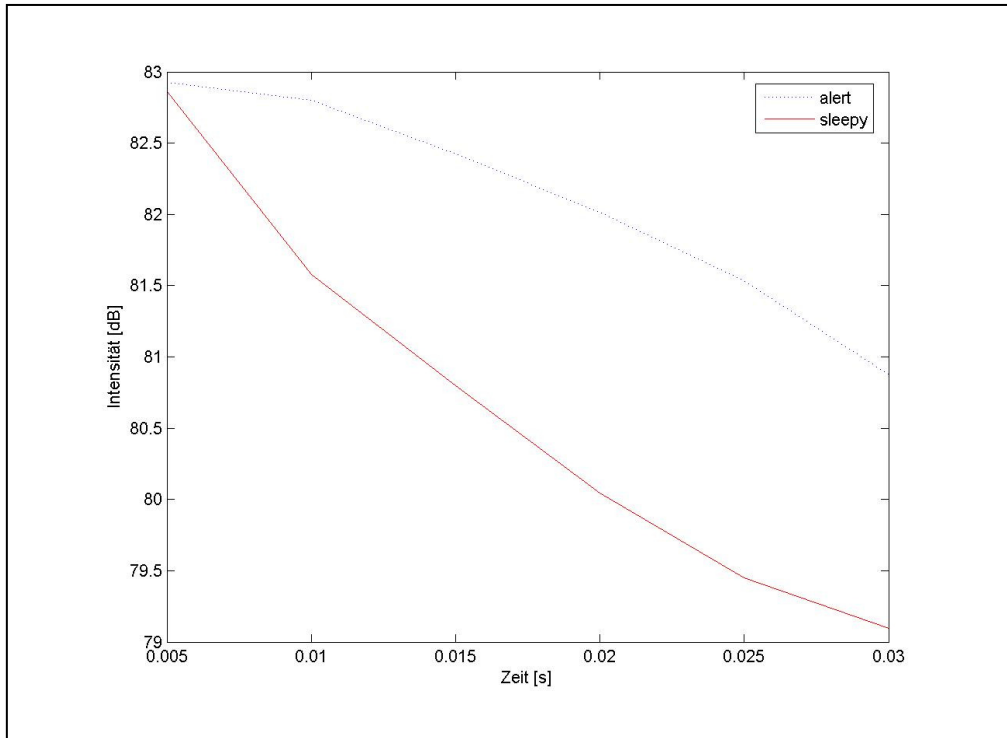


Abbildung 43: Intensitätsverlauf des extrahierten Vokals [a:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 14)

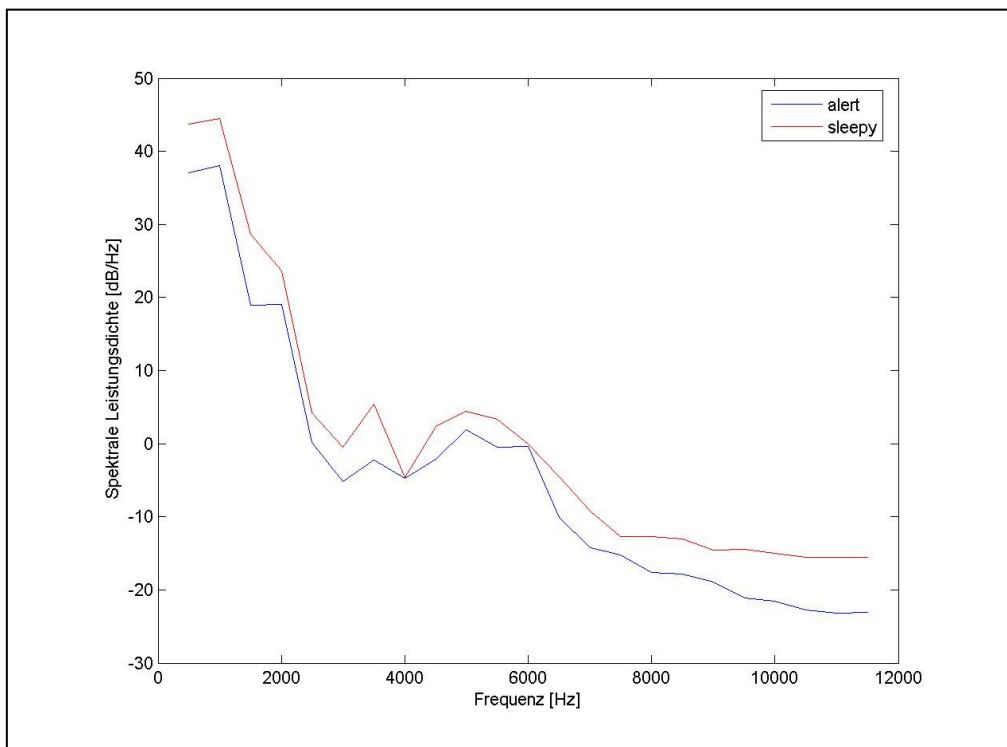


Abbildung 44: Long Term Average Spectrum des extrahierten Vokals [a:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 17)

Wie aus den Abbildungen 42, 43 und 44 ersichtlich wird, unterscheiden sich wache von schläfrigen Sprachaufnahmen in ihrem Verlauf der Rohwelleform (z.B. höhere Amplitude), in ihrem Intensitätsverlauf (z.B. höheres 25stes Perzentil, höheres 75stes Perzentil, höheres Maximum, höherer Mittelwert, höheres Minimum) und in ihrem Long Term Average Spectrum (z.B. tiefere Band Energy bei 5000-8000 Hz).

Sprachmaterial „extrahierter Vokal [o:]“: korrelative Einzelergebnisse

Korrelative Zusammenhänge zwischen Stimmkennzahlen des extrahierten Vokals [o:] und Schläfrigkeitsratings. 45 Einzelkorrelationen zwischen Stimmkennzahlen des extrahierten Vokals [o:] und selbstberichteten Schläfrigkeitsratings erreichen eine mindestens kleine Effektstärke von $r > .2$. Die fünf betragsmäßig höchsten linearen Zusammenhänge werden für die folgenden Merkmale erzielt:

f0_slope_z: -0.41 (z-transformierte Steigung der Regressionsgerade des Fundamental Frequenzverlaufs); Delta-MFCC7: 0.34 (Mittelwert der ersten Ableitung des Verlaufs des Mel-Frequency Cepstrum Coefficient 7); MFCC10_z: 0.32 (z-transformierter Mittelwert des Verlaufs des Mel-Frequency Cepstrum Coefficient 10); f0_range_z: -0.36 (z-transformierter Mittelwert des Ranges des Fundamental Frequenzverlaufs); f0_max_z: -0.33 (z-transformiertes Maximum des Fundamental Frequenzverlaufs).

Tabelle 18: Korrelative Zusammenhänge ($r > .20$) zwischen Sprachmerkmalen des extrahierten Vokals [o:] und KSS-Werten.

f0_max	f0_sd	f0_75quant	f0_range	f0_slope	f0_slope_octjump	Jitter	f2_mean	f3_mean
-.24	-.25	-.23	-.27	-.33	-.2	-.33	.25	.29
df1f2	Delta-MFCC7	Delta-delta-MFCC5	Delta-delta-MFCC8	Delta-delta-MFCC9	Delta-delta-MFCC10	int_sd	int_range	baen_hf_lf1000
-.25	.34	.21	-.26	-.27	.20	-.29	-.24	-.22
f0_max_z	f0_sd_z	f0_75quant_z	f0_mintime_rel_z	f0_range_z	f0_slope_z	f0_slope_octjump_z	Jitter_z	f1_max_z
-.33	-.36	-.30	-.24	-.36	-.41	-.29	-.23	.20
f1_maxtime_rel_z	f2_mean_z	f3_maxtime_rel_z	f3_mean_z	df1f2_z	df1f3_z	df1f4_z	df2f5_z	df3f5_z
.25	.26	.20	.27	-.25	-.23	-.26	.28	.22
df4f5_z	MFCC10_z	Delta-MFCC11_z	int_max_z	int_sd_z	int_75quant_z	hammarberg_s0_z	hammarberg_s3_z	ltas_trend_max_z
.24	.32	-.21	-.24	-.21	-.23	-.22	-.20	-.21

Im Weiteren werden die aus den korrelativen Analysen gewonnenen Hinweise auf schläfrigkeitssensitive sprachliche Einzelmerkmale ähnlich wie für das Sprachmaterial des extrahierten [a:] exemplarisch dargestellt.

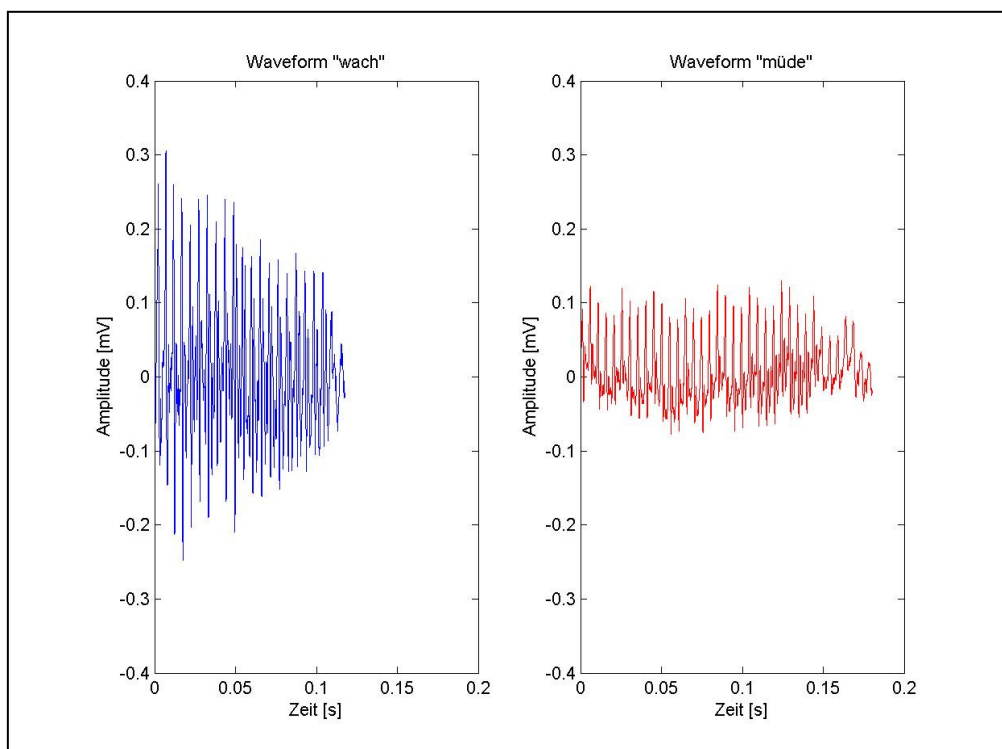


Abbildung 45: Rohwaveform des extrahierten Vokals [o:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 11)

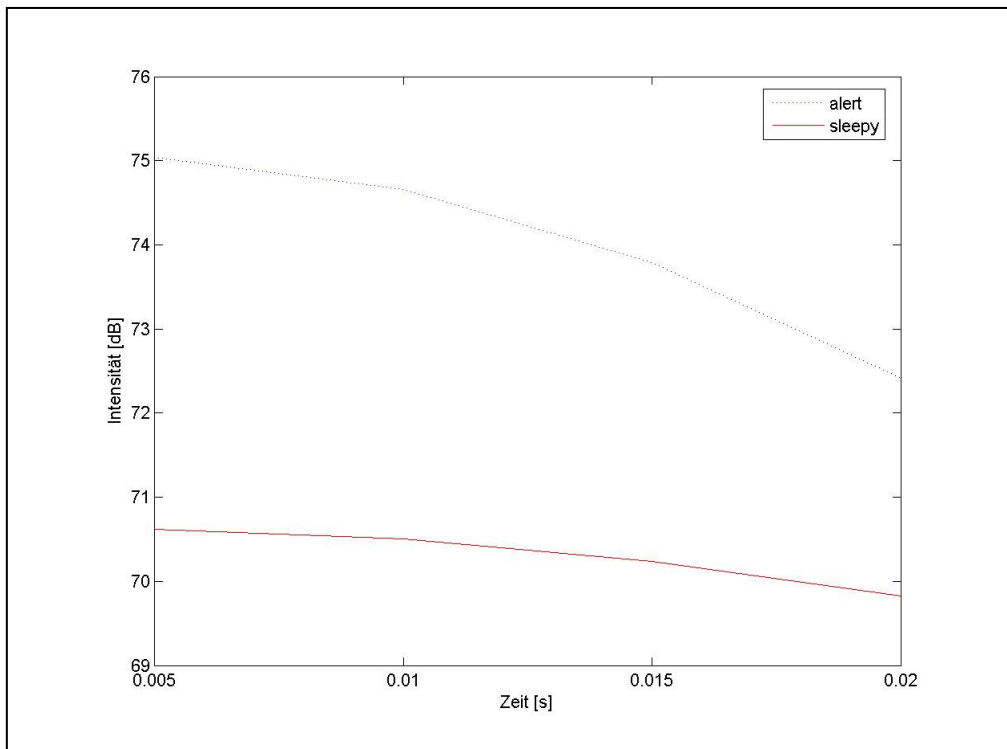


Abbildung 46: Intensitätsverlauf des extrahierten Vokals [o:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 16)

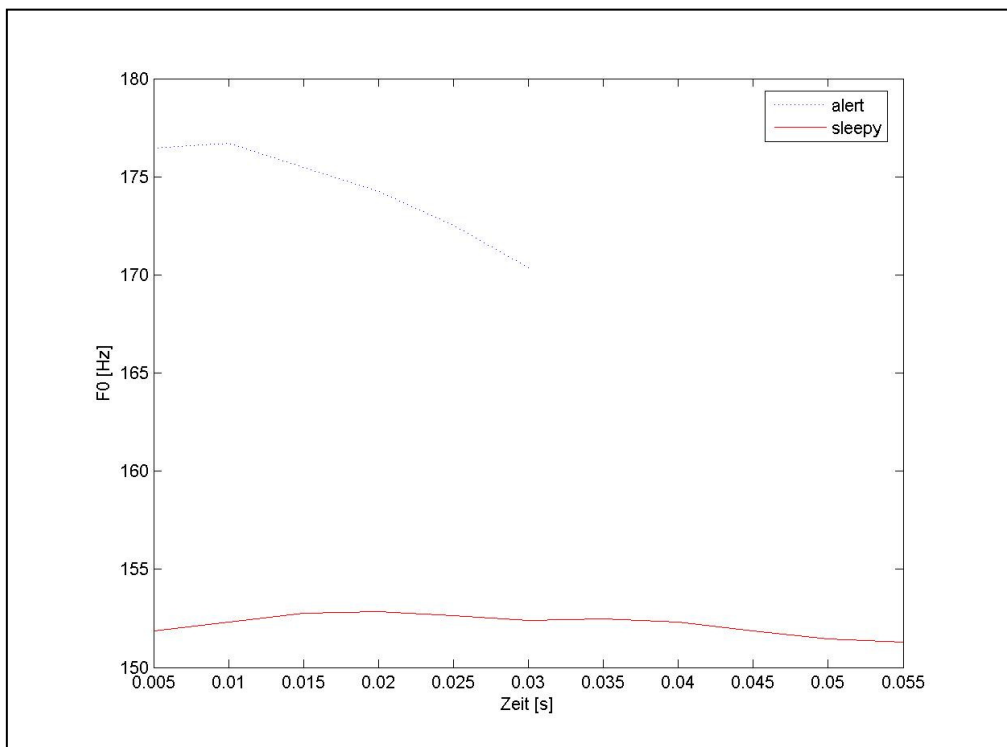


Abbildung 47: Fundamental Frequenzverlauf des extrahierten Vokals [o:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 12)

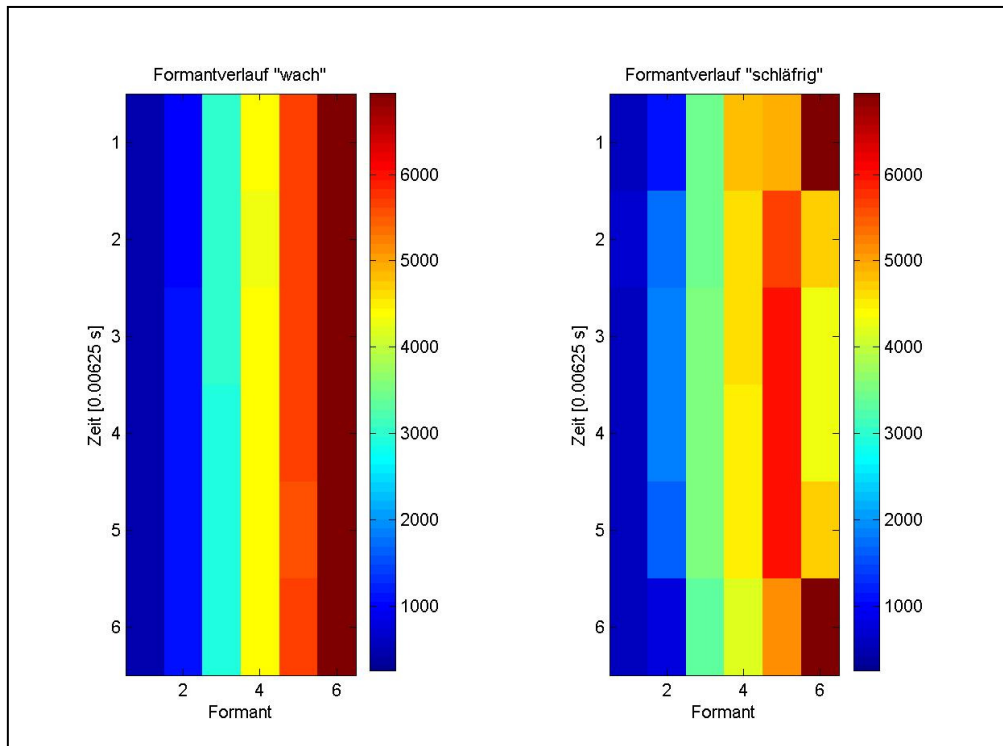


Abbildung 48: Formantverlufe des extrahierten Vokals [o:] fur ein waches ($KSS \leq 5$) und ein mudes ($KSS > 5$) Sprachbeispiel (Vp 16)

Wie aus den Abbildungen 45, 46, 47 und 48 ersichtlich wird, unterscheiden sich wache von schlufigen Sprachaufnahmen in ihrem Verlauf der Rohwaweform (z.B. hohere Amplitude), in ihrem Intensitatsverlauf (z.B. hohere Standardabweichung, hoherer Range), in ihrem Fundamental Frequenzverlauf (z.B. hohere Standardabweichung, hoheres Maximum) und in ihren Formantpositionsverlufen (z.B. tieferer Mittelwert von F2).

Sprachmaterial „Vokalphonation [a:]“: korrelative Einzelergebnisse

Korrelative Zusammenhange zwischen Stimmkennzahlen der isolierten Vokalphonation [a:] und Schlufigkeitsratings. 19 Einzelkorrelationen zwischen Stimmkennzahlen der isolierten Vokalphonation [a:] und selbstberichteten Schlufigkeitsratings erreichen eine mindestens kleine Effektstarke von $r > .2$. Die funf betragsmaig hochsten linearen Zusammenhange werden fur die folgenden Merkmale erzielt:

F0_sd_z: -.34 (Standardabweichung des z-transformierten Fundamentalfrequenz-Verlaufs);
 F0_slope_z: -.30 (Regressionssteigung des z-transformierten Fundamentalfrequenz-Verlaufs);

HNR: .26 (Mittelwert des Harmonics-to-Noise Ratio Verlaufs); F3_bandw_q25_z: .26 (z-transformiertes 25tes Perzentil der Bandwidth der dritten Formantfrequenz); F1_bandw_q50_z: .25 (z-transformiertes 50tes Perzentil der Bandwidth der ersten Formantfrequenz).

Tabelle 18: Korrelative Zusammenhänge ($r > .20$) zwischen Sprachmerkmalen der isolierten Vokalphonation [a:] und KSS-Werten.

Shimmer	HNR	F2_mintime	F2_bandw_q25	F1_bandw_q50	Df2f5	Delta-delta-MFCC 1	
-.21	.26	-.23	.25	.25	.20	-.25	
F0_max_z	F0_sd_z	F0_range_z	F0_slope_z	F2_bandw_q25_z	F3_bandw_q25_z	F3_bandw_q50_z	Df2f5_z
-.21	-.34	-.21	-.30	.20	.25	.21	.20
Delta-MFCC 6_z	Delta-MFCC 7_z	Open Quot. Glottis	Spec_sd_z				
-.20	.20	.23	-.24				

Im Weiteren werden die aus den korrelativen Analysen gewonnenen Hinweise auf schläfrigkeitssensitive sprachliche Einzelmerkmale ähnlich wie für das Sprachmaterial zuvor exemplarisch dargestellt.

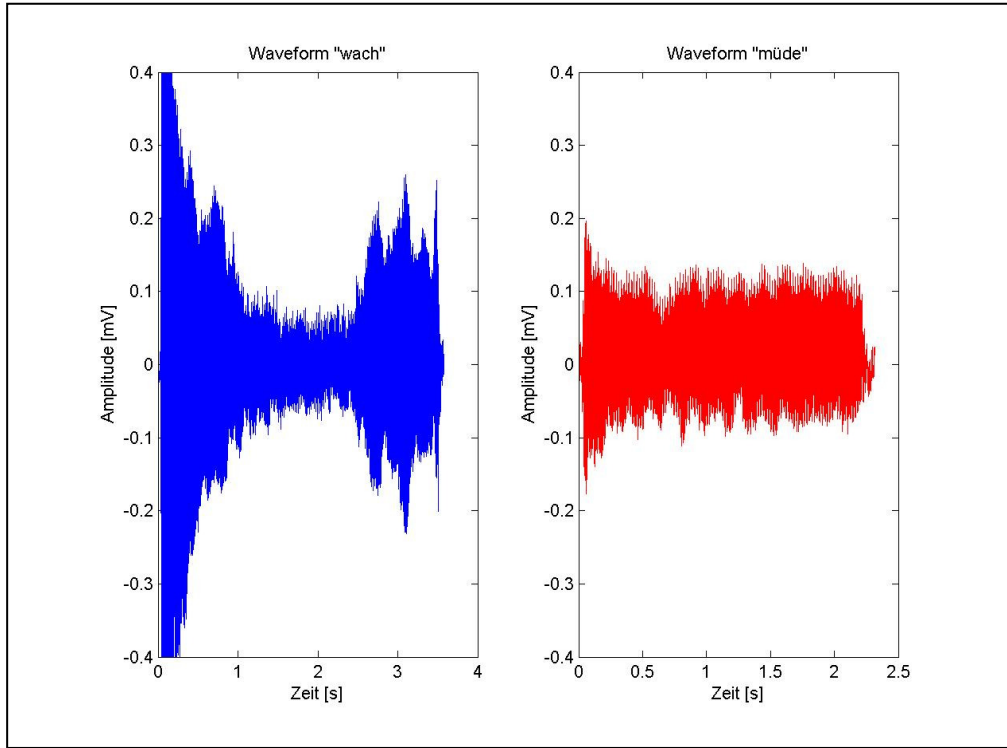


Abbildung 49: Rohwaveform Verlauf der isolierten Vokalphonation [a:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 11)

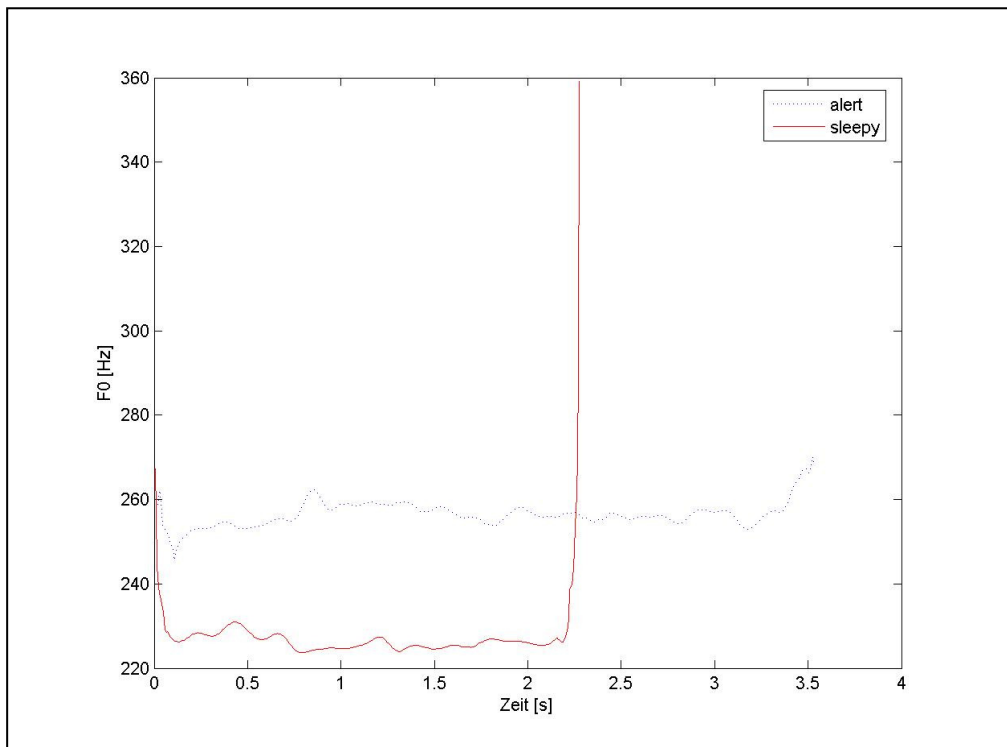


Abbildung 50: Fundamental Frequenzverlauf der isolierten Vokalphonation [a:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 11)

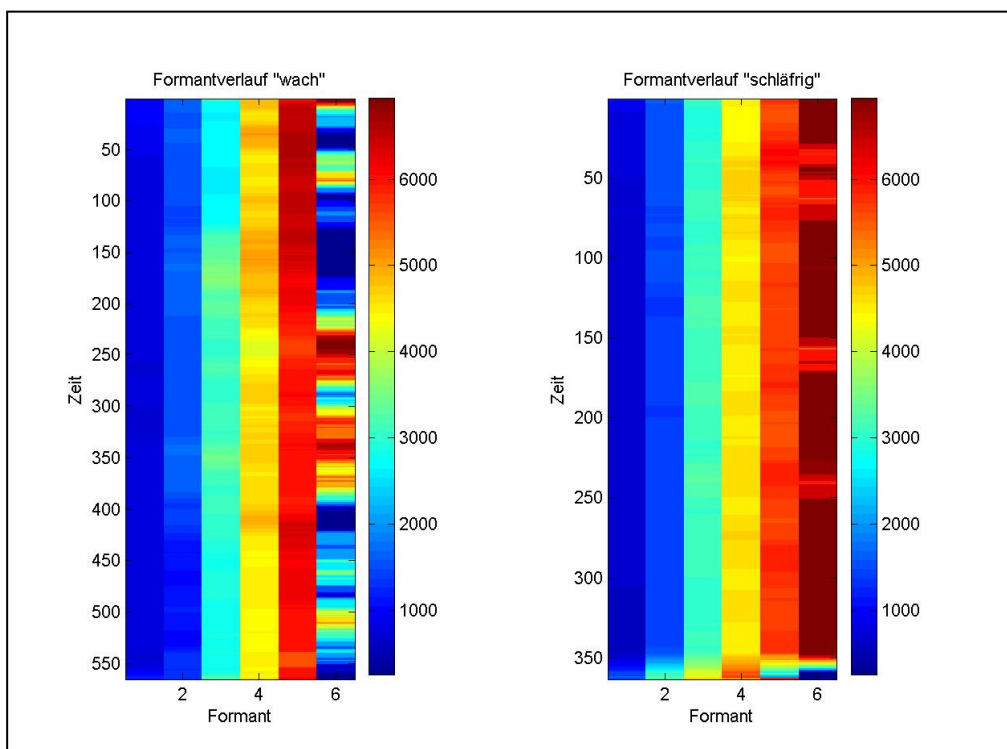


Abbildung 51: Formantverläufe der isolierten Vokalphonation [a:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 11)

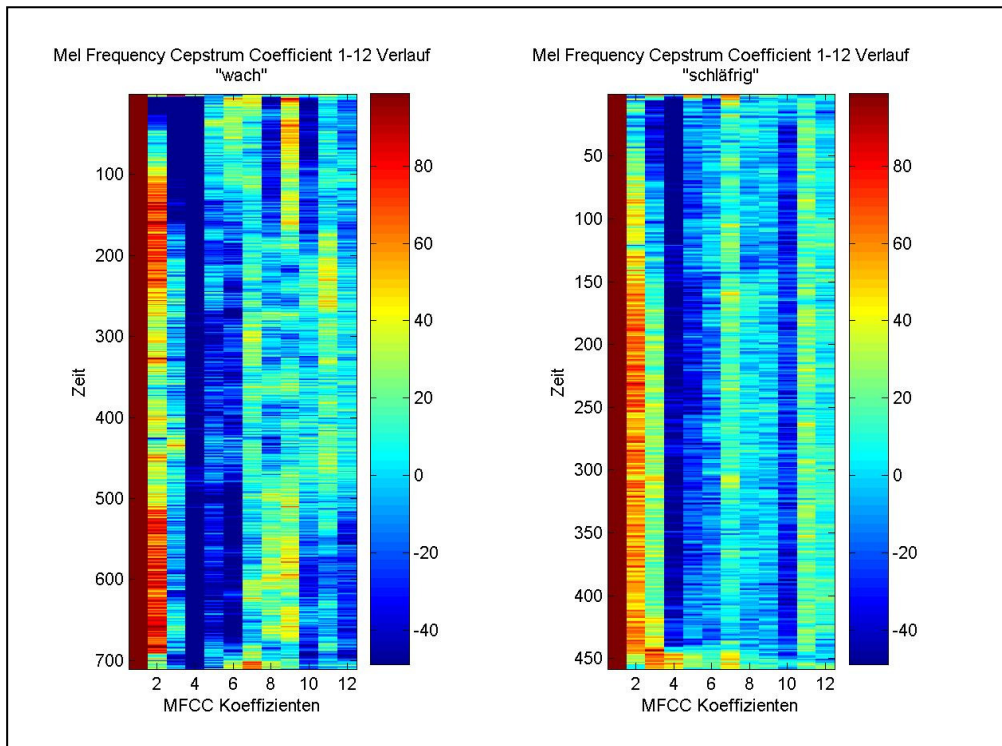


Abbildung 52: 12 MFCC Verläufe der isolierten Vokalphonation [a:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 12)

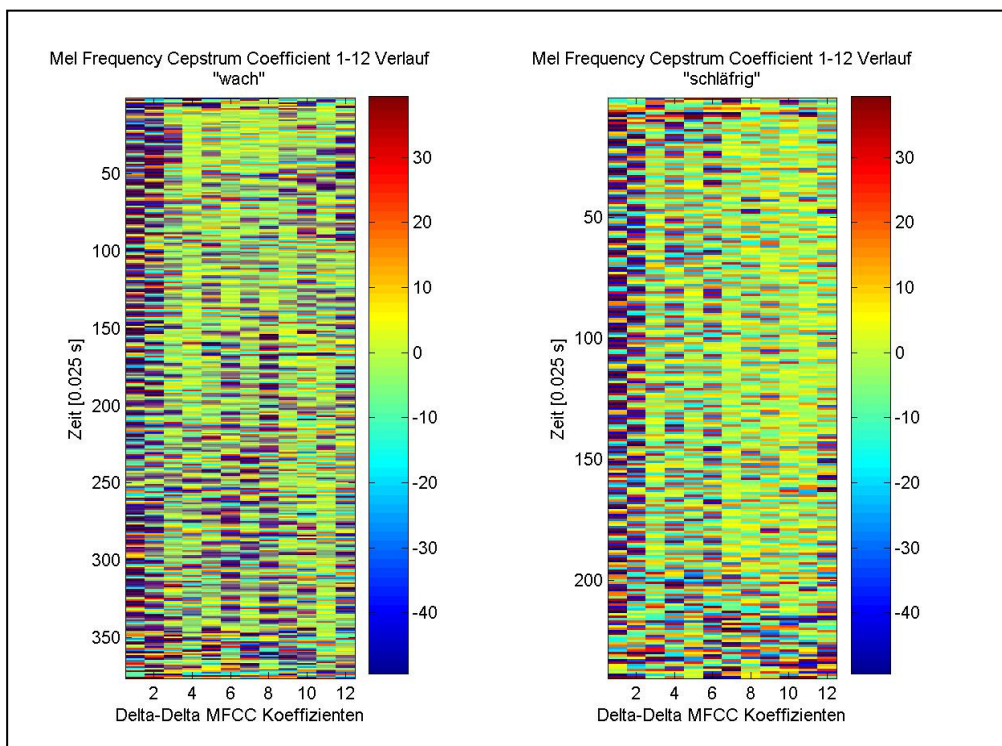


Abbildung 53: 12 Delta-Delta MFCC Verläufe der isolierten Vokalphonation [a:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 15)

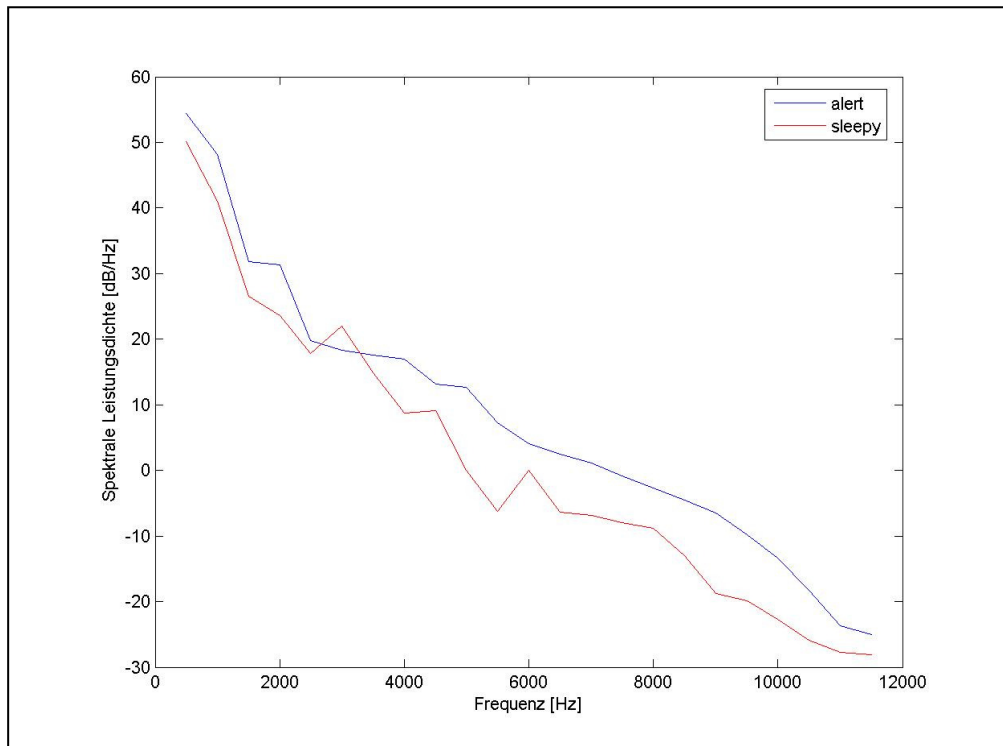


Abbildung 54: Long Term Average Speactrum der isolierten Vokalphonation [a:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 12)

Wie aus den Abbildungen 49, 50, 51, 52, 53 und 54 ersichtlich wird, unterscheiden sich wache von schläfrigen Sprachaufnahmen in ihrem Verlauf der Rohwelleform (z.B. höhere mittlere Amplitude), in ihrem Fundamental Frequenzverlauf (z.B. höherer Mittelwert, höheres 25stes Perzentil, höheres 75stes Perzentil), in ihren Formantpositionsverläufen (z.B. größere Standardabweichung in F3), in ihren MFCC Verläufen (z.B. tieferer MFCC9 Mittelwert), in ihren Delta-Delta MFCC Verläufen (z.B. höherer Delta-Delta MFCC1 Mittelwert) und in ihrem Long Term Average Spectrum Verlauf (z.B. höhere Band Energy zwischen 200 und 300 Hz).

2.2 Ergebnisse: Machine-Learning Klassifikation

In diesem Kapitel werden die Ergebnisse der auf dem beschriebenen Machine-Learning Prozedere basierenden Vorhersagen des Schläfrigkeitsniveaus präsentiert. Anders als im vorangegangenen Kapitel werden, hier ganze Kennzahlenbündel (hochdimensionale Merkmalsvektoren) für die Vorhersage genutzt (vgl. Abschnitt 4.5, Teil I). Der aus dem Fahrerassistenzsatz 6 („Am Nachmittag möchte ich nach Hamburg“) extrahierte Vokal [a:], der aus dem Fahrerassistenzsatz 9 („Rufen Sie den Notdienst an“) extrahierte Vokal [o:], sowie die isolierte 2-Sekunden Vokalphonation des gesprochenen [a:] dienen hier als Sprachmaterial.

Genutzt werden innerhalb der Machine-Learning Phase „Dimensionalitätsreduktion“ Korrelationsfilter-basierte Verfahren, PCA-basierte Dimensionalitätsreduktion (Extraktion von 100 Principle Components) sowie eine Wrapper-basierte Subset Selection mittels k-Nearest-Neighbour Verfahren mit einem Nachbarn (1-NN), Multilayer-Perceptron mit Feedforward-Netz und Backpropagation-Learning-Algorithm (MLP, 2 Hidden Layer, je 10 Neurone) und Support Vector Machine mit polynomialer Kernelfunktion (Genetic-Algorithm-basierte Suchstrategie). Die Darstellung der Vorhersageergebnisse für männliche und weibliche Sprecher beschränkt sich auf die Vorhersage von ungesehenen Testdaten mittels 10-facher Kreuzvalidierung. Die lediglich die Anpassungsfähigkeit des Klassifikators dokumentierenden Vorhersagegüten von Trainingsdaten werden im Weiteren nicht erwähnt.

Bestimmung der im Folgenden eingesetzten Klassifikators

In einer Vormessung soll zunächst bestimmt werden, welcher Klassifikator im weiteren Verlauf vorrangig zum Einsatz kommen soll. Wie Tabelle 19 dokumentiert, erzielt in einer Genetic-Algorithm-basierten Suchstrategie das k-Nearest-Neighbour Verfahren mit einem Nachbarn (1-NN) im Vergleich zum Multilayer-Perceptron mit Feedforward-Netz und Backpropagation-Learning-Algorithm (Input Layer Type: Linear; 2 Hidden Layer mit je 10 Neuronen; Hidden Layer Type: Sigmoid; Output Layer Type: Sigmoid; Training Cycles: 200; Learning Rate: 0.3; Momentum = 0.2; Error Epsilon = 0.05) und im Vergleich zu einer Support Vector Machine mit polynomialer Kernelfunktion (Gamma = 1/340; Epsilon = 0.001; Degree of Polynomial Function = 3; Cost = 0) die besten mittleren und die stabilsten Vorhersageergebnisse für ein repräsentatives Sprachmaterial („Fahrerassistenzsatz 9“). Die erfolgreiche Vorhersage des 1-NN lässt den Schluss zu, dass mit einiger Plausibilität der hier überlegene Klassifikator im Mittel auch in anderen Schläfrigkeits-Klassifikationsaufgaben zumindest durchschnittliche

Vorhersagegüten erzielt. Aus diesem Grund werden in den nachfolgenden Berechnungen lediglich 1-Nearest-Neighbour Klassifikatoren eingesetzt.

Tabelle 19: Vergleichende Gegenüberstellung der Ergebnisse der Klassifikationsverfahren k-Nearest-Neighbour, Multilayer Perceptron und Support Vector Machine (Sprachmaterial "Fahrerassistenzsatz 9")

Klassifikationsverfahren	Klassifikationsgenauigkeit
1-Nearest-Neighbour	77.32%+/- 11.18%
Multilayer Perceptron	70.71%+/-15.45%
Support Vector Machine	71.41% +/-21.87

Sprachmaterial "extrahierter Vokal [a:]": Klassifikationsergebnisse

Männliche Sprecher. Die Klassifikationsgenauigkeit der Schläfrigkeits-Vorhersage bei männlichen Sprecher beträgt 88.3% (SD 18.3%). Der positive prädiktive Wert liegt bei 83.3%, der negative prädiktive Wert bei 88.9%. Die Sensitivität des Verfahrens beträgt 71.4%, die Spezifität 94.1%. Die Korrelation zwischen den dichotomisierten KSS-Werten und der vorhergesagten Klassenzugehörigkeit liegt bei $r = .69$. Das Ergebnis des exakten Fisher-Tests bestätigt die Assoziation von vorhergesagten und selbstberichteten Schläfrigkeitszuständen ($p < .05$).

Tabelle 20: Klassifikationsergebnisse männlicher Sprecher des extrahierten Vokals [a:]. Angegeben sind die Klassifikationsgenauigkeit (Accuracy: Anzahl der richtig klassifizierten durch alle Klassifikationen), Sensitivität (Detektionsrate der Schläfrigkeit), Spezifität (1 minus False Alarm Rate) sowie der positiver und negative prädiktiver Wert (PV_pos: Wahrscheinlichkeit der tatsächlichen Schläfrigkeit, wenn Schläfrigkeit vorhergesagt wurde; PV_neg: Wahrscheinlichkeit der tatsächlichen Wachheit, wenn Wachheit vorhergesagt wurde)

	True a	True s	
Pred. a	16	2	PV_neg 88.9%
Pred. s	1	5	PV_pos 83.3%
	Spezifität 94.1%	Sensitivität 71.4%	Accuracy 88.3%

Weibliche Sprecher. Die Klassifikationsgenauigkeit der Schläfrigkeits-Vorhersage bei weiblichen Sprecher beträgt 83.5% ($SD = 12.3\%$). Der positive prädiktive Wert liegt bei 71.4%, der negative prädiktive Wert bei 88.2%. Die Sensitivität des Verfahrens beträgt 71.4%, die Spezifität 94.1%. Die Korrelation zwischen den dichotomisierten KSS-Werten und der vorhergesagten Klassenzugehörigkeit liegt bei $r = .60$. Das Ergebnis des exakten Fisher-Tests bestätigt die Assoziation von vorhergesagten und selbstberichteten Schläfrigkeitszuständen ($p < .05$).

***Tabelle 21:** Klassifikationsergebnisse weiblicher Sprecher des extrahierten Vokals [a:]. Angegeben sind die Klassifikationsgenauigkeit (Accuracy: Anzahl der richtig klassifizierten durch alle Klassifikationen), Sensitivität (Detektionsrate der Schläfrigkeit), Spezifität (1 minus False Alarm Rate) sowie der positiver und negative prädiktiver Wert (PV_pos: Wahrscheinlichkeit der tatsächlichen Schläfrigkeit, wenn Schläfrigkeit vorhergesagt wurde; PV_neg: Wahrscheinlichkeit der tatsächlichen Wachheit, wenn Wachheit vorhergesagt wurde)*

	True a	True s	
Pred. a	30	4	PV_neg 88.2%
Pred. s	4	10	PV_pos 71.4%
	Spezifität 88.2%	Sensitivität 71.4%	Accuracy 83.5%

Sprachmaterial “extrahierter Vokal [o:]“: Klassifikationsergebnisse

Männliche Sprecher. Die Klassifikationsgenauigkeit der Schläfrigkeits-Vorhersage bei männlichen Sprecher beträgt 81.7% ($SD = 18.9\%$). Der positive prädiktive Wert liegt bei 66.7%, der negative prädiktive Wert bei 83.3%. Die Sensitivität des Verfahrens beträgt 57.1%, die Spezifität 88.2%. Die Korrelation zwischen den dichotomisierten KSS-Werten und der vorhergesagten Klassenzugehörigkeit liegt bei $r = .48$. Das Ergebnis des exakten Fisher-Tests bestätigt die Assoziation von vorhergesagten und selbstberichteten Schläfrigkeitszuständen ($p < .05$).

Table 22: Klassifikationsergebnisse männlicher Sprecher des extrahierten Vokals [o:]. Angegeben sind die Klassifikationsgenauigkeit (Accuracy: Anzahl der richtig klassifizierten durch alle Klassifikationen), Sensitivität (Detektionsrate der Schläfrigkeit), Spezifität (1 minus False Alarm Rate) sowie der positiver und negative prädiktiver Wert (PV_pos: Wahrscheinlichkeit der tatsächlichen Schläfrigkeit, wenn Schläfrigkeit vorhergesagt wurde; PV_neg: Wahrscheinlichkeit der tatsächlichen Wachheit, wenn Wachheit vorhergesagt wurde)

	True a	True s	
Pred. a	15	3	PV_neg 83.3%
Pred. s	2	4	PV_pos 66.7%
	Spezifität 88.2%	Sensitivität 57.1%	Accuracy 81.7%

Weibliche Sprecher. Die Klassifikationsgenauigkeit der Schläfrigkeits-Vorhersage bei weiblichen Sprecher beträgt 77.5% ($SD = 14.0\%$). Der positive prädiktive Wert liegt bei 60.0%, der negative prädiktive Wert bei 84.9%. Die Sensitivität des Verfahrens beträgt 64.3%, die Spezifität 82.4%. Die Korrelation zwischen den dichotomisierten KSS-Werten und der vorhergesagten Klassenzugehörigkeit liegt bei $r = .48$. Das Ergebnis des exakten Fisher-Tests bestätigt die Assoziation von vorhergesagten und selbstberichteten Schläfrigkeitszuständen ($p < .01$).

Tabelle 23: Klassifikationsergebnisse weiblicher Sprecher des extrahierten Vokals [o:]. Angegeben sind die Klassifikationsgenauigkeit (Accuracy: Anzahl der richtig klassifizierten durch alle Klassifikationen), Sensitivität (Detektionsrate der Schläfrigkeit), Spezifität (1 minus False Alarm Rate) sowie der positiver und negative prädiktiver Wert (PV_pos: Wahrscheinlichkeit der tatsächlichen Schläfrigkeit, wenn Schläfrigkeit vorhergesagt wurde; PV_neg: Wahrscheinlichkeit der tatsächlichen Wachheit, wenn Wachheit vorhergesagt wurde)

	True a	True s	
Pred. a	28	5	PV_neg 84.8%
Pred. s	6	9	PV_pos 60.0%
	Spezifität 82.4%	Sensitivität 64.3%	Accuracy 77.5%

Sprachmaterial “isolierte Vokalphonation [a:]“: Klassifikationsergebnisse

Männliche Sprecher. Die Klassifikationsgenauigkeit der Schläfrigkeits-Vorhersage bei männlichen Sprecher beträgt 91.7% ($SD = 12.9\%$). Der positive prädiktive Wert liegt bei 87.5%, der negative prädiktive Wert bei 92.8%. Die Sensitivität des Verfahrens beträgt 77.8%, die Spezifität 96.3%. Die Korrelation zwischen den dichotomisierten KSS-Werten und der vorhergesagten Klassenzugehörigkeit liegt bei $r = .77$. Das Ergebnis des exakten Fisher-Tests bestätigt die Assoziation von vorhergesagten und selbstberichteten Schläfrigkeitszuständen ($p < .01$).

***Tabelle 24:** Klassifikationsergebnisse männlicher Sprecher der isolierten Vokalphonation [a:]. Angegeben sind die Klassifikationsgenauigkeit (Accuracy: Anzahl der richtig klassifizierten durch alle Klassifikationen), Sensitivität (Detektionsrate der Schläfrigkeit), Spezifität (1 minus False Alarm Rate) sowie der positiver und negative prädiktiver Wert (PV_pos: Wahrscheinlichkeit der tatsächlichen Schläfrigkeit, wenn Schläfrigkeit vorhergesagt wurde; PV_neg: Wahrscheinlichkeit der tatsächlichen Wachheit, wenn Wachheit vorhergesagt wurde)*

	True a	True s	
Pred. a	26	2	PV_neg 92.9%
Pred. s	1	7	PV_pos 87.5%
	Spezifität 96.3%	Sensitivität 77.8%	Accuracy 91.7%

Weibliche Sprecher. Die Klassifikationsgenauigkeit der Schläfrigkeits-Vorhersage bei weiblichen Sprecher beträgt 73.0% ($SD = 18.9\%$). Der positive prädiktive Wert liegt bei 53.8%, der negative prädiktive Wert bei 79.1%. Die Sensitivität des Verfahrens beträgt 43.8%, die Spezifität 85.0%. Die Korrelation zwischen den dichotomisierten KSS-Werten und der vorhergesagten Klassenzugehörigkeit liegt bei $r = .31$. Das Ergebnis des exakten Fisher-Tests bestätigt die Assoziation von vorhergesagten und selbstberichteten Schläfrigkeitszuständen ($p < .01$).

Tabelle 25: Klassifikationsergebnisse weiblicher Sprecher der isolierten Vokalphonation [a:]. Angegeben sind die Klassifikationsgenauigkeit (Accuracy: Anzahl der richtig klassifizierten durch alle Klassifikationen), Sensitivität (Detektionsrate der Schläfrigkeit), Spezifität (1 minus False Alarm Rate) sowie der positiver und negative prädiktiver Wert (PV_pos: Wahrscheinlichkeit der tatsächlichen Schläfrigkeit, wenn Schläfrigkeit vorhergesagt wurde; PV_neg: Wahrscheinlichkeit der tatsächlichen Wachheit, wenn Wachheit vorhergesagt wurde)

	True a	True s	
Pred. a	34	9	PV_neg 79.1%
Pred. s	6	7	PV_pos 53.9%
	Spezifität 85.0%	Sensitivität 43.8%	Accuracy 73.0%

3 ERGEBNISSE: SPRACHSEGMENT WORT

Nachdem im vorangegangenen Kapitel die Ergebnisse des Sprachsegments Vokal beschrieben worden sind, beginnt dieses Kapitel mit einem Kurzüberblick zu den Hauptergebnissen des Kapitels 3. Im Anschluss daran werden korrelative Zusammenhänge zwischen sprachlichen Einzelmerkmalen des Wortes „Suchen“ und Schläfrigkeitratings vorgestellt. Das nachfolgende Unterkapitel stellt die für die Validitätseinschätzung zentralen, multivariaten Machine-Learning Klassifikationsergebnisse dar.

Ergebniszusammenfassung Sprachsegment Wort

Sprachsegment Wort „suchen“. Die linearen Zusammenhänge von 32 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeitsszuständen erreichen eine Größenordnung von $r > .2$. Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt eine Vorhersagegenauigkeit von 80.3% (Sensitivität = 45.0%, Spezifität = 92.7%).

3.1 Ergebnisse Einzelkennzahlen

In diesem Kapitel werden Zusammenhänge zwischen sprachlichen Einzelkennzahlen eines Wortsegments und dem selbstberichteten Schläfrigkeitniveau (KSS) präsentiert. Die Ergebnisse des aus dem Fahrerassistenzsatz 4 („Suchen Sie bitte die nächste Raststätte heraus“) extrahierten Wortes „Suchen“ dienen hier als Sprachmaterial. Zur weiteren Illustration der gefundenen korrelativen Analyseergebnisse werden jeweils für einzelne wache und schläfrige Stimmaufnahmen Kennwertverläufe dargestellt.

Sprachmaterial „suchen“: korrelative Einzelergebnisse

Korrelative Zusammenhänge zwischen Stimmkennzahlen des extrahierten Wortes „suchen“ und Schläfrigkeitratings. 32 Einzelkorrelationen zwischen Stimmkennzahlen des extrahierten Wortes „suchen“ und selbstberichteten Schläfrigkeitratings erreichen eine mindestens kleine Effektstärke von $r > .2$. Die betragsmäßig höchsten linearen Zusammenhänge werden für die folgenden Merkmale erzielt:

f3_bandw_q75_z: .42 (z-transformiertes 75tes Perzentil der Bandwidth der dritten Formantfrequenz); df2f4_z: .37 (z-transformierte Differenz der Mittelwerte der Verläufe der zweiten und vierten Formantfrequenz); df2f5_z: .33 (z-transformierte Differenz der Mittelwerte der Verläufe der zweiten und fünften Formantfrequenz); MFCC4_z: .35 (z-transformierter Mittelwert des Verlaufs des Mel-Frequency Cepstrum Coefficient 4); MFCC6_z: -.34 (z-transformierter Mittelwert des Verlaufs des Mel-Frequency Cepstrum Coefficient 6).

Tabelle 26: Korrelative Zusammenhänge ($r > .20$) zwischen Sprachmerkmalen des extrahierten Wortes „suchen“ und KSS-Werten.

f0_mintime_rel	f3_min	f3_range	df1f2	delta-delta-MFCC8	Delta MFCC 10	Delta MFCC 11	Delta MFCC 12	f0_timemax_z	f2_mean_z
-.27	.27	-.31	-.26	.32	-.27	-.28	-.26	.25	.27
f2_sd_z	f3_maxtime_rel_z	f3_range_z	f3_mean_z	f3_sd_z	f3_bandw_q75_z	f5_mean_z	f5_bandw_q75_z	df1f2_z	df2f4_z
-.26	-.28	-.31	.26	-.29	-.27	-.28	.42	-.27	.37
df2f5_z	df3f4_z	df3f5_z	MFCC2_z	MFCC4_z	MFCC6_z	Delta-delta-MFCC8_z	Delta MFCC 10_z	MFCC11_z	glottOpen_GOG_z
.33	.29	.33	-.30	.35	-.34	.29	-.25	.33	-.25
baen200_300_z	ltas_locpeak_z								
-.31	.27								

Im Weiteren werden die aus den korrelativen Analysen gewonnenen Hinweise auf schläfrigkeitssensitive sprachliche Einzelmerkmale ähnlich wie für das Sprachmaterial zuvor exemplarisch dargestellt.

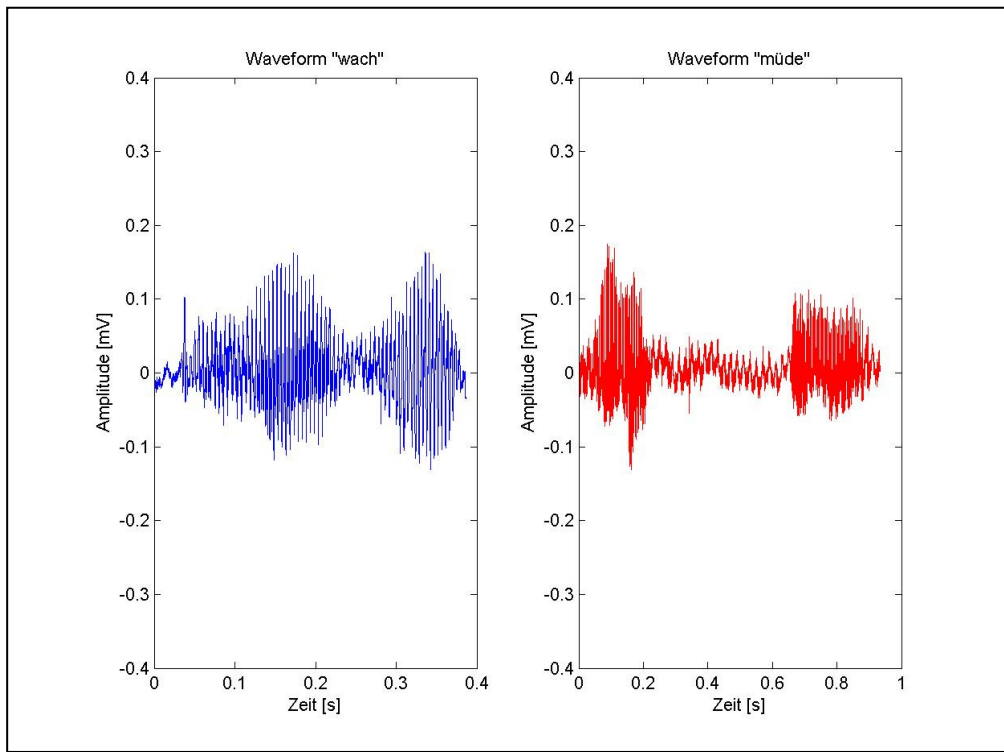


Abbildung 55: Rohwaveform des extrahierten Wortes „suchen“ für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 16)

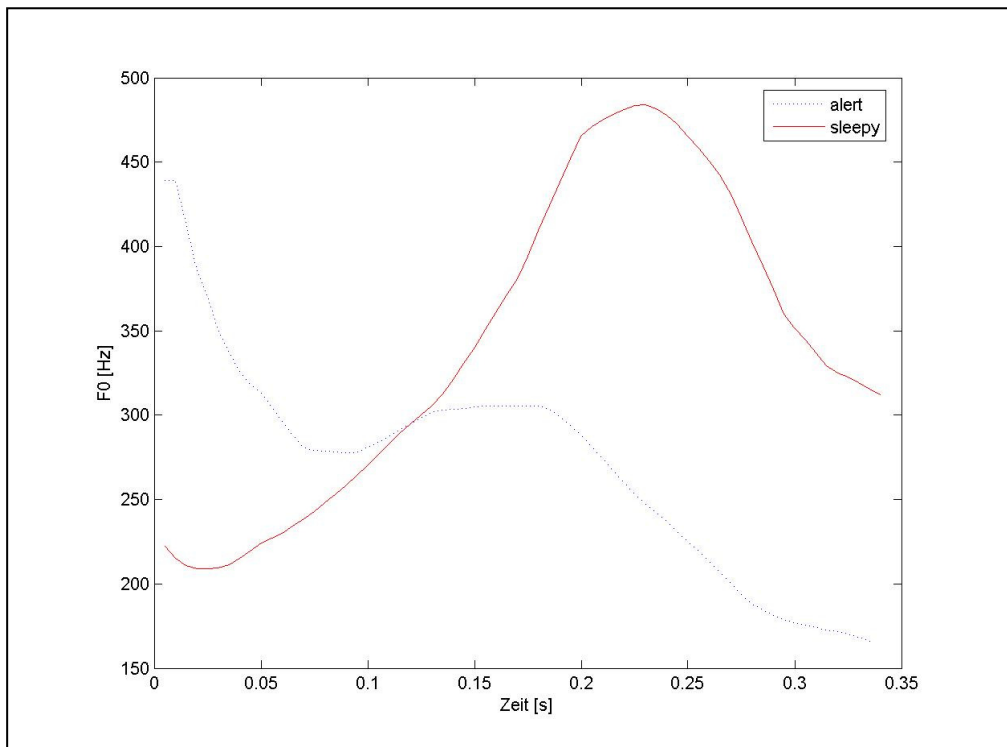


Abbildung 56: Fundamental Frequenzverlauf des extrahierten Wortes „suchen“ für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 15)

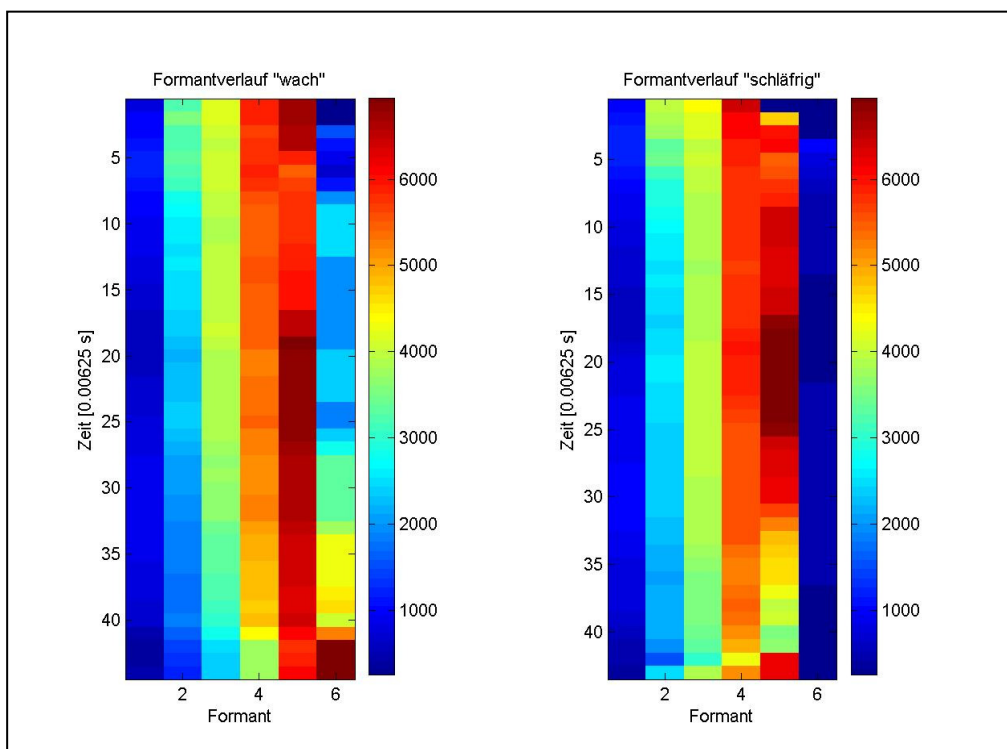


Abbildung 57: Formantverläufe des extrahierten Wortes „suchen“ für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 17)

Wie aus den Abbildungen 55, 56 und 57 ersichtlich wird, unterscheiden sich wache von schläfrigen Sprachaufnahmen in ihrem Verlauf der Rohwaveform (z.B. höhere mittlere Amplitude), in ihrem Fundamental Frequenzverlauf (z.B. spätere zeitliche Position des Minimums) und in ihren Formantpositionsverläufen (z.B. höherer Mittelwert F5).

3.2 Ergebnisse: Machine-Learning Klassifikation

In diesem Kapitel werden die Ergebnisse der auf dem beschriebenen Machine-Learning Prozedere basierenden Vorhersagen des Schläfrigkeitsniveaus präsentiert. Anders als im vorangegangenen Unterkapitel werden hier ganze Kennzahlenbündel (hochdimensionale Merkmalsvektoren) für die Vorhersage genutzt (vgl. Abschnitt 4.5, Teil I). Die Vorhersageergebnisse basieren auf dem aus dem Fahrerassistenzsatz 4 extrahierten Wort „suchen“.

Genutzt werden innerhalb der Machine-Learning Phase „Dimensionalitätsreduktion“ Korrelationsfilter-basierte Verfahren, PCA-basierte Dimensionalitätsreduktion (Extraktion von 100 Principle Components) sowie eine Wrapper-basierte Subset Selection mittels k-Nearest-Neighbour Verfahren mit einem Nachbarn (1-NN) (Genetic-Algorithm-basierte Suchstrategie). Die Darstellung der Vorhersageergebnisse für männliche und weibliche Sprecher beschränkt sich auf die Vorhersage von ungesehenen Testdaten mittels 10-facher Kreuzvalidierung. Die lediglich die Anpassungsfähigkeit des Klassifikators dokumentierenden Vorhersagewerten von Trainingsdaten werden im Weiteren nicht erwähnt.

Sprachmaterial “Fahrerassistenzsatz 4 – Wort 1“ (“suchen“): Klassifikationsergebnisse

Männliche Sprecher. Die Klassifikationsgenauigkeit der Schläfrigkeits-Vorhersage bei männlichen Sprecher beträgt 85.0% (SD 22.9%). Der positive prädiktive Wert liegt bei 100%, der negative prädiktive Wert bei 83.3%. Die Sensitivität des Verfahrens beträgt 40.0%, die Spezifität 100%. Die Korrelation zwischen den dichotomisierten KSS-Werten und der vorhergesagten Klassenzugehörigkeit liegt bei $r = .58$. Das Ergebnis des exakten Fisher-Tests bestätigt die Assoziation von vorhergesagten und selbstberichteten Schläfrigkeitszuständen ($p < .05$).

Tabelle 27: Klassifikationsergebnisse männlicher Sprecher des Fahrerassistenzsatz 4 – Wort 1“ (“suchen“). Angegeben sind die Klassifikationsgenauigkeit (Accuracy: Anzahl der richtig klassifizierten durch alle Klassifikationen), Sensitivität (Detektionsrate der Schläfrigkeit), Spezifität (1 minus False Alarm Rate) sowie der positiver und negative prädiktiver Wert (PV_pos: Wahrscheinlichkeit der tatsächlichen Schläfrigkeit, wenn Schläfrigkeit vorhergesagt wurde; PV_neg: Wahrscheinlichkeit der tatsächlichen Wachheit, wenn Wachheit vorhergesagt wurde)

	True a	True s	
Pred. a	15	3	PV_neg 83.3%
Pred. s	0	2	PV_pos 100%
	Spezifität 100%	Sensitivität 40.0%	Accuracy 85.0%

Weibliche Sprecher. Die Klassifikationsgenauigkeit der Schläfrigkeits-Vorhersage bei weiblichen Sprecher beträgt 75.5% (SD = 17.4%). Der positive prädiktive Wert liegt bei 58.3%, der negative prädiktive Wert bei 80.6%. Die Sensitivität des Verfahrens beträgt 50.0%, die Spezifität 85.3%. Die Korrelation zwischen den dichotomisierten KSS-Werten und der vorhergesagten Klassenzugehörigkeit liegt bei $r = .37$. Das Ergebnis des exakten Fisher-Tests bestätigt die Assoziation von vorhergesagten und selbstberichteten Schläfrigkeitszuständen ($p < .01$).

Tabelle 28: Klassifikationsergebnisse weiblicher Sprecher des Fahrerassistenzsatz 4 – Wort 1“ (“suchen“). Angegeben sind die Klassifikationsgenauigkeit (Accuracy: Anzahl der richtig klassifizierten durch alle Klassifikationen), Sensitivität (Detektionsrate der Schläfrigkeit), Spezifität (1 minus False Alarm Rate) sowie der positiver und negative prädiktiver Wert (PV_pos: Wahrscheinlichkeit der tatsächlichen Schläfrigkeit, wenn Schläfrigkeit vorhergesagt wurde; PV_neg: Wahrscheinlichkeit der tatsächlichen Wachheit, wenn Wachheit vorhergesagt wurde)

	True a	True s	
Pred. a	29	7	PV_neg 80.6%
Pred. s	5	7	PV_pos 58.3%
	Spezifität 85.3%	Sensitivität 50.0%	Accuracy 75.5%

4 ERGEBNISSE: SPRACHSEGMENT PHRASALE EINHEIT

Nachdem im vorangegangenen Kapitel die Ergebnisse des Sprachsegments Wort beschrieben worden sind, beginnt dieses Kapitel mit einem Kurzüberblick zu den Hauptergebnissen des Kapitels 4. Im Anschluss daran werden korrelative Zusammenhänge zwischen sprachlichen Einzelmerkmalen dreier Phrasaler Einheiten und Schläfrigkeitratings vorgestellt. Das nachfolgende Unterkapitel stellt die für die Validitätseinschätzung zentralen, multivariaten Machine-Learning Klassifikationsergebnisse dar.

Ergebniszusammenfassung Sprachsegment Phrasale Einheit

Sprachsegment Phrasale Einheit „Satz 3“. Die linearen Zusammenhänge von 35 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeitsszuständen erreichen eine mindestens kleine Effektstärke von $r > .2$. Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt eine Vorhersagegenauigkeit von 83.3% (Sensitivität = 75.0%, Spezifität = 86.8%).

Sprachsegment Phrasale Einheit „Satz 9“. Die linearen Zusammenhänge von 32 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeitsszuständen erreichen eine mindestens kleine Effektstärke von $r > .2$. Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt eine Vorhersagegenauigkeit von 82.3% (Sensitivität = 60.7%, Spezifität = 90.7%).

Sprachsegment Phrasale Einheit „Intro-Phrase“. Die linearen Zusammenhänge von 14 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeitsszuständen erreichen eine mindestens kleine Effektstärke von $r > .2$. Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt eine Vorhersagegenauigkeit von 84.5% (Sensitivität = 68.4%, Spezifität = 84.1%).

4.1 Ergebnisse: Einzelkennzahlen

In diesem Kapitel werden Zusammenhänge zwischen sprachlichen Einzelkennzahlen dreier Phrasaler Einheiten und dem selbstberichteten Schläfrigkeitniveau (KSS) präsentiert. Die Ergebnisse des Fahrerassistenzsatzes 3 („Das ist falsch, es muss ein Irrtum sein“), des Fahrerassistenzsatzes 9 („Rufen Sie den Notdienst an“) sowie der Probanden Vorstellungsintröphrase („19.50 Uhr, Vp 10, Sleepiness Scale Wert 1“) dienen hier als Sprachmaterial. Zur weiteren Illustration der gefundenen korrelativen Analyseergebnisse werden jeweils für einzelne wache und schläfrige Stimmaufnahmen Kennwertverläufe dargestellt.

Sprachmaterial „Fahrerassistenzsatz 3“: korrelative Einzelergebnisse

Korrelative Zusammenhänge zwischen Stimmkennzahlen des Fahrerassistenzsatzes 3 und Schläfrigkeitratings. 35 Einzelkorrelationen zwischen Stimmkennzahlen des Fahrerassistenzsatzes 3 und selbstberichteten Schläfrigkeitratings erreichen eine mindestens kleine Effektstärke von $r > .2$. Die höchsten linearen Zusammenhänge werden für die folgenden Merkmale erzielt:

hammerberg_z: .34 (Hammarberg Index); spec_cog_z: -.28 (z-transformierter Schwerpunkt des Amplitudenspektrums); spec_kurt_z: .34 (z-transformierter Kurtosis des Amplitudenspektrums); spec_skew_z: .45 (z-transformierter Schiefe des Amplitudenspektrums); spec_sd_z: -.36 (z-transformierter Standardabweichung des Amplitudenspektrums).

Tabelle 29: Korrelative Zusammenhänge ($r > .20$) zwischen Sprachmerkmalen des Fahrerassistenzsatzes 3 („Das ist falsch, es muss ein Irrtum sein“) und KSS-Werten.

f1_mean	f2_range	f3_mintime_rel	f3_sd	f4_mean	f5_maxtime_rel	f5_bandw_q75	df1f2
-.20	-.21	-.20	-.24	-.20	.22	.20	-.23
hammarberg_s3	spec_kurt	spec_skew	spec_sd	Delta-MFCC1	Delta-MFCC9		
.27	.21	.23	-.20	.21	.25		
f3_mintime_rel_z	f5_bandw_q75_z	imcomplclos_IC_z	energmax2k_5k_z	hammarberg_z	hammarberg_s0_z	MFCC2_z	f0_slope_octjump_z
-.20	.24	.27	-.30	.34	-.21	.20	-.28
f0_max_z	f0_mean_z	f0_range_z	int_range_z	int_max_z	int_mean_z	int_sd_z	int_75quant_z
-.22	-.22	-.22	-.21	-.20	-.22	-.21	-.26
spec_cog_z	spec_cm_z	spec_kurt_z	spec_skew_z	spec_sd_z			
-.28	-.31	.43	.45	-.36			

Im Weiteren werden die aus den korrelativen Analysen gewonnenen Hinweise auf schläfrigkeitssensitive sprachliche Einzelmerkmale ähnlich wie für das Sprachmaterial zuvor exemplarisch dargestellt.

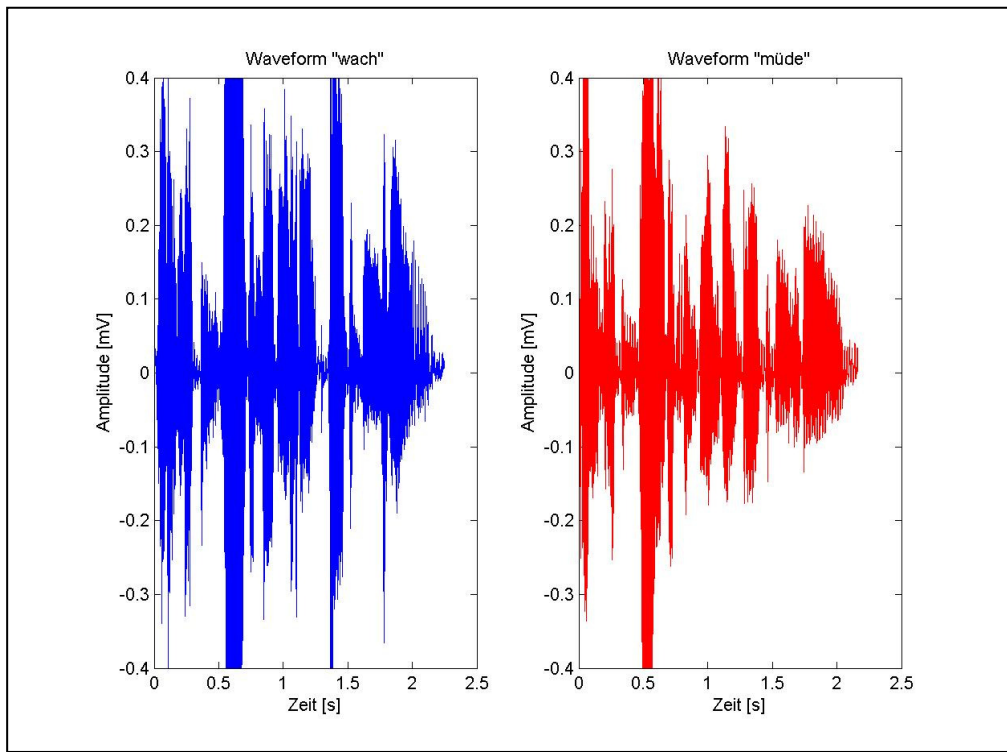


Abbildung 58: Rohwaveform des Fahrerassistenzsatzes 3 für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 14)

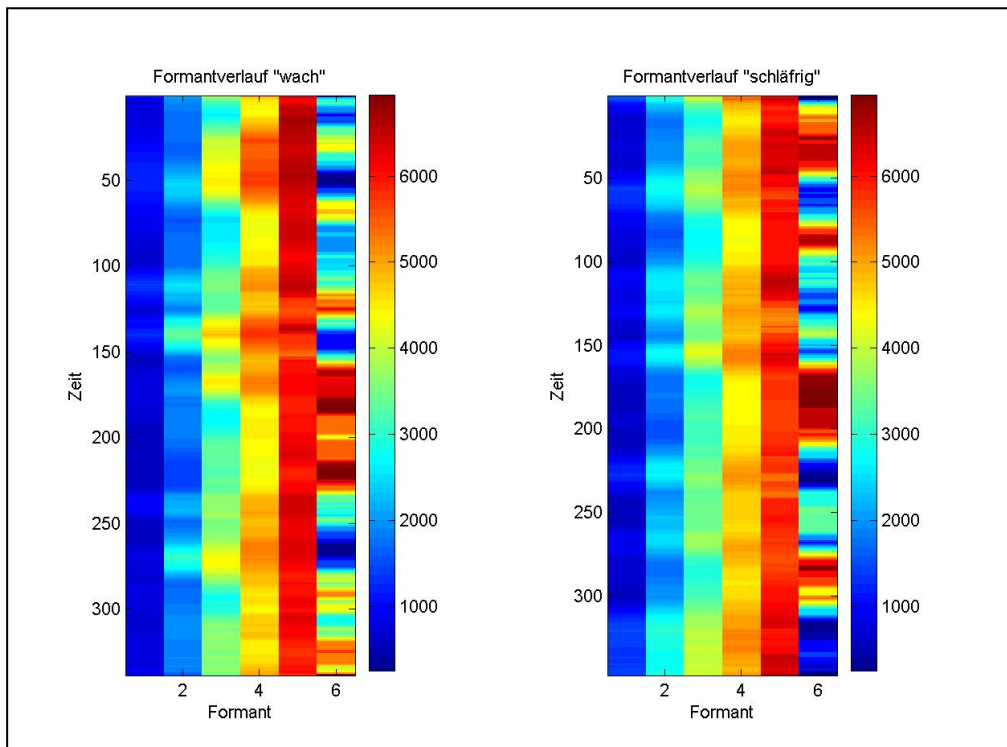


Abbildung 59: Formantverläufe des Fahrerassistenzsatzes 3 für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 23)

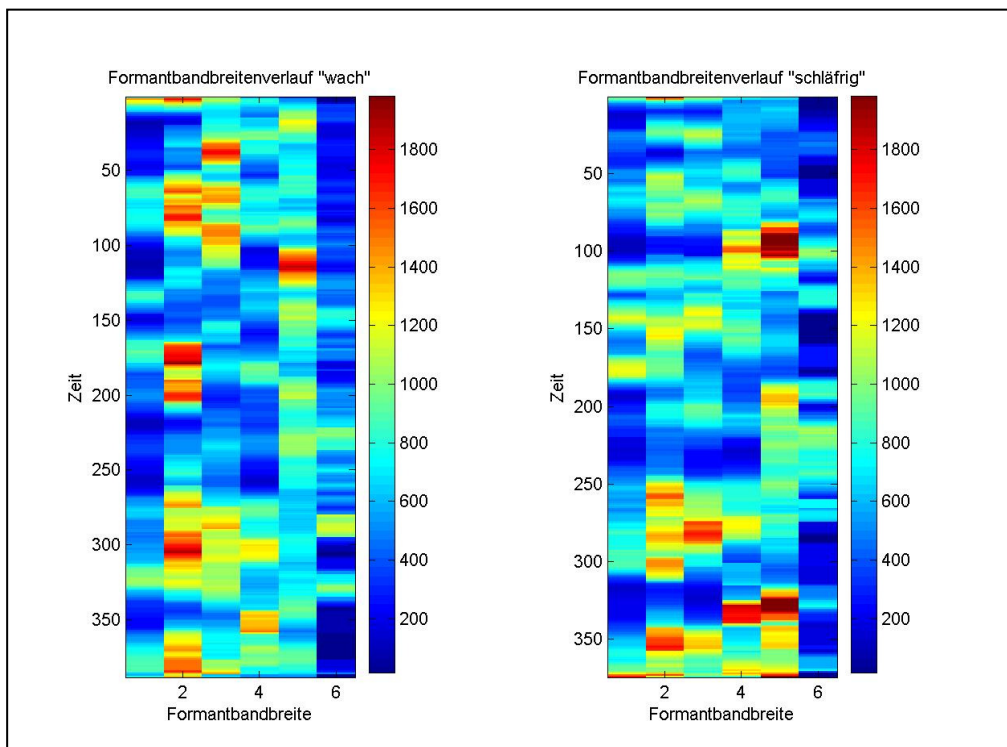


Abbildung 60: Formantbandbreitenverläufe (Formantbandwidth) des Fahrerassistenzsatzes 3 für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 10)

Wie aus den Abbildungen 58, 59 und 60 ersichtlich wird, unterscheiden sich wache von schläfrigen Sprachaufnahmen in ihrem Verlauf der Rohwaveform (z.B. höhere mittlere Amplitude), in ihren Formantpositionsverläufen (z.B. höhere Standardabweichung F3) und in ihren Formantbandbreitenverläufen (z.B. niedrigeres 75stes Perzentil Bandwidth F5).

Sprachmaterial „Fahrerassistenzsatz 9“: korrelative Einzelergebnisse

Korrelative Zusammenhänge zwischen Stimmkennzahlen des Fahrerassistenzsatzes 9 und Schläfrigkeitsratings. 32 Einzelkorrelationen zwischen Stimmkennzahlen des Fahrerassistenzsatzes 9 und selbstberichteten Schläfrigkeitsratings erreichen eine mindestens kleine Effektstärke von $r > .2$. Die fünf betragsmäßig höchsten linearen Zusammenhänge werden für die folgenden Merkmale erzielt:

f5_bandw_q75_z: .42 (z-transformiertes 75tes Perzentil der Bandwidth der dritten Formantfrequenz); df2f4_z: .37 (z-transformierte Differenz der Mittelwerte der Verläufe der zweiten und vierten Formantfrequenz); df2f5_z: .33 (z-transformierte Differenz der Mittelwerte der Verläufe der zweiten und fünften Formantfrequenz); MFCC4_z: .35 (z-transformierter Mittelwert des Verlaufs des Mel-Frequency Cepstrum Coefficient 4); MFCC6_z: -.34 (z-transformierter Mittelwert des Verlaufs des Mel-Frequency Cepstrum Coefficient 6).

Tabelle 30: Korrelative Zusammenhänge ($r > .20$) zwischen Sprachmerkmalen des Fahrerassistenzsatzes 9 („Rufen Sie den Notdienst“) und KSS-Werten.

f0_mintime_rel	f3_min	f3_range	df1f2	delta-delta-MFCC8	delta10	delta11	delta12	f0_timemax_z	f2_mean_z
-.27	.27	-.31	-.26	.32	-.29	-.29	-.26	.25	.27
f2_sd_z	f3_maxtime_rel_z	f3_range_z	f3_mean_z	f3_sd_z	f3_bandw_q75_z	f5_mean_z	f5_bandw_q75_z	df1f2_z	df2f4_z
-.32	-.28	-.31	.26	-.33	-.27	-.28	.42	-.27	.37
df2f5_z	df3f4_z	df3f5_z	MFCC2_z	MFCC4_z	MFCC6_z	delta-delta-MFCC8_z	delta10_z	MFCC11_z	glottOpen_GOG_z
.33	.29	.33	-.30	.35	-.34	.29	-.25	.33	-.25
baen200_300_z	Itas_locpeak_z								
-.31	.27								

Im Weiteren werden die aus den korrelativen Analysen gewonnenen Hinweise auf schläfrigkeitssensitive sprachliche Einzelmerkmale ähnlich wie für das Sprachmaterial zuvor exemplarisch dargestellt.

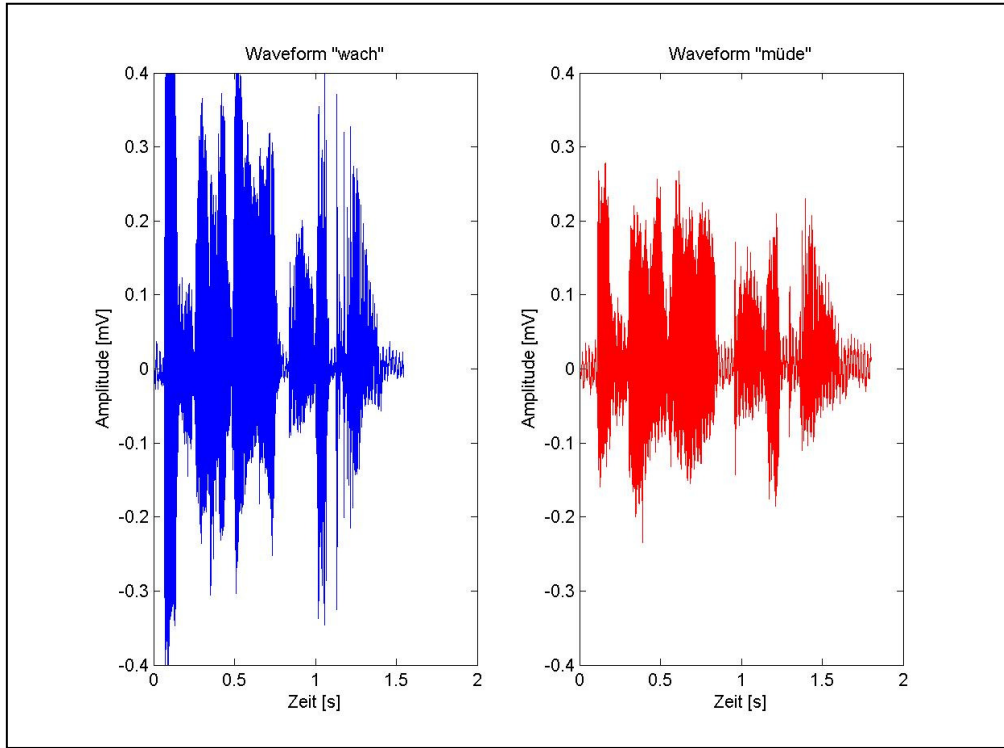


Abbildung 61: Rohwaveform des Fahrerassistenzsatzes 9 für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 14)

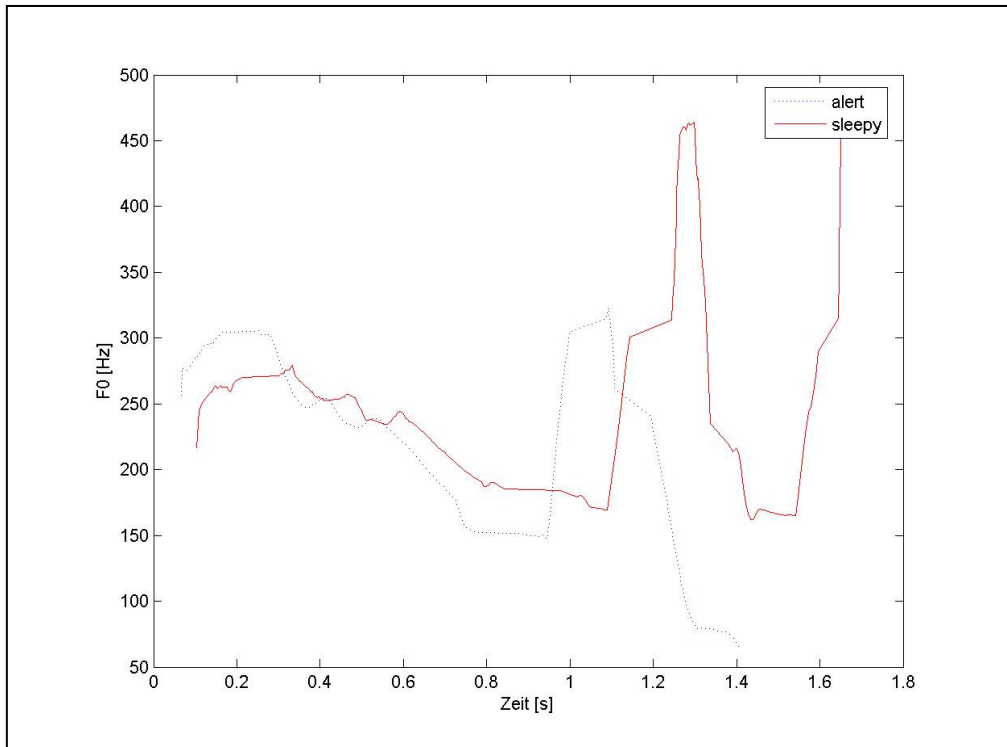


Abbildung 62: Fundamental Frequenzverlauf des Fahrerassistenzsatzes 9 für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 14)

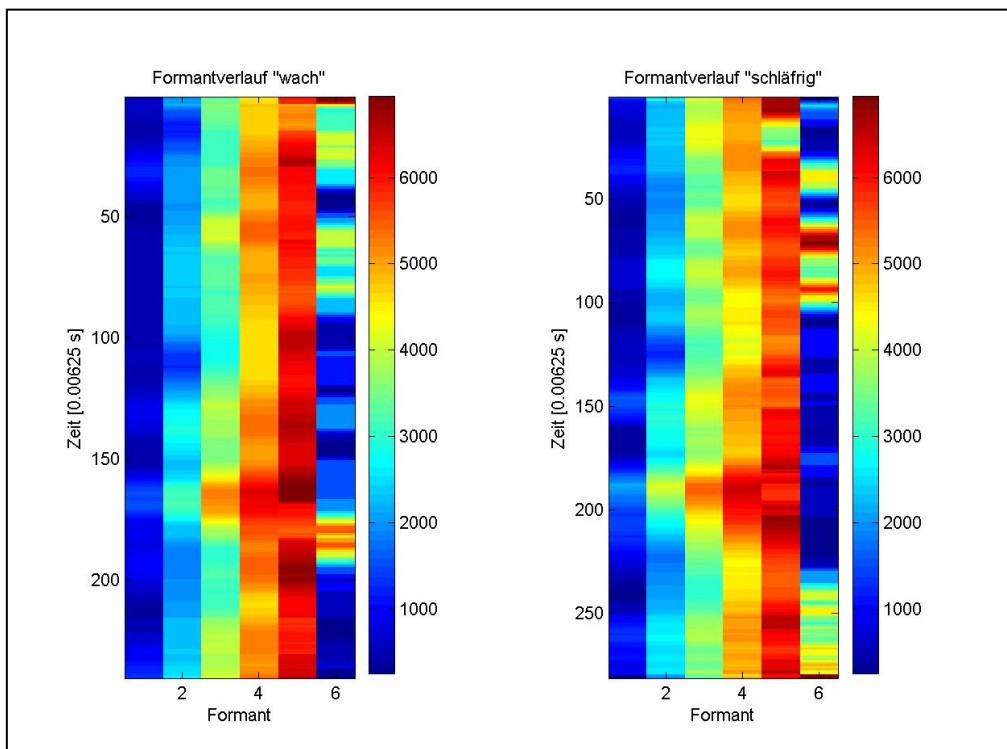


Abbildung 63: Formantverläufe des Fahrerassistenzsatzes 9 für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 10)

Wie aus den Abbildungen 61, 62 und 63 ersichtlich wird, unterscheiden sich wache von schläfrigen Sprachaufnahmen in ihrem Verlauf der Rohwaveform (z.B. höhere mittlere Amplitude), in ihrem Fundamental Frequenz Verlauf (z.B. spätere Position des Maximums) und in ihren Formantpositionsverläufen (z.B. höherer Mittelwert F5).

Sprachmaterial „Introphrase“: korrelative Einzelergebnisse

Korrelative Zusammenhänge zwischen Stimmkennzahlen der Introphrase und Schläfrigkeitsratings. 14 Einzelkorrelationen zwischen Stimmkennzahlen der Introphrase und selbstberichteten Schläfrigkeitsratings erreichen eine mindestens kleine Effektstärke von $r > .2$. Die höchsten 5 linearen Zusammenhänge werden für die folgenden Merkmale erzielt:

f3_sd: -.28 (Standardabweichung des Verlaufs der dritten Formantenfrequenz); f3_max_z: -.26 (Maximum des Verlaufs der dritten Formantenfrequenz); f4_max: -.27 (Maximum des Verlaufs der vierten Formantenfrequenz); spec_sd_z: -.26 (z-transformierte Standardabweichung des Amplitudenspektrums); hammarberg_s4_z: .33 (z-transformierter Hammarberg Index 4).

Tabelle 31: Korrelative Zusammenhänge ($r > .20$) zwischen Sprachmerkmalen der Introphrase (z.B. „19.50 Uhr, Vp 10, Sleepiness Scale Wert 2“) und KSS-Werten.

f3_sd	f1_bandw_q25_z	f1_bandw_q50_z	f1_maxtime_z	f2_maxtime_rel_z	f2_bandw_q50	f3_bandw_q50	f3_maxzwert
-.28	-.24	-.21	.25	.20	-.22	-.23	-.26
f4_max	f5_mintime_z	MFCC2	MFCC6_z	spec_sd_z	hammarberg_s4_z		
-.27	-.22	-.26	-.26	-.23	.33		

Im Weiteren werden die aus den korrelativen Analysen gewonnenen Hinweise auf schläfrigkeitssensitive sprachliche Einzelmerkmale ähnlich wie für das Sprachmaterial zuvor exemplarisch dargestellt.

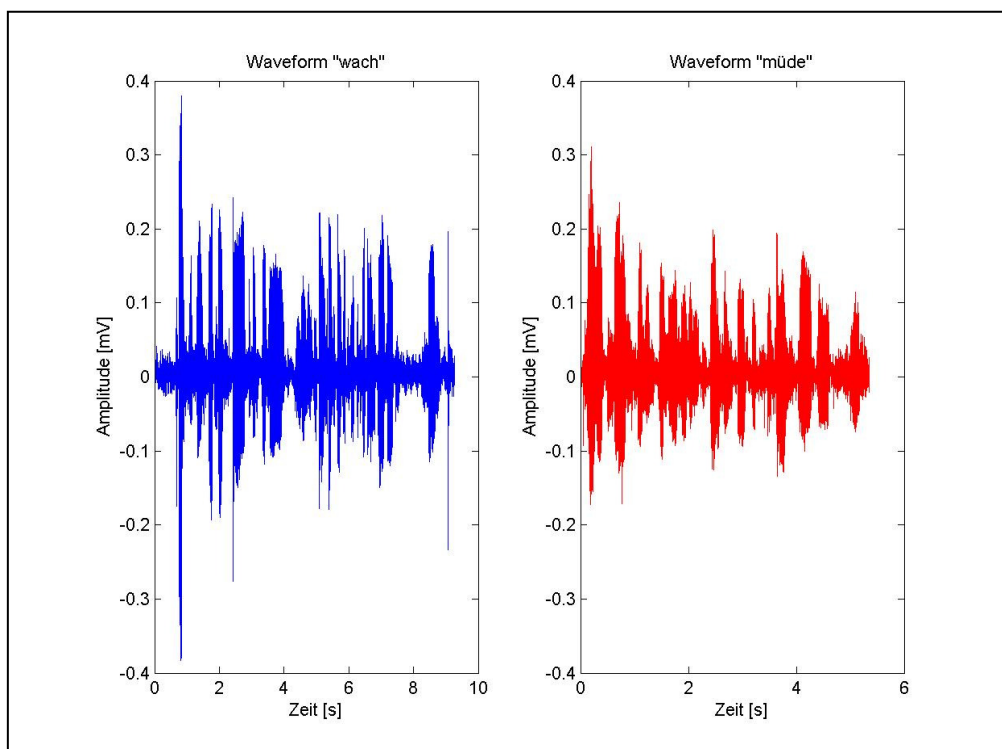


Abbildung 64: Rohwaveform der Introphrase für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 22)

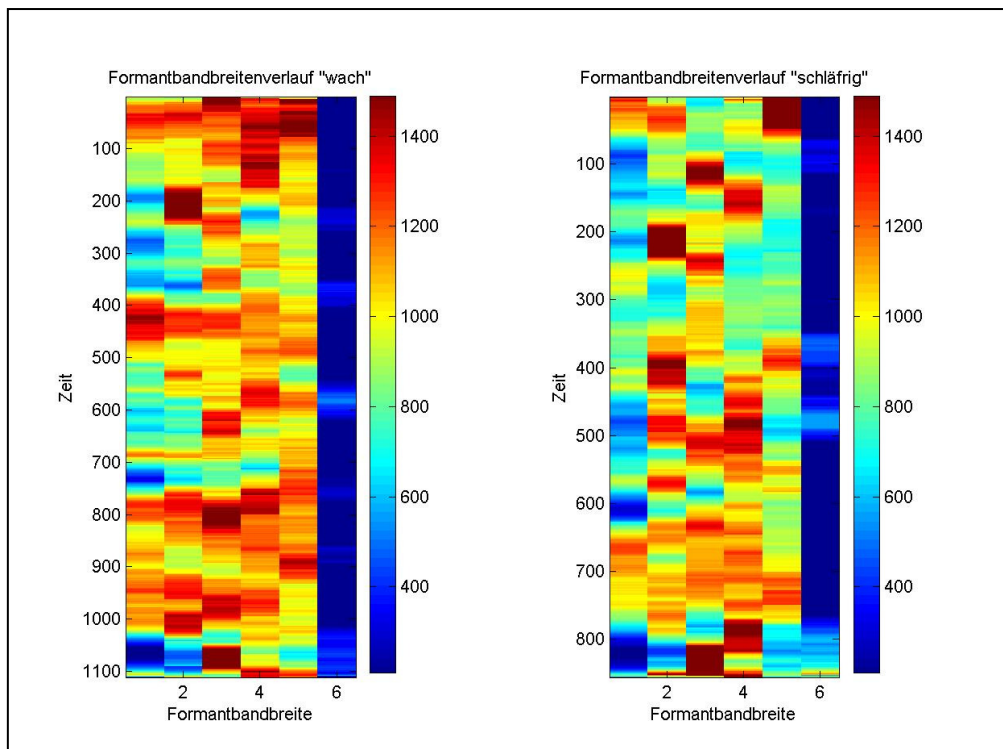


Abbildung 65: Formantbandbreitenverlauf (Formant Bandwidth) der Introphrase für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 15)

Wie aus den Abbildungen 64 und 65 ersichtlich wird, unterscheiden sich wache von schläfrigen Sprachaufnahmen in ihrem Verlauf der Rohwelleform (z.B. höhere mittlere Amplitude) und in ihren Formantbandbreitenverläufen (z.B. höheres 25stes Perzentil Bandwidth F1, höheres 50stes Perzentil Bandwidth F2).

4.2 Ergebnisse: Machine-Learning Klassifikation

In diesem Kapitel werden die Ergebnisse der auf dem beschriebenen Machine-Learning Prozedere basierenden Vorhersagen des Schläfrigkeitsniveaus präsentiert. Anders als im vorangegangenen Kapitel werden, hier ganze Kennzahlenbündel (hochdimensionale Merkmalsvektoren) für die Vorhersage genutzt (vgl. Abschnitt 4.5, Teil I). Der Fahrerassistenzsatz 3 („Das ist falsch, es muss ein Irrtum sein“), der Fahrerassistenzsatz 9 („Rufen Sie den Notdienst an“) sowie die Probandenvorstellungs-Introphrase dienen hier als Sprachmaterial.

Genutzt werden innerhalb der Machine-Learning Phase „Dimensionalitätsreduktion“ Korrelationsfilter-basierte Verfahren, PCA-basierte Dimensionalitätsreduktion (Extraktion von 100 Principle Components) sowie eine Wrapper-basierte Subset Selection mittels k-Nearest-Neighbour Verfahren mit einem Nachbarn (1-NN) (Genetic-Algorithm-basierte Suchstrategie). Die Darstellung der Vorhersageergebnisse für männliche und weibliche Sprecher beschränkt sich auf die Vorhersage von ungesehenen Testdaten mittels 10-facher Kreuzvalidierung. Die lediglich die Anpassungsfähigkeit des Klassifikators dokumentierenden Vorhersagegüten von Trainingsdaten werden im Weiteren nicht erwähnt.

Sprachmaterial “Fahrerassistenzsatz 3“: Klassifikationsergebnisse

Männliche Sprecher. Die Klassifikationsgenauigkeit der Schläfrigkeits-Vorhersage bei männlichen Sprecher beträgt 83.3% (SD 37.3%). Der positive prädiktive Wert liegt bei 63.6%, der negative prädiktive Wert bei 100%. Die Sensitivität des Verfahrens beträgt 100%, die Spezifität 76.5%. Die Korrelation zwischen den dichotomisierten KSS-Werten und der vorhergesagten Klassenzugehörigkeit liegt bei $r = .70$. Das Ergebnis des exakten Fisher-Tests bestätigt die Assoziation von vorhergesagten und selbstberichteten Schläfrigkeitszuständen ($p < .01$).

***Tabelle 32:** Klassifikationsergebnisse männlicher Sprecher des Fahrerassistenzsatz 3. Angegeben sind die Klassifikationsgenauigkeit (Accuracy: Anzahl der richtig klassifizierten durch alle Klassifikationen), Sensitivität (Detektionsrate der Schläfrigkeit), Spezifität (1 minus False Alarm Rate) sowie der positiver und negative prädiktiver Wert (PV_pos: Wahrscheinlichkeit der tatsächlichen Schläfrigkeit, wenn Schläfrigkeit vorhergesagt wurde; PV_neg: Wahrscheinlichkeit der tatsächlichen Wachheit, wenn Wachheit vorhergesagt wurde)*

	True a	True s	
Pred. a	13	0	PV_neg 100%
Pred. s	4	7	PV_pos 63.6%
	Spezifität 76.5%	Sensitivität 100.0%	Accuracy 83.3%

Weibliche Sprecher. Die Klassifikationsgenauigkeit der Schläfrigkeits-Vorhersage bei weiblichen Sprecher beträgt 83.5% (SD = 37.3%). Der positive prädiktive Wert liegt bei 87.5%, der negative prädiktive Wert bei 82.5%. Die Sensitivität des Verfahrens beträgt 50.0%, die Spezifität 97.1%. Die Korrelation zwischen den dichotomisierten KSS-Werten und der vorhergesagten Klassenzugehörigkeit liegt bei $r = .57$. Das Ergebnis des exakten Fisher-Tests bestätigt die Assoziation von vorhergesagten und selbstberichteten Schläfrigkeitszuständen ($p < .01$).

Tabelle 33: Klassifikationsergebnisse weiblicher Sprecher des Fahrerassistenzsatz 3. Angegeben sind die Klassifikationsgenauigkeit (Accuracy: Anzahl der richtig klassifizierten durch alle Klassifikationen), Sensitivität (Detektionsrate der Schläfrigkeit), Spezifität (1 minus False Alarm Rate) sowie der positiver und negative prädiktiver Wert (PV_pos: Wahrscheinlichkeit der tatsächlichen Schläfrigkeit, wenn Schläfrigkeit vorhergesagt wurde; PV_neg: Wahrscheinlichkeit der tatsächlichen Wachheit, wenn Wachheit vorhergesagt wurde)

	True a	True s	
Pred. a	33	7	PV_neg 82.5%
Pred. s	1	7	PV_pos 87.5%
	Spezifität 97.1%	Sensitivität 50.0%	Accuracy 83.3%

Sprachmaterial “Fahrerassistenzsatz 9“: Klassifikationsergebnisse

Männliche Sprecher. Die Klassifikationsgenauigkeit der Schläfrigkeits-Vorhersage bei männlichen Sprecher beträgt 85.0% (SD 18.9%). Der positive prädiktive Wert liegt bei 80.0%, der negative prädiktive Wert bei 84.2%. Die Sensitivität des Verfahrens beträgt 57.1%, die Spezifität 94.1%. Die Korrelation zwischen den dichotomisierten KSS-Werten und der vorhergesagten Klassenzugehörigkeit liegt bei $r = .57$. Das Ergebnis des exakten Fisher-Tests bestätigt die Assoziation von vorhergesagten und selbstberichteten Schläfrigkeitszuständen ($p < .01$).

***Tabelle 34:** Klassifikationsergebnisse männlicher Sprecher des Fahrerassistenzsatz 9. Angegeben sind die Klassifikationsgenauigkeit (Accuracy: Anzahl der richtig klassifizierten durch alle Klassifikationen), Sensitivität (Detektionsrate der Schläfrigkeit), Spezifität (1 minus False Alarm Rate) sowie der positiver und negative prädiktiver Wert (PV_pos: Wahrscheinlichkeit der tatsächlichen Schläfrigkeit, wenn Schläfrigkeit vorhergesagt wurde; PV_neg: Wahrscheinlichkeit der tatsächlichen Wachheit, wenn Wachheit vorhergesagt wurde)*

	True a	True s	
Pred. a	16	3	PV_neg 84.2%
Pred. s	1	4	PV_pos 80.0%
	Spezifität 94.1%	Sensitivität 57.1%	Accuracy 85.0%

Weibliche Sprecher. Die Klassifikationsgenauigkeit der Schläfrigkeits-Vorhersage bei weiblichen Sprecher beträgt 81.5% (SD = 10.9%). Der positive prädiktive Wert liegt bei 77.8%, der negative prädiktive Wert bei 82.1%. Die Sensitivität des Verfahrens beträgt 50.0%, die Spezifität 94.1%. Die Korrelation zwischen den dichotomisierten KSS-Werten und der vorhergesagten Klassenzugehörigkeit liegt bei $r = .51$. Das Ergebnis des exakten Fisher-Tests bestätigt die Assoziation von vorhergesagten und selbstberichteten Schläfrigkeitszuständen ($p < .01$).

Tabelle 35: Klassifikationsergebnisse weiblicher Sprecher des Fahrerassistenzsatz 9. Angegeben sind die Klassifikationsgenauigkeit (Accuracy: Anzahl der richtig klassifizierten durch alle Klassifikationen), Sensitivität (Detektionsrate der Schläfrigkeit), Spezifität (1 minus False Alarm Rate) sowie der positiver und negative prädiktiver Wert (PV_pos: Wahrscheinlichkeit der tatsächlichen Schläfrigkeit, wenn Schläfrigkeit vorhergesagt wurde; PV_neg: Wahrscheinlichkeit der tatsächlichen Wachheit, wenn Wachheit vorhergesagt wurde)

	True a	True s	
Pred. a	32	7	PV_neg 82.5%
Pred. s	2	7	PV_pos 77.8%
	Spezifität 94.1%	Sensitivität 50.0%	Accuracy 81.5%

Sprachmaterial “Intro“: Klassifikationsergebnisse

Männliche Sprecher. Die Klassifikationsgenauigkeit der Schläfrigkeits-Vorhersage bei männlichen Sprecher beträgt 92.0% (SD 13.3%). Der positive prädiktive Wert liegt bei 77.8%, der negative prädiktive Wert bei 95.1%. Die Sensitivität des Verfahrens beträgt 77.8%, die Spezifität 95.1%. Die Korrelation zwischen den dichotomisierten KSS-Werten und der vorhergesagten Klassenzugehörigkeit liegt bei $r = .73$. Das Ergebnis des exakten Fisher-Tests bestätigt die Assoziation von vorhergesagten und selbstberichteten Schläfrigkeitszuständen ($p < .01$).

***Tabelle 36:** Klassifikationsergebnisse männlicher Sprecher der Introphrase. Angegeben sind die Klassifikationsgenauigkeit (Accuracy: Anzahl der richtig klassifizierten durch alle Klassifikationen), Sensitivität (Detektionsrate der Schläfrigkeit), Spezifität (1 minus False Alarm Rate) sowie der positiver und negative prädiktiver Wert (PV_pos: Wahrscheinlichkeit der tatsächlichen Schläfrigkeit, wenn Schläfrigkeit vorhergesagt wurde; PV_neg: Wahrscheinlichkeit der tatsächlichen Wachheit, wenn Wachheit vorhergesagt wurde)*

	True a	True s	
Pred. a	39	2	PV_neg 95.1%
Pred. s	2	7	PV_pos 77.8%
	Spezifität 95.1%	Sensitivität 77.8%	Accuracy 92.0%

Weibliche Sprecher. Die Klassifikationsgenauigkeit der Schläfrigkeits-Vorhersage bei weiblichen Sprecher beträgt 76.9% (SD = 11.8%). Der positive prädiktive Wert liegt bei 62.5%, der negative prädiktive Wert bei 80.0%. Die Sensitivität des Verfahrens beträgt 41.7%, die Spezifität 90.3%. Die Korrelation zwischen den dichotomisierten KSS-Werten und der vorhergesagten Klassenzugehörigkeit liegt bei $r = .37$. Das Ergebnis des exakten Fisher-Tests bestätigt die Assoziation von vorhergesagten und selbstberichteten Schläfrigkeitszuständen ($p < .01$).

Tabelle 37: Klassifikationsergebnisse weiblicher Sprecher der Introphrase. Angegeben sind die Klassifikationsgenauigkeit (Accuracy: Anzahl der richtig klassifizierten durch alle Klassifikationen), Sensitivität (Detektionsrate der Schläfrigkeit), Spezifität (1 minus False Alarm Rate) sowie der positive und negative prädiktive Wert (PV_pos: Wahrscheinlichkeit der tatsächlichen Schläfrigkeit, wenn Schläfrigkeit vorhergesagt wurde; PV_neg: Wahrscheinlichkeit der tatsächlichen Wachheit, wenn Wachheit vorhergesagt wurde)

	True a	True s	
Pred. a	56	14	PV_neg 80.0%
Pred. s	6	10	PV_pos 62.5%
	Spezifität 90.3%	Sensitivität 41.7%	Accuracy 76.9%

5 ERGEBNISSE: SPRACHSEGMENT MULTIPLE PHRASALE EINHEIT

Nachdem im vorangegangenen Kapitel die Ergebnisse des Sprachsegments Phrasale Einheit beschrieben worden sind, beginnt dieses Kapitel mit einem Kurzüberblick zu den Hauptergebnissen des Kapitels 5. Im Anschluss daran werden korrelative Zusammenhänge zwischen sprachlichen Einzelmerkmalen Multipler Phrasaler Einheiten und Schläfrigkeitratings vorgestellt. Das nachfolgende Unterkapitel stellt die für die Validitätseinschätzung zentralen, multivariaten Machine-Learning Klassifikationsergebnisse dar.

Ergebniszusammenfassung Sprachsegment Multiple Phrasale Einheit

Sprachsegment multiple Phrasale Einheit „freie Bildbeschreibung“. Die linearen Zusammenhänge von 14 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeitzuständen erreichen eine mindestens kleine Effektstärke von $r > .2$. das Signifikanzniveau ($r < .23$). Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt eine Vorhersagegenauigkeit von 81.9% (Sensitivität = 61.9%, Spezifität = 88.9%).

5.1 Ergebnisse: Einzelkennzahlen

In diesem Kapitel werden Zusammenhänge zwischen sprachlichen Einzelkennzahlen des Segments Multiple Phrasale Einheit und dem selbstberichteten Schläfrigkeitniveau (KSS) präsentiert. Die Ergebnisse der aus der freien Bildbeschreibung entnommenen Multiplen Phrasalen Einheit dienen hier als Sprachmaterial. Zur weiteren Illustration der gefundenen korrelativen Analyseergebnisse werden jeweils für einzelne wache und schläfrige Stimmaufnahmen Kennwertverläufe dargestellt.

Sprachmaterial „freie Bildbeschreibung“: korrelative Einzelergebnisse

Korrelative Zusammenhänge zwischen Stimmkennzahlen der freien Bildbeschreibung und Schläfrigkeitratings. 14 Einzelkorrelationen zwischen Stimmkennzahlen der Bildbeschreibung und selbstberichteten Schläfrigkeitratings erreichen eine mindestens kleine Effektstärke von $r > .2$. Die fünf betragsmäßig höchsten linearen Zusammenhänge werden für die folgenden Merkmale erzielt:

Delta-Delta-MFCC1: .27 (Mittelwert der zweiten Ableitung des Verlaufs des Mel-Frequency Cepstrum Coefficient 1); Delta-MFCC7: .26 (Mittelwert der ersten Ableitung des Verlaufs des Mel-Frequency Cepstrum Coefficient 7); MFCC9: .25; (Mittelwert des Verlaufs des Mel-Frequency Cepstrum Coefficient 9); Itas_maxfreq: -.30 (Maximum Frequenz des Long Term Average Spectrums); f0_sd_z: .25 (z-transformierte Standardabweichung des Verlaufs der Fundamental Frequenz).

Tabelle 38: Korrelative Zusammenhänge ($r > .20$) zwischen Sprachmerkmalen der freien Bildbeschreibung und KSS-Werten.

f1_maxtime_rel	f1_mean	delta-delta-MFCC1	delta7	MFCC9	baendi0_500_500_4000	ltas_maxfreq
-.20	.20	.27	.26	.25	-.20	-.30
f0_sd_z	f0_75quant_z	f0_slope_z	f4_mintime_rel_z	MFCC3_z	delta-delta-MFCC5_z	baendi0_500_500_4k_z
.25	.20	.23	.24	-.22	.23	-.20

Im Weiteren werden die aus den korrelativen Analysen gewonnenen Hinweise auf schläfrigkeitssensitive sprachliche Einzelmerkmale ähnlich wie für das Sprachmaterial zuvor exemplarisch dargestellt.

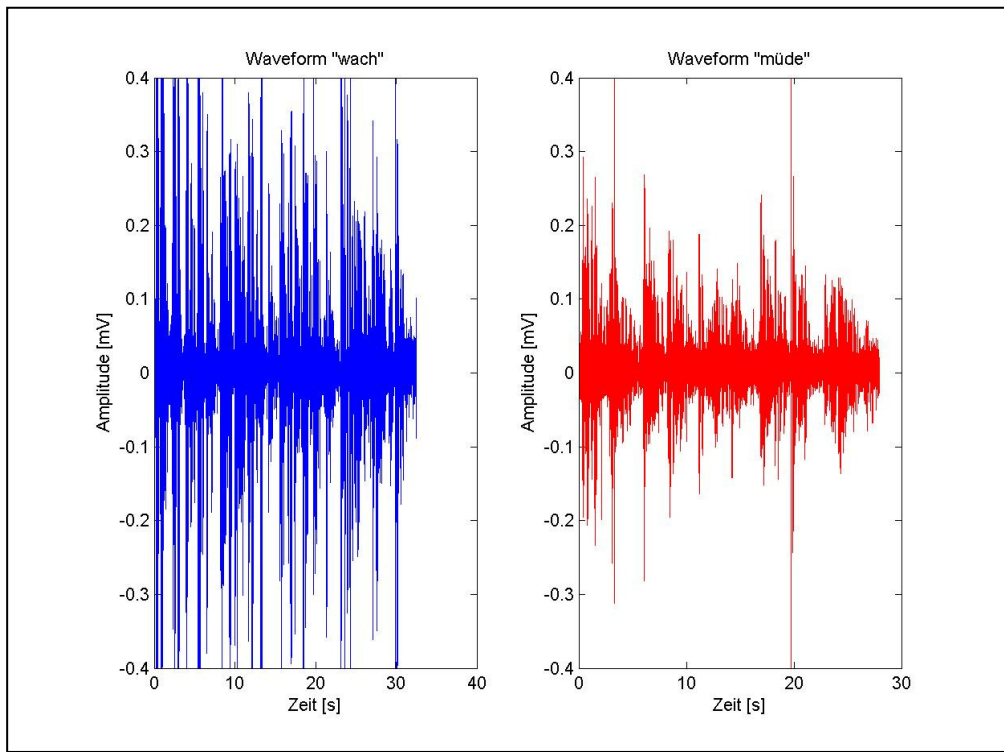


Abbildung 66: Rohwaveform der freien Bildbeschreibung für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 17)

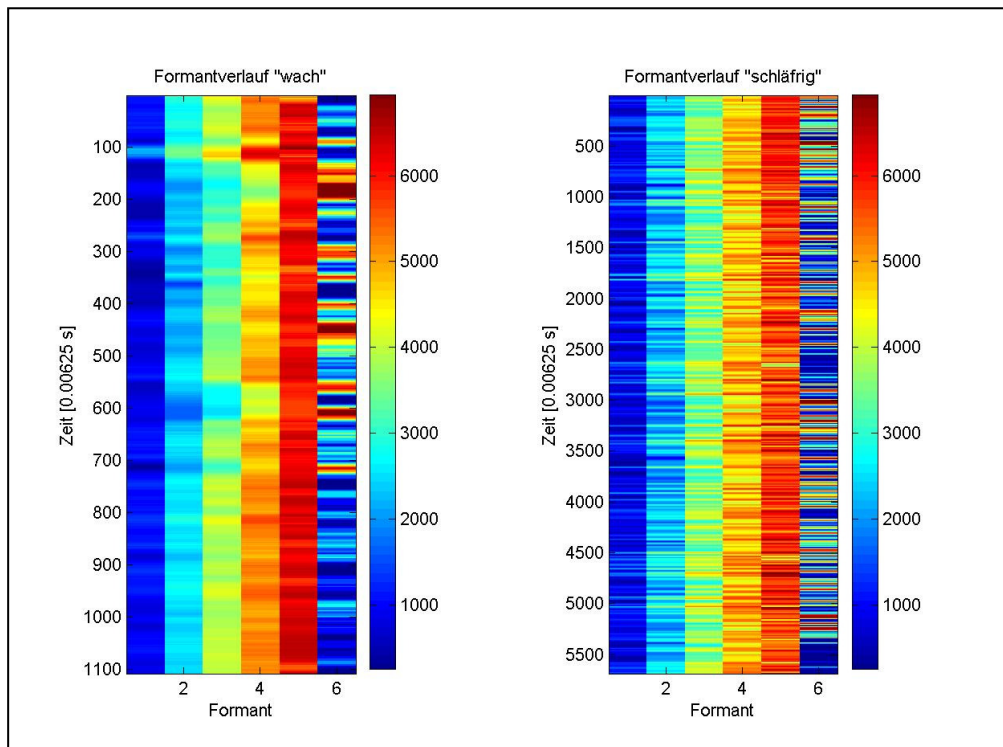


Abbildung 67: Formantverläufe der freien Bildbeschreibung für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 14)

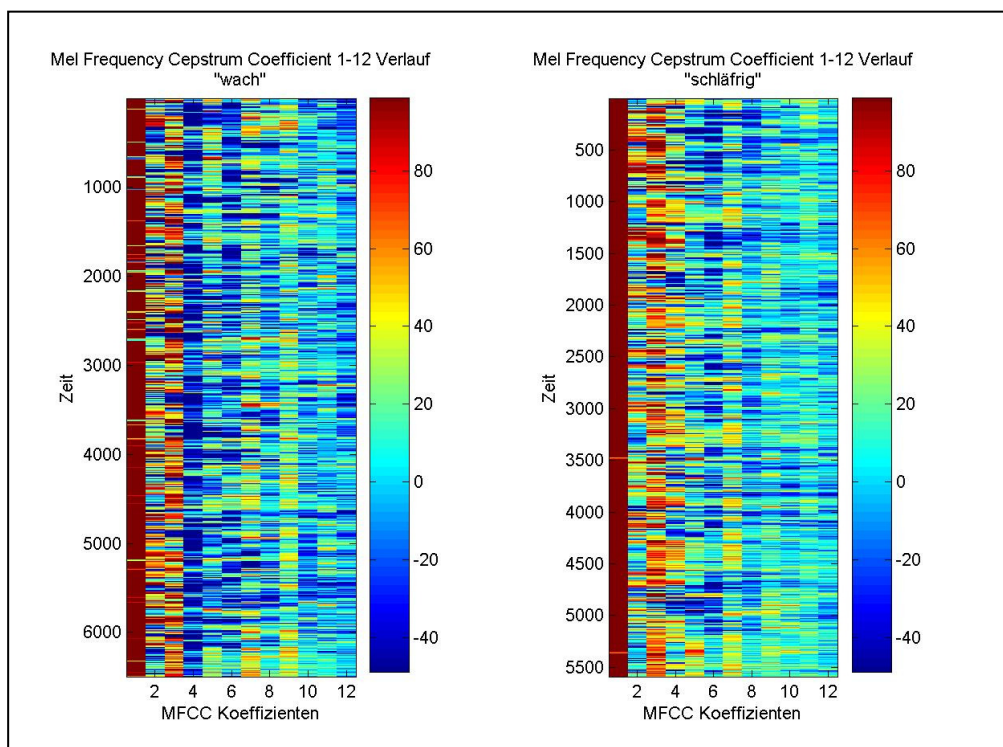


Abbildung 68: 12 MFCC Verläufe der freien Bildbeschreibung für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 11)

Wie aus den Abbildungen 66, 67 und 68 ersichtlich wird, unterscheiden sich wache von schläfrigen Sprachaufnahmen in ihrem Verlauf der Rohwaveform (z.B. höhere mittlere Amplitude), in ihren Formantpositionsverläufen (z.B. niedrigerer Mittelwert F1) und in ihren MFCC Verläufen (z.B. niedrigerer Mittelwert MFCC9).

5.2 Ergebnisse: Machine-Learning Klassifikation

In diesem Kapitel werden die Ergebnisse der auf dem beschriebenen Machine-Learning Prozedere basierenden Vorhersagen des Schläfrigkeitsniveaus präsentiert. Anders als im vorangegangenen Kapitel werden, hier ganze Kennzahlenbündel (hochdimensionale Merkmalsvektoren) für die Vorhersage genutzt (vgl. Abschnitt 4.5, Teil I). Die hier bestimmten Vorhersageergebnisse basieren auf dem Sprachmaterial der freien Bildbeschreibung. Das Kapitel schließt mit der Darstellung der Klassifikationsergebnisse aller Sprachmaterialien in einen Gesamtüberblick.

Genutzt werden innerhalb der Machine-Learning Phase „Dimensionalitätsreduktion“ Korrelationsfilter-basierte Verfahren, PCA-basierte Dimensionalitätsreduktion (Extraktion von 100 Principle Components) sowie eine Wrapper-basierte Subset Selection mittels k-Nearest-Neighbour Verfahren mit einem Nachbarn (1-NN) (Genetic-Algorithm-basierte Suchstrategie). Die Darstellung der Vorhersageergebnisse für männliche und weibliche Sprecher beschränkt sich auf die Vorhersage von ungesehenen Testdaten mittels 10-facher Kreuzvalidierung. Die lediglich die Anpassungsfähigkeit des Klassifikators dokumentierenden Vorhersagegütern von Trainingsdaten werden im Weiteren nicht erwähnt.

Männliche Sprecher. Die Klassifikationsgenauigkeit der Schläfrigkeits-Vorhersage bei männlichen Sprecher beträgt 90.0% (SD = 15.2%). Der positive prädiktive Wert liegt bei 75.0%, der negative prädiktive Wert bei 91.7%. Die Sensitivität des Verfahrens beträgt 60.0%, die Spezifität 95.6%. Die Korrelation zwischen den dichotomisierten KSS-Werten und der vorhergesagten Klassenzugehörigkeit liegt bei $r = .60$. Das Ergebnis des exakten Fisher-Tests bestätigt die Assoziation von vorhergesagten und selbstberichteten Schläfrigkeitszuständen ($p < .05$).

***Tabelle 39:** Klassifikationsergebnisse männlicher Sprecher der freien Bildbeschreibung. Angegeben sind die Klassifikationsgenauigkeit (Accuracy: Anzahl der richtig klassifizierten durch alle Klassifikationen), Sensitivität (Detektionsrate der Schläfrigkeit), Spezifität (1 minus False Alarm Rate) sowie der positiver und negative prädiktiver Wert (PV_pos: Wahrscheinlichkeit der tatsächlichen Schläfrigkeit, wenn Schläfrigkeit vorhergesagt wurde; PV_neg: Wahrscheinlichkeit der tatsächlichen Wachheit, wenn Wachheit vorhergesagt wurde)*

	True a	True s	
Pred. a	22	2	PV_neg 91.7%
Pred. s	1	3	PV_pos 75.0%
	Spezifität 95.6%	Sensitivität 60.0%	Accuracy 90.0%

Weibliche Sprecher. Die Klassifikationsgenauigkeit der Schläfrigkeits-Vorhersage bei weiblichen Sprecher beträgt 77.0% (SD = 15.4%). Der positive prädiktive Wert liegt bei 60.0%, der negative prädiktive Wert bei 82.9%. Die Sensitivität des Verfahrens beträgt 56.2%, die Spezifität 85.0%. Die Korrelation zwischen den dichotomisierten KSS-Werten und der vorhergesagten Klassenzugehörigkeit liegt bei $r = .42$. Das Ergebnis des exakten Fisher-Tests bestätigt die Assoziation von vorhergesagten und selbstberichteten Schläfrigkeitszuständen ($p < .01$).

Tabelle 40: Klassifikationsergebnisse weiblicher Sprecher der freien Bildbeschreibung. Angegeben sind die Klassifikationsgenauigkeit (Accuracy: Anzahl der richtig klassifizierten durch alle Klassifikationen), Sensitivität (Detektionsrate der Schläfrigkeit), Spezifität (1 minus False Alarm Rate) sowie der positiver und negative prädiktiver Wert (PV_pos: Wahrscheinlichkeit der tatsächlichen Schläfrigkeit, wenn Schläfrigkeit vorhergesagt wurde; PV_neg: Wahrscheinlichkeit der tatsächlichen Wachheit, wenn Wachheit vorhergesagt wurde)

	True a	True s	
Pred. a	34	7	PV_neg 82.9%
Pred. s	6	9	PV_pos 60.0%
	Spezifität 85.0%	Sensitivität 56.2%	Accuracy 77.0%

Tabelle 41: Gesamtüberblick über die Klassifikationsergebnisse in % (bis auf die Korrelationen) von Schläfrigkeit Zuständen mittels diverser Sprachmaterialien. (PV_pos: positiver prädiktiver Wert; PV_neg: negativer prädiktiver Wert; M: Männer; F: Frauen). Korrelationen in einem betragsmäßigen Wertebereich über $r \geq .36$ (Männer) und über $r \geq .27$ (Frauen) sind mit einem „*“ markiert und auf dem Niveau von $p < .05$ signifikant (zweiseitige Testung)

	Extrahiertes [a:]		Extrahiertes [o:]		Isoliertes [a:]		Wort 1		Satz 3		Satz 9		Intro		Bild	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F
Korrelation	.69*	.60*	.48*	.46*	.77*	.31*	.58*	.37*	.70*	.57*	.57*	.51*	.73*	.37*	.61*	.42*
Acc. Rate	88.33	83.50	81.67	77.50	91.67	73.00	85.00	75.50	83.33	83.33	85.00	81.50	92.00	76.94	90.00	77.00
Spezifität	94.12	88.24	88.24	82.35	96.30	85.00	100.00	85.29	76.47	97.06	94.12	94.12	77.78	90.32	95.65	85.00
Sensitivität	71.43	71.43	57.14	64.29	77.78	43.75	40.00	50.00	100.00	50.00	57.14	50.00	95.12	41.67	60.00	56.25
PV_neg	88.89	88.24	83.33	84.85	92.86	79.07	83.33	80.56	100.00	82.50	84.21	82.05	95.12	80.00	91.67	82.93
PV_pos	83.33	71.43	66.67	60.00	87.50	53.85	100.00	58.33	63.64	87.50	80.00	77.78	77.78	62.50	75.00	60.00

6 DISKUSSION

Das zentrale Ziel der vorliegenden Arbeit ist der empirische Nachweis des Zusammenhangs zwischen akustischen Stimmkennwerten und Schläfrigkeit Zuständen. Die Ergebnisse der Studie werden im Folgenden nach Sprachmaterialien getrennt zusammengefasst und mit bisherigen empirischen Befunden und theoretischen Prognosen abgeglichen. Darüber hinaus wird die Aussagekraft der hier präsentierten Validierungsstudie kritisch eingeschätzt, Einschränkungen herausgearbeitet und Forschungsdesiderate abgeleitet. Das Kapitel schließt mit einem Ausblick auf das Anwendungspotenzial der akustischen Stimmanalyse innerhalb psychologischer und nicht-psychologischer Anwendungsfelder.

6.1 Zusammenfassung der Ergebnisse

Sprachsegment extrahierter Vokal [a:]. Die linearen Zusammenhänge von 42 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeit Zuständen erreichen eine mindestens kleine Effektstärke von $r > .2$. Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt eine Vorhersagegenauigkeit von 80.7% (Sensitivität = 80.4%, Spezifität = 80.9%).

Sprachsegment extrahierter Vokal [o:]. Die linearen Zusammenhänge von 45 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeit Zuständen erreichen eine mindestens kleine Effektstärke von $r > .2$. Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt eine Vorhersagegenauigkeit von 79.6% (Sensitivität = 60.7%, Spezifität = 83.3%).

Sprachsegment isolierte Vokalphonation [a:]. Die linearen Zusammenhänge von 19 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeit Zuständen erreichen eine mindestens kleine Effektstärke von $r > .2$. Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt eine Vorhersagegenauigkeit von 83.3% (Sensitivität = 53.6%, Spezifität = 94.1%).

Sprachsegment Wort „suchen“. Die linearen Zusammenhänge von 32 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeit Zuständen erreichen eine mindestens kleine

Effektstärke von $r > .2$. Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt eine Vorhersagegenauigkeit von 80.3% (Sensitivität = 45.0%, Spezifität = 92.7%).

Sprachsegment Phrasale Einheit „Satz 3“. Die linearen Zusammenhänge von 35 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeit Zuständen erreichen eine mindestens kleine Effektstärke von $r > .2$. Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt eine Vorhersagegenauigkeit von 83.3% (Sensitivität = 75.0%, Spezifität = 86.8%).

Sprachsegment Phrasale Einheit „Satz 9“. Die linearen Zusammenhänge von 32 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeit Zuständen erreichen eine mindestens kleine Effektstärke von $r > .2$. Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt eine Vorhersagegenauigkeit von 82.3% (Sensitivität = 60.7%, Spezifität = 90.7%).

Sprachsegment Phrasale Einheit „Intro-Phrase“. Die linearen Zusammenhänge von 14 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeit Zuständen erreichen eine mindestens kleine Effektstärke von $r > .2$. Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt eine Vorhersagegenauigkeit von 84.5% (Sensitivität = 68.4%, Spezifität = 84.1%).

Sprachsegment multiple Phrasale Einheit „freie Bildbeschreibung“. Die linearen Zusammenhänge von 14 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeit Zuständen erreichen eine mindestens kleine Effektstärke von $r > .2$. Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt eine Vorhersagegenauigkeit von 81.9% (Sensitivität = 61.9%, Spezifität = 88.9%).

Zusammenfassung aller verwendeten Sprachsegmente. Die linearen Zusammenhänge von durchschnittlich 29.13 sprachlichen Einzelmerkmalen und selbstberichteten Schläfrigkeit Zuständen erreichen das Effektstärkenkriterium ($r < .2$). Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergibt durchschnittlich eine Vorhersagegenauigkeit von 82.0% (Sensitivität = 63.4%, Spezifität = 87.8%).

Tabelle 42: Gesamtüberblick über die Klassifikationsergebnisse (bis auf die Korrelationen in % angegeben) von Schläfrigkeit Zuständen mittels diverser Sprachmaterialien. (PV_pos: positiver prädiktiver Wert; PV_neg: negativer prädiktiver Wert; Einzelkorr: Anzahl der inhaltlich bedeutsamen Einzelkorrelationen. Korrelationen in einem betragsmäßigen Wertebereich über $r \geq .23$ (mit „*“ markiert) sind auf dem Niveau von $p < .05$ signifikant (zweiseitige Testung))

	Extrah. [a:]	Extrah. [o:]	Isol. [a:]	Wort 1	Satz 3	Satz 9	Intro	Bild	GESAMT
Korrelation	0.57*	0.47*	0.54*	0.48*	0.64*	0.54*	0.55*	0.62*	0.55*
Accuracy Rate	80.71	79.59	83.25	80.25	83.33	82.34	84.47	81.94	81.99
Spezifität	80.95	85.30	94.12	92.65	86.77	90.65	84.05	88.89	87.78
Sensitivität	80.39	60.72	53.57	45.00	75.00	60.77	68.40	61.90	63.40
PV_neg	91.11	84.09	83.13	81.95	91.25	85.97	87.56	87.50	86.43
PV_pos	62.96	63.34	78.89	79.17	75.57	70.68	70.14	65.00	71.53
Einzelkorr.	42	45	19	32	35	32	14	14	29.13

Wie aus Tabelle 42 zu entnehmen ist, wurde somit das Hauptziel der Arbeit, aus sprachlicher Interaktion Schläfrigkeit zu detektieren, erreicht.

Sprachmaterialspezifische Vorhersagegüte. Die vorgestellten Ergebnisse dokumentieren eine relativ homogene Gesamtvorhersagevalidität und Spezifität für alle Sprachmaterialien. Grundsätzlich erscheinen somit alle verwendeten Sprachsegmenteinheiten (Vokal, Wort, Phrasale Einheit, multiple Phrasale Einheit) geeignet, Schläfrigkeit vorherzusagen. Die Sensitivität hingegen variiert deutlicher und erzielt für das extrahierte [a:] die besten Detektionsraten. Hinsichtlich dieser Detektionsraten schneidet das Sprachmaterial „suchen“ am schlechtesten ab. Der Grund hierfür könnte darin liegen, dass die eindeutige Bestimmung der Wortgrenzen innerhalb der Segmentierungsphase mit Unsicherheiten behaftet ist und daher zur Fehlervarianz beiträgt. In dieser Größenordnung unerwartet ist auch die Vorhersagepräzision für die bezüglich der verwendeten Worte nicht restringierten freien Bildbeschreibungsaufgabe.

Geschlechtsspezifische Vorhersagegüte. Die vorgestellten Ergebnisse dokumentieren eine deutlich bessere Vorhersagevalidität der Sprachkennwerte für Männer. Da die Anzahl der untersuchten männlichen Stimmproben gering war, wäre zu erwarten gewesen, dass auch eine im Vergleich zu den Frauen verringerte Klassifikationsgüte zu verzeichnen ist. Eine mögliche Ursache für diese Diskrepanz könnte darin begründet liegen, dass Männer systematisch ihre Schläfrigkeit unterschätzen und erst bei Vorliegen starker Schläfrigkeit diese zugeben. Die dann aufgrund dieser Selbstreporte als schläfrig kategorisierten männlichen Sprecher, besitzen eine extremere und somit eindeutiger zu vorhersagende Schläfrigkeit. Des Weiteren erscheint es denkbar, dass bei den weiblichen Sprechern mehr oder auch schwieriger zu diskriminierende Schläfrigkeitssubformen vorlagen. Männliche Sprecher könnten in diesem Sinne bei Schläfrigkeit homogenere, einfachere und eindeutiger zu prognostizierende Reaktionsformen präsentieren.

Mittlere Sensitivität. Im Gegensatz zu der sehr guten Spezifität ist die Sensitivität lediglich als zufriedenstellend einzuordnen. Grund hierfür könnte in einigen Sprachmaterialien die ungleiche Verteilung von schläfrigen und wachen Sprachproben sein. Der vorliegende Überhang von wachen Sprachproben erleichtert es den Machine-Learning Verfahren diese zu identifizieren. Ein weiterer Grund für die mittlere Sensitivität kann in der vorgenommenen Dichotomisierung der Schläfrigkeitswerte gesehen werden. Vor allem Klassifikationsfehler im

Trennbereich zwischen KSS=5 (nicht schläfrig) und KSS=6 (schläfrig) sind für die mittlere Sensitivität verantwortlich. Die Höhe der gefundenen Zusammenhänge, war vor dem Hintergrund, dass aufgrund des gewählten Messdesigns, nur verhältnismäßig wenige schläfrige Messungen erhoben werden konnten (ca. 30%) und die Schläfrigkeit oft nicht maximal ausgeprägt war, überraschend hoch. Es ist zu erwarten, dass bei einer Monotonisierung der Messsituation und einer Verlängerung der Schlafdeprivation extremere Schläfrigkeitszustände erzielt worden wären. Diese extremeren Schläfrigkeitszustände wiederum wären eindeutiger klassifizierbar. Darüber hinaus erleichtert eine vergrößerte Lernstichprobe die klare Identifikation von Schläfrigkeits-Subgruppen und verbessert somit die zu erwartende Sensitivität.

Vergleich der Ergebnisse mit den Annahmen des Kognitiv-phonetischen Mediatormodells. Abgeleitet aus dem Kognitiv-physiologischen Mediatormodell zu schläfrigkeitsinduzierten Stimmveränderungen wurde erwartet, dass sich schläfrigkeitsinduzierte Veränderungen der Intensität (Lautstärke), Grundfrequenz (F_0), Formanten (F_1 - F_5) und der Bandbreite der Formanten ergeben. Die zentralen aus den Kognitiv-physiologischen Mediatormodell abgeleiteten Veränderungen konnten an unterschiedlichen Sprachsegmenten (z.B. Phrasale Einheit: Fahrerassistenzsatz 9) nachgewiesen werden. Darüber hinaus zeigten sich jedoch auch für eine Vielzahl weiterer im Modell nicht beschriebener Kennzahlenbereiche schläfrigkeitsinduzierte Veränderungen (z.B. MFCC, Δ MFCC, $\Delta\Delta$ MFCC, Hammarberg Indices).

Vergleich der Ergebnisse mit bisherigen empirischen Befunden. Wichtige Befunde der bisherigen Forschungen zur akustischen Detektion von Schläfrigkeitszuständen konnten in der hier vorliegenden Studie repliziert werden. Der von Morris et al. (1960) und Nwe et al. (2006) dokumentierte Zusammenhang von schlafdeprivierter Sprache und abnehmender Variation des Tonhöhenverlaufs konnte in der vorliegenden Untersuchung nachgewiesen werden. Der von Whitmore und Fisher (1996) gefundene Zusammenhang zwischen wachsender Schläfrigkeit und einer Abnahme der Tonhöhe wurde ebenfalls nachgewiesen. Des Weiteren finden sich auch die von Harrison & Horne (1997) postulierten Intensitäts-bezogenen stimmlichen Veränderungen. Der in der Untersuchung von Greely (2006) dokumentierte Befund, sich verändernder MFCCs, konnte in der vorliegenden Arbeit ebenfalls bestätigt werden. Keine Entsprechung findet sich für die in der Studie von Morris et al. (1960) gefundenen ungewöhnlichen Lautstärke-Variationen, die hier lediglich für als generelle Lautstärken-Abnahme in Erscheinung traten.

Einige zentrale Befunde zur stimmlichen Detektion der Aktivierungsdimension der Emotion wurden in der hier vorliegenden Studie ebenfalls repliziert. So zeigt sich übereinstimmend, dass Schläfrigkeit mit tiefen F_0 -Werten, flachen Intonationsverläufen, niedrigen Intensitäten, niedrigen Positionen des ersten Formanten, schmaler Bandbreite von F_1 , sowie geringer spektraler Energie der Frequenzen über 500Hz verbunden ist. Damit bestätigen die Ergebnisse die Befunde von Juslin & Laukka (2001) und Scherer (1986). Ebenso konnten die Befunde von Laukka et al. (2005), wonach niedrige Aktiviertheit mit niedrigem F_0 -Mittelwert und schmaler F_0 -Bandbreite assoziiert ist, bestätigt werden. Nicht bestätigt wurden die Befunde zu schläfrigkeitsinduzierten Intensitätsschwankungen von Juslin & Laukka (2001) und zum F_0 -Maximum von Laukka et al. (2005).

6.2 Fehleranalyse, Datenqualität und Validitätsgefährdung

Interne Validitätsüberprüfung

Der nächste Abschnitt widmet sich der Einschätzung der Datengüte, der wichtigsten methodologischen Einschränkungen, potentiellen Fehlerquellen sowie Validitätsgefährdungen. Unerlässlich für die Beurteilung der Aussagekraft der Studie ist die Beantwortung der Frage: *Ist die Untersuchungsanlage geeignet die Validität der Stimmanalyse zu überprüfen?* Zu unterscheiden sind in diesem Zusammenhang die Fehlersphären der (a) Instrumenten-immanenten sowie der (b) Versuchsdesign-immanenten Fehlerquellen. Einzuschätzen ist daher zum einen die Frage nach dem Goldstandard-Charakter des Validierungsankers Karolinska Sleepiness Scale (*Ground Truth*) und zum anderen die Frage nach konfundierenden Größen, die für sich oder in Kombination als Alternativerklärung der Stimmerkmalveränderungen in Frage kommen.

Instrumenten-immanente Fehlerperspektive (Goldstandard-Charakter des Validierungsankers). Die als Selbstreport-Validierungsanker verwendete Karolinska Sleepiness Scale unterliegt den klassischen Einschränkungen von Selbstreportverfahren: willentliche Beeinflussbarkeit der Ergebnisse verbunden mit möglichen kognitiven und motivationalen Verzerrungen. Auch wenn eine willentliche Beeinflussung der Ergebnisse nicht generell auszuschließen ist, kann bei der vorliegenden Untersuchung nicht davon ausgegangen werden, dass die Schläfrigkeitsratings willentlich durch die Versuchspersonen verfälscht wurden. Eine vertrauensvolle Versuchsleiter- und Probandenbeziehung sicherte die instruktionsgerechte Compliance und somit eine sorgfältige Schläfrigkeitsbeurteilung.

Designimmanente Fehlerquellen (Störvariablen-bezogene Alternativerklärungen). Selbst wenn die Messung der Ground Truth Schläfrigkeit mittels KSS valide vollzogen wurde, bleibt die Frage, in wieweit Alternativerklärungen für die sich im Nachtverlauf verändernden Stimmergebnisse verantwortlich gemacht werden können. Zu diesen Störvariablen zählen wachsende Vertrautheit des Sprachmaterials, wachsende Frustrationszustände, sinkende ängstliche Anspannung, wachsende Selbstaktivierungsbemühungen und schläfrigkeitstereotyp gerechte Verzerrungen der Sprache. Gemeinsam ist diesen Störvariablen die zugrundeliegende Vermutung, dass ihre Veränderung und nicht die der veränderten Schläfrigkeit ursächlich für die veränderten Sprachkennwerte sind.

Aufgrund der mit dem *within-subject* Messdesign verbundenen Messwiederholungen, ist von einer wachsenden Vertrautheit mit dem Sprachmaterial auszugehen. Aus ursprünglich gelesenen, werden zunehmend freigesprochene Sprechtexte. Das Einführen eines anfänglichen Trainingsdurchgangs begegnet dieser Schwierigkeit. Des Weiteren sind mit steigender Wiederholungszahl des Sprachmaterials auch Sättigungs- und Frustrationsempfindungen zu vermuten. Diese Vermutung konnte anhand von parallel erhobenen Frustrations-Messungen nicht bestätigt werden. Denkbar sind auch Kombinationszustände von Schläfrigkeit und ängstlicher (aus der Untersuchungssituation resultierender) Anspannung, was vor allem auf die ersten wachen Messungen zutreffen könnte. Aus diesem Grund wurde die erste Trainingsmessung nicht gewertet. Folglich kann davon ausgegangen werden, dass die verbleibenden Messungen in einem vergleichbar entspannten, wenig ängstlichen Anspannungszustand vorgenommen worden sind. Denkbar erscheint, dass ein im Nachtverlauf steigender Aktivierungsbedarf über Tabakkonsum befriedigt werden kann. Mögliche Nebenwirkungen dieses Zigarettenkonsums auf die Stimmqualität sind zwar nicht bekannt, aber dennoch nicht auszuschließen. Die Nikotinkarrenz-Vorgabe sowie ihre fortwährende Kontrolle machen ein instruktionswidriges Rauchen jedoch unwahrscheinlich.

Des Weiteren könnten volitionale Selbstaktivierungsversuche selbst, wie das Anspannen der Muskulatur müde Stimmen kurzfristig maskieren (vgl. Engle-Friedmann, 2003). Dieser Zusammenhang bleibt jedoch völlig hypothetisch. Die letzte Alternativerklärung für eine sich über den Nachtverlauf verändernde Stimme ist die Vermutung der willentlich verzerrten, Schläfrigkeitsstereotyp-gerechten Sprache. Diese Vermutung könnte abschließend nur über die Verwendung von Studiendesigns erfolgen, die eine bewusste Manipulation der dargestellten Schläfrigkeit instruieren (*faking-good* oder *faking-bad* Designs). Zusammenfassend sollte festgehalten werden, dass der Einfluss einiger Störvariablen in weiteren Untersuchungen gezielt kontrolliert werden sollte. Dienlich könnte es in diesem Sinne daher sein, die hier gewonnenen Ergebnisse des *within-subject* Designs mit einem *between-subject* Design (z.B. mit einem randomisierten Kontrollgruppendesign, RCT) zu replizieren.

Externe Validitätsüberprüfung (Generalisierbarkeit)

Die Überprüfung der externen Validität bezieht sich auf die Frage, inwieweit die gefundenen Ergebnisse auf andere Personengruppen und andere Anwendungskontexte generalisierbar sind. Als wichtiges Bestimmungsstück der externen Validität kann daher die Realitätsnähe des verwendeten Untersuchungssettings gelten. Dieser als ökologische Validität bezeichneter Sachverhalt sichert die Übertragbarkeit von Laborbedingungen auf natürliche Situationen.

Bewertung der erzielten Klassifikationsleistungen. Bei der Bewertung der erzielten Klassifikationsleistungen sollten insbesondere externe Validitätsüberlegungen eine Rolle spielen. Für die Einschätzung der erzielten Detektionsraten sind daher insbesondere folgende Fragen zu beantworten:

- Gibt es eine an die anvisierte Anwendungssituation angepasste *Diversivität der Stichprobe* an die Zusammensetzung der Zielpopulation (bzgl. allgemeiner Sprechermerkmale wie z.B. Alter, Geschlecht und Dialekt)?
- Gibt es eine an die anvisierte Anwendungssituation angepasste *Diversivität von Zuständen* (z.B. Langeweile, Motivationslosigkeit, die mit Schläfrigkeit verwechselt werden könnten)? Oder die False Alarm Rate beeinflussen könnten? Sind alle Zustände -und nur genau die- repräsentativ erfasst, zwischen denen das System differenzieren können soll?
- Sind zu erwartende Zustandskombinationen erfasst, die die Detektion des Schläfrigkeitszustandes maskieren könnten und somit die Sensitivität des Instruments beeinflussen könnten (z.B. *schläfrigkeitsmaskierende Zustandskombinationen* wie schläfrig und ängstlich, schläfrig und gereizt, schläfrig und bemüht)?
- Besteht Übereinstimmung in den gemachten Sprechsituationsannahmen (sprecherbezogene Trainingsmöglichkeiten, Sprecherüberlappung, sprachinhaltsbezogene Restriktionen, wie z.B. vorgegebenes Sprachmaterial) und der realen Anwendungssituation?

Zur Einschätzung der anwendungsbezogenen Detektionsrate (Sensitivität) muss gewährleistet sein, dass eventuelle schläfrigkeitmaskierende Zustandskombinationen aktueller internaler Befindlichkeitszustände (wie Stress, innere Anspannung, Angst und weitere körperliche Zustände wie Erkältung, Milch getrunken, Schmerzen) auch in repräsentativer Form erfasst worden sind. Aufgrund der testbezogenen Notwendigkeit des Wachbleibens und des verwendeten Messdesigns, das fortlaufende Testanforderungen vorsah, ist im vorliegenden Fall von einem volitionalen Ankämpfen (“Fighting against Sleepiness“) gegen Schläfrigkeit auszugehen. Dieses bewusste Unterdrücken von Schläfrigkeit durch gezielte Aktivierung, kann zur Maskierung von Schläfrigkeit führen. Es ist davon auszugehen, dass die “Fighting against Sleepiness“ Komponente auch in alltäglichen Situationen, wie bei nächtlichem Autofahren, häufig vorkommt. Die gefundene Detektionsrate stellt demnach vermutlich eine reelle Annäherung an den tatsächlichen Wert dar.

Problematik der Stichprobengröße. Eine Determinante der externen Validität liegt wie bereits beschrieben in Art und Größe der verwendeten Stichprobe vor. Welche Einschränkungen sind dementsprechend mit geringen Stichprobengrößen im Allgemeinen und welche im Speziellen mit dieser Untersuchung verbunden? Im Zuge dieser Überlegungen finden sich zwei Problemkreise geringer Stichprobengrößen.

Signifikanzstatistische Zufallsabsicherung. Der vermeintlich größte Widerstand gegen niedrige Fallzahlen resultiert aus forschungspragmatischen Zwängen der Signifikanzstatistik (zur Irrelevanz und Trivialität der gängigen signifikanz-statistischen Praxis siehe Brandstätter, 1999; Chow, 1988, 1991a, 1991b, 1996, 1998; Cohen, 1994; Cortina & Dunlap, 1997; Cowles, 1989; Falk & Greenbaum, 1995; Folger, 1989; Gigerenzer, 1993; Gigerenzer & Murray, 1987; Hager, 2000; Macdonald, 1997; Oakes, 1987; Pollard, 1993; Schmidt, 1996; Sedlmeier, 1996; Serlin, 1987; Shaver, 1993; Sohn, 1993; Wampold, Davis & Good, 1990). So ermöglicht die mit kleinen Fallzahlen einhergehende niedrige Power der Test das Auffinden „signifikanter“ Ergebnisse erst bei sehr großen Effektstärken. Einen Ausweg aus diesem Dilemma weisen zum einen Power erhöhende Messwiederholungsdesigns und zum anderen stichprobensensitive und auf kleine Fallzahlen ausgelegte Randomisierungsverfahren.

Repräsentativitätsproblem. Innerhalb der Repräsentationsproblematik lassen sich zwei Facetten unterscheiden, die Stichprobenrepräsentativität und die Kontext-Repräsentativität. Die Stichprobenrepräsentativität bezieht sich auf Auswahlverzerrungen, d.h. auf die Frage, inwieweit sich die ausgewählte Stichprobe deutlich von der Population unterscheidet, die sie repräsentieren soll. Bei Anwendung dieses Sachverhalts auf die vorliegende Studie kann durchaus von einer für Psychologie Studierende repräsentativen Stichprobe ausgegangen werden. Eine mögliche Ausweitung des Geltungsbereichs auf andere Populationen erscheint durchaus plausibel, benötigt jedoch noch weitere Studien mit Variation von Sprecher-Trait und State Merkmalen zu ihrer Untermauerung. Des Weiteren ist die als Kontext-Repräsentativität bezeichnete Generalisierungsweite der Ergebnisse auf alternative Anwendungssituationen zu berücksichtigen. Um die hier vorliegende Kontext-Repräsentativität zu beurteilen, sind vielfältige Variationen der Aufnahmesituation und Nebengeräuschbedingungen erforderlich.

Um die Übertragbarkeit auf diverse Anwendungskontexte zu gewährleisten, müssten Aufgabenmerkmale (z.B. aktuelle Mental Workload Situation), Umfeldfaktoren (z.B. Hintergrund-Lautstärke), Sprechsituationen (informelle vs formale Sprechsituation, Human Computer Kommunikation), Sprechmodus (flüsternd, laut, hyperartikuliert), Sprechinhalt (Erweiterung des einbezogenen Vokabulars), Sprecherkalibrierungsmöglichkeiten (nur einmalige Sprachaufnahmen) systematisch variiert und in ihrer Wirkung auf die Prognosegüte der ASA bewertet werden. Aufgrund der artifiziellen Laborsituation und der eingeschränkten Sprechsituationen sind Generalisierungen der gefundenen Ergebnisse auf andere Sprechsituationen nur mit Vorsicht zu führen. Die getestete Stichprobe lässt nur bedingt Rückschlüsse über schläfrigkeitsbezogene Vorhersagegüte bei Sprechern unter 20 Jahren und über 40 Jahren zu. Darüber hinaus sind keine Probanden mit ausgeprägtem Dialekt oder nicht-Muttersprachler untersucht worden. Generalisierungen der gefundenen Ergebnisse auf andere Sprechergruppen können daher nur mit allergrößter Vorsicht unternommen werden. In der vorliegenden Untersuchung lagen bzgl. der oben genannten Kategorien folgende Spezifikationen vor: keine zusätzlichen Mental Workload induzierenden Aufgaben; keine Hintergrundgeräusche oder Nebensprecher; Monologsituation mit vorgegebenem Sprachinhalt und gespanntem Sprechen; wiederholte Sprachaufnahmen aller Sprecher.

Zusammenfassendes Urteil der Datenqualität. Die Analyse der Instrumenten- und Design-immanenten Fehlerquellen dokumentiert eine gute bis zufriedenstellende interne Validität der Untersuchung. Die angegebenen Detektionsraten und False Alarm Raten unterliegen somit keinen systematischen Verzerrungen. Lediglich die Übertragbarkeit der Ergebnisse auf andere Sprecher und Anwendungskontexte ist mit einer großen Unsicherheit verknüpft.

6.3 Forschungsdesiderate

Um die stimmbasierte Echtzeit-Detektion weiterzuentwickeln, sollten sich zukünftige Forschungsbemühungen primär (a) mit der Weiterentwicklung der Acoustic Sleepiness Analysis entlang des 5-stufigen Messprozederes, (b) mit weiteren Validierungsszenarien und (c) mit dem forschungsstrategischen Aspekt des Aufbaus einer leistungsfähigen Sprachanalyse-Infrastruktur (frei zugängliche Sprachanalysesoftware und Schläfrigkeitssprachdatenbanken) beschäftigen. Darüber hinaus könnten flankierend die Beforschung von kognitiv-phonetischen Mediatormechanismen, die auditiv-perzeptuelle Analyse sowie die Sprachsynthese-gestützte Analyse wertvolle Hinweise zur Identifikation weiterer akustischer Kenngrößen liefern.

Weiterentwicklung der Acoustic Sleepiness Analysis. Aufgabe zukünftiger Forschungsvorhaben ist die Optimierung des 5-stufigen Messprozesses der ASA, um eine leistungsfähigere Vorverarbeitung, Merkmalsberechnung, Dimensionalitätsreduktion und Klassifikation zu ermöglichen und damit die Gesamtleistung der ASA zu verbessern (siehe Abbildung 69). Als eingehendes Sprachmaterial sollten unterschiedliche Aufnahmequalitäten und Umgebungsgeräuschbedingungen, sowie vorverarbeitende Noise-Filter getestet werden. Hinsichtlich der Segmentierung sollten ausgesprochen schläfrigkeitssensitive phonetische Einheiten und Konsonantencluster (z.B. Glide-Konsonant [r], verschiedene Konsonant-Vokal-Konsonant Kombinationen auf diversen Satz- und Betonungspositionen) identifiziert werden. Die Phase der Merkmalsberechnung könnte um weitere Timing, Dauer, Pausen und Rhythmus Kennzahlen, sowie Kennzahlen der nicht-linearen Zeitreihenanalyse erweitert werden (Embedded Space Features: durchschnittliche Winkel oder Länge der Embedded Space Vektoren, Delay Vector Variance, Lyapunov Exponenten, Korrelationsdimension, Automutual Information, Time Resolved Density, Fractal Dimensions, Multiscale Entropies und Recurrence Quantification Analysis) (vgl. z.B. Golz et al., 2007). Des Weiteren kommt gerade bei zusätzlich erweiterten Merkmalssetgrößen den Dimensionalitätsreduktions-Verfahren eine besondere Bedeutung zu. (z.B. Sequentiell Backward Selection, Sequentiell Forward Floating Selection, Sequentiell Backward Floating Selection, LDA basierte Dimensionsreduktion). Weitere Machine-Learning Algorithmen und Mustererkennungsverfahren wie Maximum-Likelihood Bayes Classifier, Fuzzy Membership Indexing, Hidden Markov Models und Gaussian Mixture Density Models könnten zusätzlich in den Prozess der Klassifikation integriert werden. Ferner sollten neben Detektions- auch Prädiktionsalgorithmen gefunden werden, die eine Vorwarnzeit vor Erreichen eines kritischen Schläfrigkeitsszustands ermöglichen.

Darüber hinaus könnte eine multimodale Fusion schläfrigkeitsbezogener Messkanäle (z.B. behaviorale, pupillographische, EEG-basierte und Fahrzeugbedienungs-bezogene Indikatoren) sowie die Einbringung von aus Schläfrigkeitsmodellen geschätzten Apriori-Schläfrigkeitswahrscheinlichkeiten (z.B. mittels zirkadianem Rhythmustyp, Schlafdauer, Schlafqualität, Time-since-Sleep und Time-on-Task Informationen), einen signifikanten Zugewinn an Erkennungssicherheit bieten.

Eine besondere Herausforderung wird es sein, die Stimmmerkmalsanalyse kontextbezogen durchzuführen und dabei Variationen des Sprecherzustands (z.B. sitzend, stehend, liegend; mit Kopfschmerzen, Milch getrunken, geraucht; nervös, gestresst, frustriert; gelangweilt, demotiviert oder traurig; Verfälschungsmotivation wacher oder müder zu erscheinen), Variationen der Sprechereigenschaften (Alter, Geschlecht, Muttersprache, Dialekt), Variationen der parallel ausgeführten Aufgabenanforderungen (aktuelle Mental Workload Situation), Variationen des Störgeräuschniveaus (simultane Fremdsprecher, Motorengeräusche, Hall in kleinen Räumen, Aufnahme über Freisprecheinrichtung, Funkverkehr- und Telefonqualität) und Variationen der Sprechsituationen (klarheits- lautstärke- und dialogpartnerbezogene Sprechmodusvariationen, wie informelle vs. formale Sprechsituation, Human Computer Kommunikation) zu berücksichtigen. Diese Variationen könnten die Detektionsrate und die False Alarm Fehlerrate der Schläfrigkeitsmessung beeinflussen. Daher sollte es das Ziel sein, die Variationen, soweit sie in der Anwendungssituation hinreichend prävalent sind, zu berücksichtigen.

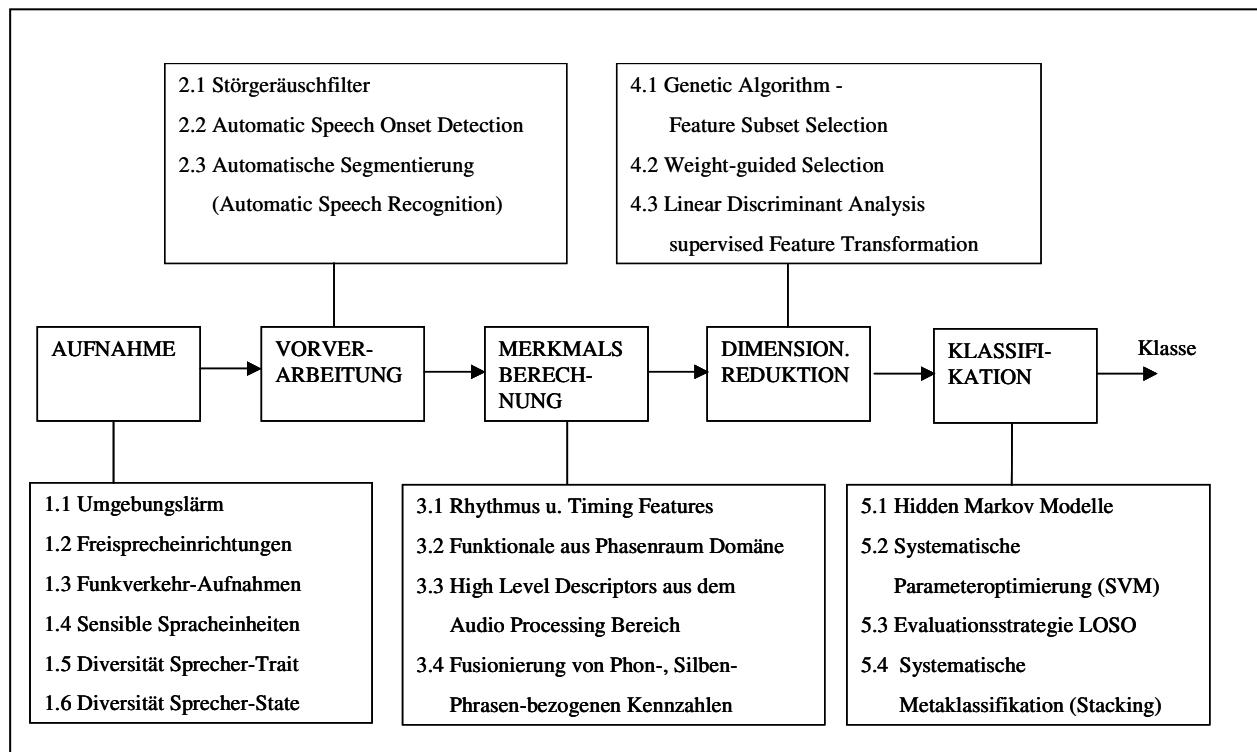


Abbildung 69: Mögliche Weiterentwicklungsperspektiven der akustischen Schläfrigkeitsmessung

Alternative Validierungsszenarien. Um eine größere Menge eindeutig schläfriger Stimmmaterialien zu bekommen, sollte ein zukünftiges Untersuchungsdesign die zirkadian gesteuerte Schlaf-Wach-Rhythmik berücksichtigen und gegebenenfalls den Chronotyp der Probanden bei der Auswahl der Stichprobe einbeziehen. Möglich wäre auch, dass in der vorliegenden Untersuchung die Arousal unterstützende DVD-Bedingung dazu geführt hat, dass die Schläfrigkeit der Probanden unterdrückt wurde. In einer follow-up Studie könnte die Zeit der Schlafdeprivation verlängert und eine strengere Monotoniesituation hergestellt werden.

Zukünftige Untersuchungen sollten es sich zum Ziel machen, die Aussagekraft der Validierungsergebnisse weiter zu erhöhen, indem instrumenten- und designbezogene Studienvariationen vorgenommen werden. Ein zukünftiges Untersuchungsdesign könnte randomisierte Kontrollgruppen einbeziehen und zur Absicherung eine Übereinstimmung diverser Validierungsanker verlangen (z.B. MSLT, Unstable Tracking Test oder PST, sowie EOG oder EEG basierte Verfahren, wie dem Alpha-Attenuation Test). Denkbar ist auch eine Verschiebung des Messziels in Richtung Prognose von Mikroschlafepisoden. Darüber hinaus kann die wie-

derholte Messung von Probanden in mehreren Nächten einen Robustheitsgewinn bringen, indem unsystematische Sprecherzustands-Variationen einer Messnacht über den Verlauf mehrerer Messnächte gemittelt werden (siehe auch Validierungsansätze Leave-One-Speaker-Out, Leave-one-Session-Out).

Forschungsstrategische Aspekte: Akquisition schläfrigkeitsbezogener Daten. Dem Aufbau einer forschungsförderlichen Infrastruktur wird zukünftig wachsende Bedeutung zukommen. Dazu zählen frei zugängliche Programm-Skripte für die Bestimmung von Sprachkennzahlen und frei zugängliche Sprachdatenbanken für Befindlichkeitszustände. Im Weiteren werden daher Anforderungen und Gütemerkmale einer schläfrigkeitsbezogenen Sprachdatenbank aufgelistet (vgl. auch Schuller, 2006):

- *Validität der Validierungsanker.* Lediglich eindeutige Sprachbeispiele, die durch Übereinstimmung von Selbst- und Fremdeinschätzungen (evtl. ergänzt durch behaviorale und physiologische Daten) gekennzeichnet sind, sollten herangezogen werden. Denkbar ist auch eine Erweiterung um gespielte Schläfrigkeitsbeispiele.
- *Perzeptionstests.* Um eine Vergleichsbasis mit der menschlichen Erkennensleistung zu ermöglichen, können mit repräsentativen Testpersonen Schläfrigkeitsperzeptionstest durchgeführt werden.
- *Granularität der Kategorisierung.* Die Annotation von Daten, beziehungsweise ihre Zuordnung zu Klassen ist ein essentieller Bestandteil einer Datenbank. Verschiedene Granularitätsstufen, d.h. Schläfrigkeitsklasseneinteilungen (z.B. zwei, drei oder zehn Schläfrigkeitsklassen) sollten, um unterschiedliche Anwendungen möglich zu machen, bereitgestellt werden.
- *Ideale Erfassungsbedingungen.* Weitere Konfundierungen z.B. durch Umgebungslärm sollten je nach gefordertem Realismus gewählt werden. Unter Berücksichtigung des noch frühen Forschungsstadiums sollten möglichst störbedingungsfreie Settings (keine simultanen Fremdsprecher, wenig Umgebungslärm) gewählt werden. Nicht zu vermeidende Störeinflüsse könnten zumindest in separaten Datensets erfasst werden.
- *Präzisierung der Aufnahmesituation.* Die genaue Beschreibung der Situationscharakteristika und Randbedingungen der Aufnahmen ermöglicht erst die Einschätzung des Übertragbarkeitsbereichs der erzielten Ergebnisse.
- *Diversität.* Wünschenswert ist eine hohe Diversität bezüglich veränderlicher Personenzustände (momentane Befindlichkeit, gesprochene Sprache) und stabiler Personeneigenschaften

(Muttersprache, Altersklassen, Geschlecht, Bildungsgrad). Diese Sprachdaten sollten über die gewöhnlichen Standards hinaus auch die genannten Zustands-Kombinationen beinhalten (wie z.B. ein müder Sprecher, der zugleich ängstlich, erkältet, gestresst ist, Alkohol/ Milch getrunken hat oder unter Schmerzeinfluss steht). Diese Diversität ermöglicht es eine auf die jeweilige Anwendungssituation zugeschnittene Auswahl der Sprachproben zu treffen.

- *Hohe Gesamtanzahl und Gleichverteilung.* In der Regel ist mit einer hohen Gesamtzahl an Lern- und Testbeispielen auch eine höhere Gesamtleistung eines Erkennungssystems auf Grund exakterer Modellierung der Problemstellung zu erwarten. Eine gleichmäßige Verteilung der Schläfrigkeitskategorien wiederum, vermeidet die Gefahr der Bevorzugung einer überrepräsentierten Klasse. Für die selteneren schläfrigen Sprachbeispiele werden ansonsten nur geringe Detektionsraten erzielt.

- *Zugänglichkeit.* Um internationale Vergleichbarkeit und Reproduzierbarkeit zu gewährleisten, ist die freie Zugänglichkeit der Daten von zentraler Bedeutung.

6.4 Anwendungsbezogenes Transferpotenzial

Entwicklung von stimmbasierten psychologischen Messverfahren auf Grundlage des akustischen Schläfrigkeitsmess-Prozederes

Die Methodik der akustischen Schläfrigkeitsmessung bietet ein reichhaltiges Transferpotenzial. So eröffnet die akustische Echtzeit-Erfassung von internen Zuständen eine Reihe von vielversprechenden Anwendungsfeldern. Viele Alltags-Anwendungskontexte erfordern einen den Tätigkeitsvollzug nicht beeinträchtigenden, belästigungsarmen Messzugang. Die Implementierbarkeit der Stimmessung in diese Alltagskontexte unterstreicht daher zusätzlich die potenzielle Bedeutung stimmbasierter psychologischer Messinstrumente, die in den folgenden Anwendungsfeldern (siehe Abbildung 70) in realistischer Reichweite liegt:

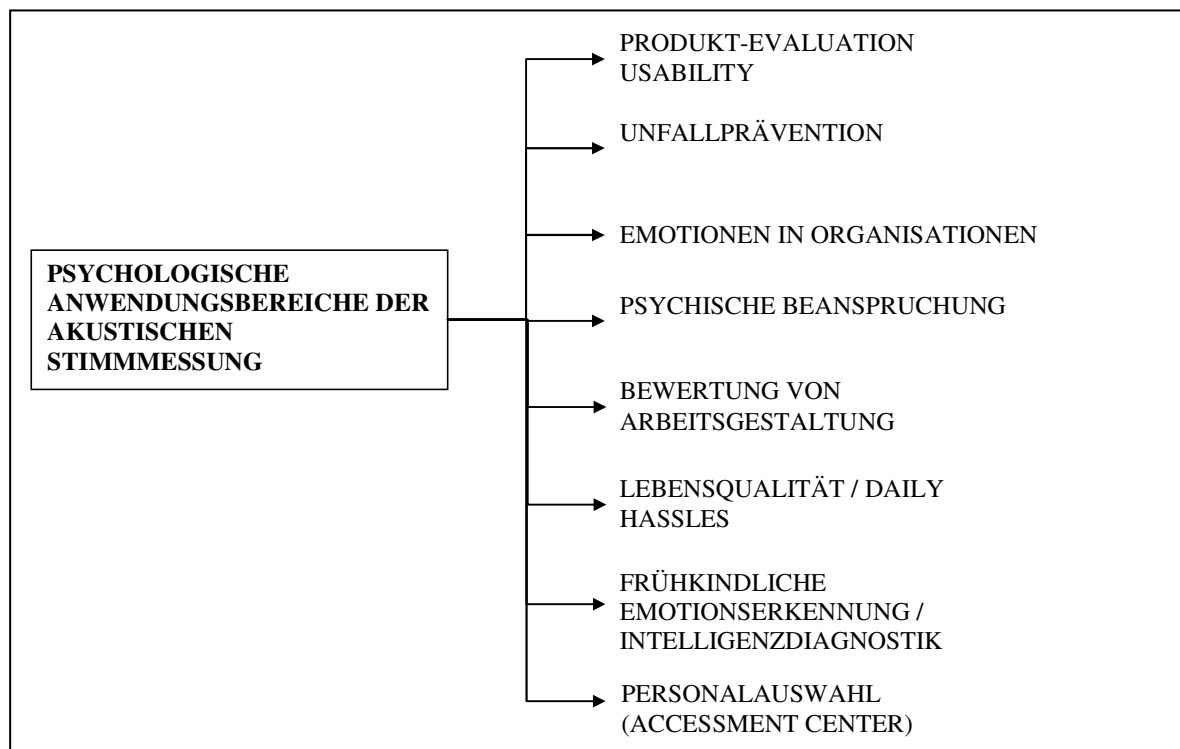


Abbildung 70: Vielfalt von potenziellen psychologischen Anwendungssituationen der akustischen Stimmessung

In *Markt- und Werbewirkungsforschung* könnte die berichtete Produktbewertung objektiviert werden. Eine entsprechende Produkt-Evaluation, die Erfassung des Nutzererlebens (z.B. Angenehmheit und Aktivierung) wäre auch unter widrigen („feuchten“) situativen Rahmenbedingungen möglich (z.B. Schaumbäder-Bewertung). In Kaufsituationen könnte die automatische Erfassung von internen Zuständen einerseits das Wohlfühl und die Zufriedenheit des

Kunden mit einem Produkt, und andererseits die Güte der Dienstleistungs- und Servicequalität auf Verkäuferseite bestimmen helfen (z.B. Leistungsbewertung von Call Center Agents).

Im Anwendungsfeld der *Mensch-Maschine-Interaktion* (Human-Computer-Interaction) und *Emotional Ergonomics* (Wieland, 2002) stellt sich die Aufgabe der Akzeptanzsicherung von maschineller Kommunikation. Die Integration sozial-intelligenter Kommunikation kann über die Erkennung des Benutzerzustands die kommunikative Natürlichkeit und somit den Komfort erhöhen, indem eine der momentanen Nutzer-Stimmung angepasste „empathische“ Sprachausgabe erfolgt (Giles et al., 1987; Wieland & Kappes, 2001). Darüber hinaus kann das Ziel der Effizienzsteigerung von Mensch-Maschine-Interaktion erreicht werden, indem im Falle eines irritierten Benutzers Hilfe angeboten wird oder im Fall eines verärgerten oder belästigten Nutzers (Ang et al., 2002) Maßnahmen zur Fehlerauflösung eingeleitet werden. In Edutainment-Programmen kann die Detektion von Langeweile oder Stress für die Anpassung der Lehrgeschwindigkeit genutzt werden. Zusätzliche Anwendungsfelder ergeben sich im Bereich der Unterhaltungselektronik, wie der Suche nach emotionsspezifischen Film- und Musikstücken in multimedialen Datenbanken oder im Bereich emotions-sensitiv reagierender Videospiele und Roboter-Spielzeuge.

Ähnliches gilt für die Bewertung des Nutzererlebens im Bereich der *Usability* und *Software-Ergonomie*. Die sprachbasierte Emotionserkennung macht die Erfassung von z.B. Empathie in Kundengesprächen oder aber auch die Detektion von aggressiv-feindseligen und hochverängstigten Stimmungslagen im öffentlichen Raum möglich (Kriminalitätsprävention). Die Bestimmung der psychischen Beanspruchung oder der Aktivierungs- und Befindlichkeitslage von Mitarbeitern ist bei der Bewertung des Gestaltungsbedarfs sinnvoll. Darüber hinaus könnte auch der Erfolg *arbeitsgestalterischer Bemühungen* objektiv bewertbar werden (z.B. Wirkung einer pausengestalterischen Intervention). Auch in klinisch-therapeutischen Anwendungskontexten kann die stimmliche Erfassung von z.B. Ängsten, Traurigkeit, Selbstsicherheit oder Wut im Bereich der *Diagnose, Therapieunterstützung, Erfolgskontrolle und Rückfallprophylaxe* wichtige Hilfestellung liefern. So könnte ein Echtzeit-Selbstmonitoring Frühindikatoren sich anbahnender kritischer Zustände rechtzeitig detektieren und Warnungen ausgeben.

Innerhalb von *eignungsdiagnostischen Beratungs- oder Auswahl-situationen* könnte eine aus State-Variablen abgeleitete, effiziente und vorurteilsfreie Trait-Messung von eignungsdiagnostisch relevanten Persönlichkeitsmerkmalen (z.B. Gewissenhaftigkeit, soziale Verträglichkeit, Integrität, Belastungssouveränität, Macht- und Leistungsmotiv) stattfinden (vgl. Wieland, Braun, Hüttges, Timm & Willamowski, 2002). Neben Persönlichkeitsmerkmalen ist aber auch die Diagnostik von Kompetenzen und kognitiven Fähigkeiten - wie z.B. die *Hochbegabendiagnostik* mittels frühkindlicher Lautäußerungen („intelligentes Lallen“) - denkbar. Nicht zuletzt ergäbe sich für psychologische Feldstudien die Möglichkeit, in ökologisch validen Settings, belästigungsfrei und über einen längeren Untersuchungszeitraum Daten zu erheben (vgl. auch Fahrenberg, Myrtek, Pawlik & Perrez, 2007).

Weitere Transferfelder, die von der hier dargestellten Messmethodik (insbesondere die Machine-Learning Phasen der Merkmalsberechnung, Dimensionalitätsreduktion und Klassifikation) profitieren könnten, sind neben medizinischen Fragestellungen, u.a. auch forensische, zoologische und seismologische (siehe Abbildung 71). Das hier vorgestellte Machine-Learning Procedere kann darüber hinaus allgemeine Prognose- und Klassifikationsfragestellungen z.B. im Bereich der psychologischen Modellbildung, der physiologischen Messinstrumententwicklung oder der finanzwirtschaftlichen Wertpapier-Kursprognose bereichern (vgl. Abschnitt 1.3).

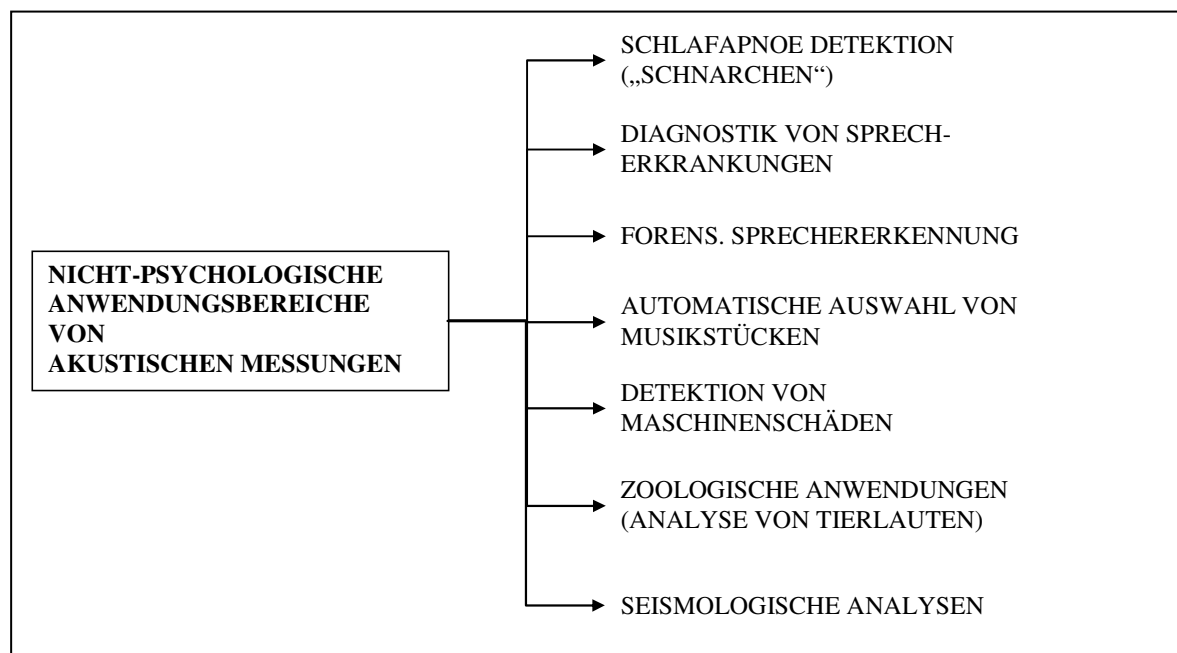


Abbildung 71: Vielfalt potenzieller nicht-psychologischer Anwendungsfelder der hier dargestellten 5-stufigen akustischen Messmethodik

Integration von Schläfrigkeits-Detektions-Systemen in die übergeordnete Architektur von zustandsadaptiven Unfallrisiko-Management-Systemen. Eine wichtige Aufgabe zukünftiger Forschungsbemühungen sollte die Integration von Schläfrigkeitswarn-Systemen in eine umfassende Detektion und Prädiktion sicherheitskritischer Zustände innerhalb von z.B. Überwachungstätigkeiten des Straßen-, Schienen-, Schiffs- und Luftverkehrs (Piloten, Fluglotsen, Lkw-Fahrer, aber auch Kraftwerkoperator und Kranführer-Tätigkeiten etc.) darstellen. Die integrative Architektur eines solchen Unfallrisiko-Management-Systems besteht grundsätzlich aus einem Detektions- und einem Interventionsmodul (siehe Abbildung 72). Im Sinne eines umfassenden Monitorings sicherheitskritischer Zustände kommt einerseits der Erfassung von als Nebenaufgaben wirksam werdenden, Zusatzregulationsaufwand erfordernden Distraktor-Zuständen, wie innere Anspannung, Wut, Schmerz und Angst eine wichtige Bedeutung zu (vgl. auch Handlungs-/ und Selbstregulationsmodell; Wieland-Eckelmann, 1994). Andererseits ist die Detektion aufmerksamkeitsreduzierter Zustände wie Schläfrigkeit und Monotoniezustände (Tagträume, Driving without Awareness, Highway Hypnosis) zu überwachen.

Ansatzpunkte für eine Reaktion auf einen detektierten oder prädiktierten Distraktor-Zustand sind (a) die Beseitigung des Distraktor-Zustand selbst, (b) das Vermeiden weiterer fahrzeugbedienungsbezogener Distraktoren sowie (c) die Reduktion fahraufgabenbezogener Anforderungen. Die Reduktion eines identifizierten Distraktor-Zustands kann im Sinne der internalen und externalen Selbstregulation² (vgl. modell der kognitiv-aktionale Handlungs- und Selbstregulation; Wieland & Baggen, 1999) erfolgen. Die Grenzen zwischen internaler und externaler Selbstregulation sind hier fließend, so kann z.B. bei identifizierter ängstlicher Anspannung das System Hilfestellung zur Selbstregulation liefern, indem es beruhigende Selbstinstruktionen vorschlägt (internale Selbstregulation; vgl. Wieland-Eckelmann, 1992).

² Internale Selbstregulation (emotionsbezogene Regulation) beschreibt kognitive und nervöse Regulationsvorgänge, die darauf gerichtet sind, interne Zustandsgrößen (z.B. innere Anspannungen, Gedanken der Besorgtheit, körperliche Erregung) so zu verändern, dass sie mit gewünschten Soll-Zuständen übereinstimmen. Art und Ausmaß dieser Form der Emotionsregulierung steht dabei in engem Zusammenhang zu intrapsychischen habituellen Bewältigungsstilen (Wieland-Eckelmann, Bösel & Badorrek, 1989; Wieland-Eckelmann, 1992, 1996). Von den Vorgängen der internalen Selbstregulation lassen sich Formen der internalen Regulation abgrenzen. Eine zentrale Funktion dieser Regulationsform besteht darin, z.B. über eine vergrößerte Anstrengung (im Sinne von Task-Effort) den Grad der für die Aufgabenbewältigung erforderlichen energetischen Aktiviertheits-Zustände, den Aufgabenerfordernissen anzupassen (vgl. auch „Control Model of State Regulation“; Hockey, 1986).

Des Weiteren kann im Sinne der externalen Selbstregulation eine Distraktor-Zustandsveränderung über das Einspielen von entspannender Musik oder über an die Emotion des Fahrzeugführers angepasste Sprachausgaben (Nass et al., 2005) bewirkt werden.

Das Vermeiden weiterer fahrzeugbedienungsbezogener Distraktoren kann mithilfe von adaptiven Fahrerassistenzsystemen erreicht werden. Diese können den Fahrer vor weiteren zusätzlich Verarbeitungskapazität bindenden, Zusatzaufgaben abschirmen. Das Abstellen der Infotainmentsysteme (z.B. Radio herunterregeln, Telefonate nicht mehr durchstellen) gehört in diese Maßnahmenkategorie ebenso wie das Vereinfachen und Fokussieren der Cockpit Displays und Bedienelemente. Die Reduktion der direkten aufgabenbezogenen Fahranforderungen -- also die Vereinfachung der Fahraufgabe -- kann durch Unterstützungs- und Substitutionssysteme teilweise bis vollständig erzielt werden. Mögliche wäre beispielsweise eine Steigerung der Sensitivität von diversen Warn- und Reaktionssystemen des Auffahrschutzes (z.B. Collision Warning, Road Departure Warning), automatische Begrenzung der Höchstgeschwindigkeit oder der vollständigen Übernahme von Fahraufgaben (z.B. Automatic Cruise Control, Automatic Lateral Control).

Neben der Detektion von Distraktor-Zuständen ist das Monitoring aufmerksamkeitsreduzierter Zustände von sicherheitsbezogener Relevanz. Im Fall einer Detektion ressourcenreduzierter Zustände wie Schläfrigkeit, Alkoholisiertheit und Monotoniezustände kommt das Interventionsmodul zum Einsatz. Zu den Reaktionsmöglichkeiten zählen die oben bereits genannten Ansätze des Vermeidens weiterer fahrzeugbedienungsbezogener Distraktoren und der Reduktion fahraufgabenbezogener Anforderungen. Darüber hinaus ist eine direkte Verbesserung der aufmerksamkeitsreduzierten Zustände denkbar (vgl. auch internale und externe Regulation; Wieland-Eckelmann, 1992). Im Fall der Schläfrigkeitsdetektion wäre bei einer identifizierten leichten Schläfrigkeit eine Warnmeldung sinnvoll, da die Bewusstmachung des sicherheitskritischen Schläfrigkeitszustandes eine anstrengungs-getriggerte Mobilisierung von Reservekapazitäten provozieren kann. Andere Interventionsmöglichkeiten eines Schläfrigkeitssystems sind die stimulierende Modifikation des Fahrzeuginnenwelt im Sinne einer Aktivierung des Fahrers beispielsweise über ätherische Duftstoffe, Lenkrad- und Sitzvibrationen, reduzierte Innenraumtemperaturen, aktivierende Komfortreduktionen, gezielte Lichtbestrahlungen, verstärkte Sauerstoffzufuhren oder das Einspielen beschwingter Musik. Grundsätzlich sollten Schläfrigkeits-Gegenmaßnahmen adaptiv an den Intensitätsgrad der Schläfrigkeit angepasst werden. Eine hilfreiche Unterscheidung führt Sagberg et al. (2004)

mit den vier Stufen, Preserving Phase, Low Stimulation Phase, High Stimulation Phase und Protection Phase ein. Für weiterreichende Betrachtungen zur Implementierung und assoziierten Problematiken von Hypovigilanz Systemen siehe Hagenmeyer (2007).

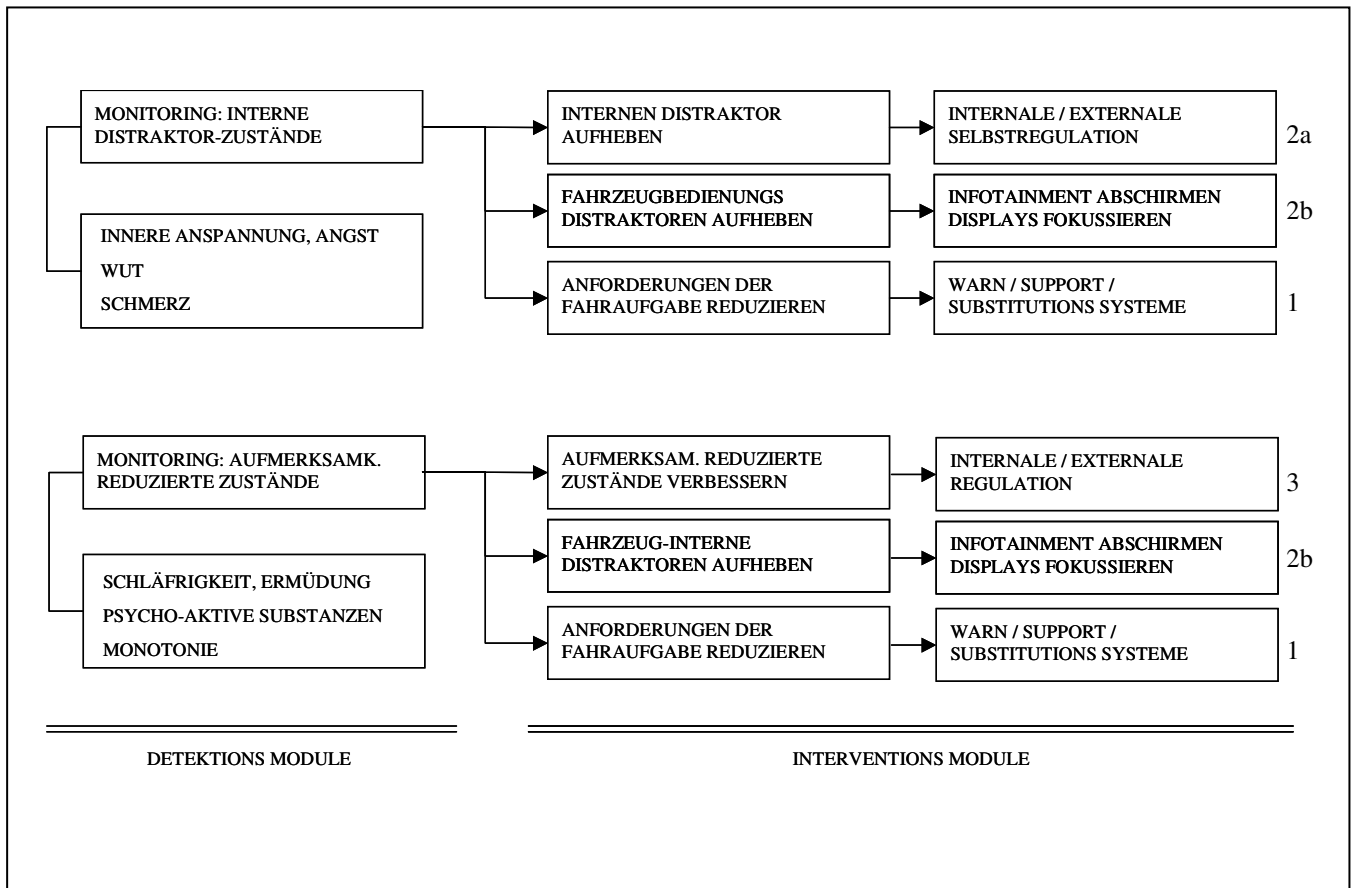


Abbildung 72: Integration von Schläfrigkeitssystemen in die erweiterte Architektur eines zustandsadaptiven Unfallrisiko-Management-Systems. Dargestellt sind die Detektionsmodule „Interne Distraktor-Zustände“ und „Aufmerksamkeitsreduzierte Zustände“ sowie nach (1) Hauptaufgabenanforderungen, (2) Zusatzanforderungen und (3) Ressourcen geordnete Interventionsansatzpunkte

7 ZUSAMMENFASSUNG

Ziel der vorliegenden Arbeit war es, im Rahmen einer experimentellen Pilotstudie das stimmanalytische Verfahren der *Acoustic Sleepiness Analysis* (ASA) zur Diagnostik von Schläfrigkeit zu validieren. Der Untersuchung lag die Annahme zugrunde, dass es schläfrigkeitsinduziert zu einer Veränderung auditiv-perzeptiver Stimmmerkmale und akustischer Kennwerte aus den Bereichen Prosodie, Artikulation und Stimmqualität kommt. Es wurde daher mithilfe eines Machine-Learning Prozederes ermittelt, inwieweit sich Schläfrigkeitszustände aus 340 akustischen Kennwerten vorhersagen lassen.

Die Ergebnisse aktueller Studien dokumentieren, dass bei ungefähr 20% aller Verkehrsunfälle Schläfrigkeit eine wesentliche Unfallursache darstellt. So ereignet sich die Mehrzahl der schweren Straßenverkehrsunfälle nach längeren Wachzeiten, mangelndem Nachtschlaf oder in den frühen Morgenstunden (vgl. Abschnitt 1.2). Nächtliches Autofahren, insbesondere in den frühen Morgenstunden, ist besonders kritisch, da zirkadian gesteuert die Schläfrigkeit hier einen Maximalwert erreicht und sich die kognitive Leistungsfähigkeit auf einem Tiefpunkt befindet. Aus diesem Grund ist die Erfassung von kritischen Schläfrigkeitszuständen sowie die Prognose von Mikroschlafepisoden eine vordringliche Aufgabe der Unfallprävention.

Zur Erfassung von Schläfrigkeitszuständen werden derzeit Selbsteinschätzungsverfahren (Stanford Sleepiness Scale, Karolinska Sleepiness Scale, Samn-Perelli Müdigkeitsskala Visuelle Analog Skalen), kognitive Leistungstests (Psychomotorischer Vigilanz Test, Unstable Tracking Test), behaviorale Verfahren (Wierwille Rating, Observer Drowsiness Test), Blick- und Pupillen-basierte Verfahren (Pupillographischer Schläfrigkeitstest, Perclos, Sakkadenbewegungen), elektrophysiologische Messverfahren (EDA), EEG-basierte Verfahren (Karolinska Drowsiness Scoring, Alpha Attenuation Test) und Einschlaf latenz-basierte Verfahren (Multipler Schlaflatenztest, Maintenance Wakefulness Test) eingesetzt.

Wichtige Kriterien zur Bewertung von Schläfrigkeitsmessverfahren sind (a) die Validität (auch Sensitivität und Spezifität), (b) die Ökonomie in Messvorbereitung, -betreuung u. –auswertung (bzgl. des zeitlichen, personellen und finanziellen Aufwands), (c) die messvorgangsbezogene Belästigung des Messobjekts (durch Trainingsphasen zur individuellen Baseline-Bestimmung, durch Re-Kalibrierungsphasen vor jedem Messbeginn oder durch Sensor-

applikationen am Körper) und (d) die Entwicklungsmöglichkeiten des Messansatzes in Richtung einer automatischen Echtzeit-Erfassung von Schläfrigkeit in realen Messanwendungssituationen. Anders als bei Selbsteinschätzungsverfahren und kognitiven Leistungsverfahren ermöglicht der akustische Stimmanalyseansatz eine beliebig wiederholbare, den Tätigkeitsvollzug nicht beeinträchtigende und verfälschungsresistente Erfassung von Schläfrigkeit. Im Gegensatz zu physiologischen Ansätzen bietet der akustische Stimmanalyseansatz, die Vorteile eines berührungs- und sensorapplikationsfreien, ökonomischen sowie belästigungs- und kalibrierungsarmen Messzugangs. Der laufende Tätigkeitsvollzug wird nicht beeinträchtigt, da eine „Hands-free“ und „Eyes-free“ Messsituation vorliegt (vgl. Abschnitt 1.5). Die ASA liefert somit im Vergleich zu bisherigen Messansätzen einen eigenständigen Mehrwert und stellt vor diesem Hintergrund einen vielversprechenden neuen Messansatz für die Echtzeit-Detektion von Schläfrigkeit dar.

Zur Konstruktion der akustischen Schläfrigkeitsmessung leisten in einem interdisziplinären Forschungsfeld folgende Wissenschaftsdisziplinen einen signifikanten Beitrag: Linguistik, Phonetik, Phoniatrie, Schlafmedizin, Arbeitspsychologie, Signal- und Audio Processing, Pattern Recognition sowie Künstliche Intelligenzforschung (vgl. Kapitel 4). Bezüglich der linguistischen und phonetischen Grundlagen sind von speziellen Interesse, sowohl die kognitiven (vgl. Abschnitt 2.2), als auch die phonetischen Aspekte der Sprachproduktion wie das dreistufige Sprachproduktionsmodell von Atmung, Phonation und Artikulation. Dargestellt werden diese Modelle der Sprachproduktion, da sie ein Bindeglied zwischen schlaftrigkeitsinduzierter physiologischer Veränderung und resultierenden stimmlichen Veränderungen darstellen.

Um denkbare stimmliche Veränderungen abbilden zu können, werden im Abschnitt 2.3 wichtige auditiv-perzeptive Beschreibungskategorien, wie Prosodie, Artikulation und Stimmqualität diskutiert. Unter Prosodie werden hier Tonhöhe, Intonationsverläufe, Lautstärke, Tempo, Dauer, Pausen und Rhythmus-Phänomene subsumiert. Zur Kategorie der Artikulationsbezogenen Phänomene zählen Assimilation, Lenisierung, Zentralisierung, Reduktion und Verschleifung. Die Stimmqualitätskategorie beinhaltet stimmliche Beschreibungsgrößen wie Stimmklang, Stimmfülle, Behauchtheit und Nasalität. Quantifiziert werden kann ein Teilbereich dieser Phänomene mittels zuordenbarer Signalverarbeitungskennzahlen (z.B. Bestimmung der Tonhöhe über die Fundamental Frequenz der Wellenform). Wichtige akustische Kennzahlen sind in diesem Zusammenhang Intensität, Sprechrage, Formantfrequenzen (F1-

F5) und die ihnen zugeordneten Formantbandwidth, Jitter, Shimmer, Spektraler Slope, Harmonics-to-Noise Ratio und Mel Frequency Cepstral Coefficients (vgl. Abschnitt 2.4).

Das Kognitiv-phonetische Mediatormodell verknüpft die oben genannten Ebenen der (a) schläfrigkeitsinduzierten kognitiv-physiologischen Veränderungen, der (b) Sprachproduktionsmodelle, sowie der (c) auditiv-perzeptiven und akustischen Stimmmerkmale zu einem Erklärungsmodell schläfrigkeitsinduzierter Stimmveränderungen. Von zentraler Bedeutung für die Sprachproduktion sind schläfrigkeitsassoziierte kognitive Veränderungen der reduzierten selektiven Aufmerksamkeit, des beeinträchtigten Gedächtnisabrufs und der verlangsamten kognitiven Verarbeitungsgeschwindigkeit. Aus physiologischer Perspektive sind eine generelle muskuläre Relaxation und eine verringerte Körperkerntemperatur Auslöser der wesentlichen schläfrigkeitsinduzierten Veränderungen der Sprachproduktionssysteme (z.B. eine reduzierte Atemfrequenz- und -tiefe, verringerte Stimmlippenspannung, vergrößerte Elastizität der Vokaltraktwände, reduzierte Wärmeabstrahlung in den Vokaltrakt, eingefallener gekrümmter Vokaltrakt, abgesenktes Velum, veränderte Bewegungsmuster der Artikulatoren Lippen, Kiefer und Zunge). Abgeleitet aus den Veränderungen des Sprachproduktionssystems werden eine Reihe spezifischer Veränderungen des auditiv-perzeptiven Stimmindrucks, wie etwa ein verlangsamtes Sprechen, häufige Sprechpausen, diverse Sprechfehler oder eine unflüssige prosodische Phrasierung erwartet. Aufgrund vielfältiger bislang ungeklärter Interaktionsprozesse können auf der Ebene der akustischen Kennwerte nur grobe Vorhersagen zu schläfrigkeitsbedingten Veränderungen getroffen werden (z.B. eine verringerte Intensität, eine reduzierte Fundamental Frequenz oder eine Verschiebung der Formant-Positionen).

Neben diesen theoretisch abgeleiteten Annahmen liefern zahlreiche empirische Befunde aus dem Bereich der sprachlichen Emotionserkennung wichtige Hinweise zu schläfrigkeitsinduzierten Veränderungen akustischer Kennwerte. Empirische Befunde einer geringen emotionsbezogenen Aktivierungsdimension beinhalten: eine niedrige Fundamental Frequenz, geringen Fundamental Frequenz Range, geringe Intensität, geringer Intensität Range, geringe Sprechgeschwindigkeit, lange Sprechpausen, tiefe Frequenzpositionen des ersten Formanten (F_1), schmaler Bandwidth von F_1 , sowie geringe Energie des Amplitudenspektrums über 500Hz (vgl. Abschnitt 3.2).

Den wichtigsten Neuigkeitswert der akustischen Schläfrigkeitsanalyse stellt die Implementierung des Messablaufs in den Mustererkennungsprozess des Machine-Learning dar. Dieses Machine-Learning Vorgehen der Acoustic Sleepiness Analysis untergliedert sich in 5 Verarbeitungsstufen: (a) *Sprachaufnahme*, (b) *Vorverarbeitung*, (c) *Merkmalsberechnung*, (d) *Dimensionalitätsreduktion* und (e) *Klassifikation*. Im Anschluss an die Sprachaufnahme folgt eine filternde und segmentierende Vorverarbeitung. Eine darüber hinaus gehende Innovation der ASA stellt die verbesserte Nutzung der Rohsignalinformationen über die Generierung großer redundanter Merkmalssets dar. Der operative Teil der Merkmalsberechnung erfolgt über die Fensterung des Rohsignals, die Berechnung von fensterbezogenen Einzelkennwerten, die Fusionierung der fensterbezogenen Einzelkennwerten zu geglätteten Merkmalskonturverläufen und die Berechnung von Konturverlauf beschreibenden Funktionalen (High-Level Features). Zusätzlich werden akustische Merkmale über die automatische Generierung evolutionärer Algorithmen (Genetic Algorithms) bestimmt.

In der Mustererkennung spielt die Dimensionalität des Eingangsvektors eine wichtige Rolle für die Verbesserung der Gesamtvorhersageleistung, die Identifikation relevanter Stimmkennwerte und die Reduktion von Rechenzeit. Differenziert werden kann die Dimensionalitätsreduktion in (a) *Feature Subset Selection* und (b) *Feature Transformation* Methoden. Die Feature Subset Selection schließt wenig brauchbare Kennzahlen vollständig aus. Die Feature Transformation (z.B. Principle Component Analysis, Linear Discriminant Analysis) hingegen versucht unter Beibehaltung der ursprünglichen Kennzahlenvielfalt, die ursprünglichen Kennzahlen lediglich in einen niedriger dimensional Raum zu projizieren.

Das Ziel des wichtigen Phasenabschnitts Klassifikation ist es, die Vorhersageleistung einzelner Stimmindikatoren durch simultane Bündelung ihres Informationsgehaltes mittels Fusionierungsalgorithmen zu erhöhen. Die Verfahren der automatischen Mustererkennung ordnen nach einer Lernphase unbekannte Eingabemuster einer bereits bekannten Klasse zu (z.B. über Lineare Diskriminanzanalysen). Anstelle einer direkten Klassifikation der Waveform des Stimmsignals (die Klassifikation der Roh-Waveform selbst wäre aufgrund der großen Anzahl an Abtastwerten unmöglich), werden die spezifischen Informationen des Wellenmusters zu Stimmkennzahlen verdichtet und in Form eines Merkmalsvektors zusammengefasst. Zur Lösung von Klassifikationsproblemen stehen eine Vielzahl von Klassifikationsverfahren zur Verfügung. Einige der am häufigsten zur akustischen Emotionserkennung eingesetzten Klassifikationsverfahren sind: Lineare Diskriminanzanalyse, Künstliche Neuronale Netze Sup-

port-Vektor-Maschinen und K-Nearest-Neighbour Verfahren. Verbessert werden kann die Klassifikationsleistung durch eine Kombination der Klassifikationsverfahren in multiplen Instanzen (Ensemble-Klassifikation). Die Klassifikationsergebnisse der Level-0-Klassifikatoren werden in einem nachfolgenden Schritt mittels einer finalen Instanz, dem übergeordneten Level-1-Klassifikator verarbeitet und zu endgültigen Klassenzuordnungsentscheidungen kombiniert (vgl. Bagging, Boosting, Stacking).

Zur Beurteilung der Validität des ASA Verfahrens wurden in einem experimentellen Within-Subject Schlafdeprivationsdesign 24 Studenten untersucht. Als Probandeninklusionskriterien fungierten unauffällige Schlafmuster, Alkohol-, Nikotin- und Koffein-Karenz am Untersuchungstag. In der Zeit von 20.00h bis 04.00h wurde im Stundenrhythmus Sprachmaterial digital aufgezeichnet (22.05 kHz; 16 Bit; mono) und die selbst- sowie fremdberichtete Schläfrigkeit erfasst (KSS). Die zwei jeweils wachsten und schläfrigesten Messungen jedes Probanden gingen in die weitere Analyse ein. Als Sprachmaterial kamen zum Einsatz: die Vokalphonation eines 3 Sekunden lang gehaltenen [a:], eine Probandenvorstellungs-Introphrase, eine Bildbeschreibung sowie 9 Fahrerassistenzsystem-Sätze, wie sie im Kontext von „In-Car-Speech-Dialog-Systemen“ denkbar wären (z.B. „Suchen Sie die nächste Raststätte heraus“ oder „Rufen Sie den Notdienst an“). Kategorisiert wurden die Sprachmaterialien für die weitere Analyse in die Sprachsegmenteinheiten: Vokal, Wort, Phrasale Einheit und Multiple Phrasale Einheit.

Die sich anschließende Merkmalsberechnung wurde mit der Sprachanalysesoftware Praat durchgeführt. Insgesamt wurden bei der Merkmalsextraktion jeweils 169 Rohkennwerte pro Sprachsample extrahiert und anschließend sprecherspezifisch z-normalisiert (338 Kennwerte, 169 Rohdaten, 169 z-normalisierte Kennwerte). In der vorliegenden Untersuchung wurden 26 prosodiebezogene Kennwerte extrahiert. Diese berechneten sich aus den Konturverläufen der Grundfrequenz (F_0) und Intensität. Folgende Funktionale (High-Level Features) wurden aus ihnen abgeleitet: Mittelwert, 2.-4. Quartil, Standardabweichung; Maximum, Minimum, Bandbreite sowie Position und Funktionswerte der Maxima und Minima. Darüber hinaus wurden die Verläufe von Jitter und Shimmer extrahiert und Mittelwerte berechnet. Des Weiteren wurden 107 spektrale Kennwerte berechnet, darunter Frequenz und Bandbreite der Formanten 1-5 sowie die Frequenz und Amplitude der ersten beiden Harmonischen. Ferner wurden 4 Hammarberg Indices, das LTAS (Long Term Average Spectrum) für 6 Frequenzbänder (125-200Hz, 200-300Hz, 500-600Hz, 1000-1600Hz, 5000-8000Hz), das Verhältnis

von niedriger Frequenzenergie unter 500Hz und unter 1000Hz, die Steigung der Regressionsgeraden (Slope) der spektralen Energie über 1000Hz, die Harmonics-to-Noise Ratio (HNR) sowie spektrale Tilt-Features ('Open Quotient', 'Glottal Opening', 'Skewness of Glottal Puls', 'Rate of Glottal Closure') bestimmt. Schlussendlich wurden 36 spektrale Kennwerte (II) berechnet, die sich aus den Mittelwerten der 12 Mel Frequency Cepstral Coefficients Verläufe, sowie ihrer ersten und zweiten Ableitungen zusammensetzten (12 Δ MFCC, 12 $\Delta\Delta$ MFCC).

Im Rahmen der Dimensionalitätsreduktion wurden die (unsupervised) Feature Transformation Methode Principal Component Analysis (PCA) angewendet. Ein weiterer Verfahrensschritt beinhaltete die filterbasierte Feature Subset Selection (supervised) mittels eines Korrelationsfilters. In Rahmen der Hauptselektion wurde Wrapper-basiert über evolutionäre Suchalgorithmen (Genetic Algorithm) schrittweise ein optimaler Merkmalsverbund selektiert. Für jedes Sprachmaterial wurde das Sub-Feature-Set mit der höchsten Vorhersageleistung mittels iterativer Anwendung der Klassifikationsverfahren k-Nearest-Neighbour, Multilayer-Perceptron (spezielle Form eines künstlichen neuronalen Netzes) und Support-Vector-Machine ermittelt. Jeder der genannten Dimensionalitätsreduktionsschritte wurde für männliche und weibliche Sprecher getrennt berechnet, um geschlechtsspezifische Schläfrigkeitsvariationen besser modellieren zu können.

Auf das resultierende selektierte Feature-Set wurden die Klassifikationsverfahren Multilayer-Perceptron mit Feedforward-Netz und Backpropagation-Learning-Algorithm (2 Hidden-Layers mit jeweils 10 Knoten), Support Vector Machine mit radialer Kernelfunktion und das k-Nearest-Neighbour Verfahren ($k = 1$) angewendet, um die Gesamtvorhersageleistung der ASA zu ermitteln und die prädiktive Validität des Verfahrens zu bestimmen. Die Daten wurden in Trainings- und Test-Set getrennt (90% Trainingsdaten, 10% Testdaten). Das Test-Set umfasste ausschließlich ungesehene Daten, die im Trainings-Set nicht enthalten waren. Alle Klassifikatoren wurden zehnmal trainiert. Dazu wurden die Testdaten in 10 disjunkte Sets gleicher Größe unterteilt. Jedes dieser 10 Sets war in einem der 10 Trainingsdurchläufe nicht Teil der Trainingsdaten und wurde stattdessen als Test-Set verwendet. Die so erzielte 10-fache Kreuzvalidierung reduzierte den Rechenaufwand bei gleichzeitiger Sicherung der strikten Sprecherunabhängigkeit. Als finaler Klassifikationsfehler wurde der Mittelwert aus den Klassifikationsfehlern der 10 Test-Sets verwendet. Als weitere Validierungsindizien werden

neben den Klassifikationsgütern auch korrelative Zusammenhänge von Selbstreportgrößen und einzelnen Sprachkennwerten herangezogen.

Der Ergebnisteil präsentiert die Validierungsergebnisse geordnet nach Sprachsegmenteinheiten (vgl. Kapitel 2- Kapitel 5). Die Validierungsergebnisse für das Sprachsegment „Vokal“ (extrahiertes [a:], extrahiertes [o:] und isoliertes [a:]) zeigen für durchschnittlich 35.33 Einzelmerkmale mindestens kleine Effekstärken Korrelationen mit KSS-Werten ($r > .2$). Die stärksten Zusammenhänge finden sich für das Merkmal „z-transformierte Steigung der Regressionsgrade des Fundamentalfrequenzverlaufs“ ($r = -.41$). Die Mustererkennungsverfahren-bezogene Vorhersagegüte liegt bei einer Klassifikationsgenauigkeit von durchschnittlich 81% (Sensitivität = 65%, Spezifität = 87%, $r = .52$). Die Validierungsergebnisse für das Sprachsegment „Wort“ (extrahiertes Wort „Suchen“) zeigen für 32 Einzelmerkmale mindestens kleine Effekstärken Korrelationen mit KSS-Werten. Die stärksten Zusammenhänge finden sich für das Merkmal „z-transformiertes 75tes Perzentil der Bandwidth der dritten Formantfrequenz“ ($r = .42$). Die Mustererkennungsverfahren-bezogene Vorhersagegüte liegt bei einer Klassifikationsgenauigkeit von durchschnittlich 80% (Sensitivität = 45%, Spezifität = 93%, $r = .48$). Die Validierungsergebnisse für das Sprachsegment „Phrasale Einheit“ (Fahrerassistenzsystem-Satz 3, Fahrerassistenzsystem-Satz 9, Probandenvorstellungs-Introphrase) zeigen für durchschnittlich 27 Einzelmerkmale mindestens kleine Effekstärken Korrelationen mit KSS-Werten. Die stärksten Zusammenhänge finden sich für das Merkmal „z-transformiertes 75tes Perzentil der Bandwidth der dritten Formantfrequenz“ ($r = .42$). Die Mustererkennungsverfahren-bezogene Vorhersagegüte liegt bei einer Klassifikationsgenauigkeit von durchschnittlich 83% (Sensitivität = 68%, Spezifität = 87%, $r = .58$). Die Validierungsergebnisse für das Sprachsegment „Multiple Phrasale Einheit“ (freie Bildbeschreibung) zeigen für 14 Einzelmerkmale mindestens kleine Effekstärken Korrelationen mit KSS-Werten. Die stärksten Zusammenhänge finden sich für das Merkmal „Mittelwert der zweiten Ableitung des Verlaufs des Mel Frequency Cepstral Coefficient 1“ ($r = .27$). Die Mustererkennungsverfahren-bezogene Vorhersagegüte liegt bei einer Klassifikationsgenauigkeit von durchschnittlich 82% (Sensitivität = 62%, Spezifität = 89%, $r = .62$). Die bezogen auf die Gesamtklassifikationsleistung erzielten Ergebnisse sind somit über alle Sprachmaterialien ähnlich gut, auch wenn die Anzahl der bedeutsamen Einzelsprachkennwerte deutlich variieren. Zusammengefasst kann festgehalten werden, dass somit das Hauptziel der Arbeit, aus sprachlicher Interaktion Schläfrigkeit Zustände zu detektieren, anhand diverser Sprachsegmente erreicht wurde.

Der Abgleich der gefundenen Ergebnisse mit den Prognosen des Kognitiv-phonetischen Mediatormodells zeigt eine gute Übereinstimmung. Eine ähnlich gute Übereinstimmung zeigt auch der Vergleich mit den Befunden der Aktivierungsdimensionen-bezogenen Emotionsdetektion. Im Gegensatz zu der sehr guten Spezifität ist die Sensitivität der ASA lediglich als zufriedenstellend einzuordnen. Grund hierfür könnte die ungleiche Verteilung von schläfrigen und wachen Sprachproben sein. Der vorliegende Überhang von wachen Sprachproben erleichtert es den Machine-Learning Verfahren diese zu identifizieren. Ein weiterer Grund für die mittlere Sensitivität kann in der vorgenommenen Dichotomisierung der Schläfrigkeitwerte gesehen werden. Vor allem Klassifikationsfehler für die vielen im Trennbereich zwischen $KSS = 5$ (nicht schläfrig) und $KSS = 6$ (schläfrig) liegenden Sprachproben können in diesem Sinne voraussichtlich für die mittlere Sensitivität verantwortlich gemacht werden.

Ist die Untersuchungsanlage geeignet die Validität der Stimmanalyse zu überprüfen? Unerlässlich für die Beurteilung der Aussagekraft der Studie ist die Analyse der Fehlersphären der (a) Instrumenten-immanenten sowie der (b) Versuchsdesign-immanenten Fehlerquellen. Einzuschätzen ist daher zum einen die Frage nach dem Goldstandard-Charakter des Validierungssankers Karolinska Sleepiness Scale (*Ground Truth*) und zum anderen die Frage nach konfundierenden Größen, die für sich oder in Kombination als Alternativerklärung der Stimmmerkmalsveränderungen in Frage kommen. Die Diskussion der internen Validität der Untersuchung zeigt, dass sowohl instrumenten- als auch designimmanente Fehlerquellen gut kontrolliert werden konnten. Die externe Validität der Ergebnisse hingegen lässt sich nur schwer abschätzen. Die Generalisierbarkeit der Ergebnisse bezogen auf Variationen von Sprecher-Traitmerkmalen (Nicht-Muttersprachler, Dialekträger, Senioren), auf Variationen von Sprecher-Statemerkmale (erkältet, geraucht, gereizt, traurig) sowie auf Variationen von Sprechsituationen (Störgeräusche, simultane Fremdsprecher, Telefon- und Funkverkehrsqualität) lässt sich vor dem Hintergrund unseres derzeitigen Kenntnisstand nur schwer abschätzen (Abschnitt 6.2).

Im abschließenden Forschungsausblick werden die genannten Einschränkungen aufgegriffen und Perspektiven zur Weiterentwicklung der Acoustic Sleepiness Analysis eröffnet (vgl. Abschnitt 6.3). Zukünftige Forschungsbemühungen sollten sich demnach (a) mit der Weiterentwicklung der Acoustic Sleepiness Analysis entlang des 5-stufigen Messprozederes, (b) mit weiteren Validierungsszenarien und (c) mit dem forschungsstrategischen Aspekt des Aufbaus einer leistungsfähigen Sprachanalyse-Infrastruktur (frei zugängliche Sprachanalysesoftware

und Schläfrigkeits-Sprachdatenbanken) beschäftigen. Die Weiterentwicklung der ASA könnte beinhalten, unterschiedliche Aufnahmequalitäten und Umgebungsgeräuschbedingungen als eingehendes Sprachmaterial zu testen sowie vorverarbeitende Noise-Filter anzuwenden. Hinsichtlich der Segmentierung sollten ausgesprochen schläfrigkeits-sensitive phonetische Einheiten identifiziert werden. Die Phase der Merkmalsberechnung sollte dringend um Timing, Dauer, Pausen und Rhythmus Kennzahlen, sowie Kennzahlen der nicht-linearen Zeitreihenanalyse erweitert werden. Des Weiteren kommt gerade bei zusätzlich erweiterten Merkmalssetgrößen den Dimensionalitätsreduktions-Verfahren eine besondere Bedeutung zu (z.B. Sequentiell Forward Floating Selection, LDA basierte Dimensionsreduktion). Weitere Machine-Learning Algorithmen und Mustererkennungsverfahren wie Maximum-Likelihood Bayes Classifier, Fuzzy Membership Indexing, Hidden Markov Models und Gaussian Mixture Density Models könnten zusätzlich in den Prozess der Klassifikation integriert werden.

Bezogen auf alternative Validierungsszenarien sollten zukünftige Untersuchungsdesigns darum bemüht sein, größere Menge eindeutig schläfriger Stimmmaterialien zu erhalten, indem die Zeit der Schlafdeprivation verlängert und eine strengere Monotoniesituation hergestellt wird. Um die anwendungspraktische Bedeutung der erzielten Detektionsraten einzuschätzen, sind in Zukunft insbesondere folgende Fragen zu beantworten: Gibt es eine an die anvisierte Anwendungssituation angepasste *Diversivität von Zuständen* (z.B. Traurigkeit, Langeweile, Motivationslosigkeit), die mit Schläfrigkeit verwechselt werden und somit die False Alarm Rate steigern könnten? Sind alle Zustände - und nur genau die - repräsentativ erfasst, zwischen denen das System differenzieren können soll? Sind zu erwartende Zustandskombinationen erfasst, die die Detektion der Schläfrigkeit maskieren und somit die Sensitivität des Instruments beeinflussen könnten (z.B. *schläfrigkeitsmaskierende Zustandskombinationen* wie schläfrig und ängstlich, schläfrig und gereizt, schläfrig und bemüht)?

Die Methodik der akustischen Schläfrigkeitsmessung bietet ein reichhaltiges Transferpotenzial. So eröffnet die akustische Echtzeit-Erfassung von internen Zuständen eine Reihe von vielversprechenden Anwendungsfeldern, wenn belästigungsarme, den Tätigkeitsvollzug nicht beeinträchtigende Messzugänge erforderlich sind. Die Implementierbarkeit des akustischen Messzugangs in diese Alltagskontexte unterstreicht daher zusätzlich die potenzielle Bedeutung stimmbasierter psychologischer Messinstrumente. Zukunftsfähige Anwendungsfelder der akustischen Stimmessung bieten in diesem Sinne Fragestellungen der Markt- und Werbewirkungs-, Mensch-Maschine-Interaktions-, Usability-, klinisch-therapeutischen oder auch

eignungsdiagnostischen Forschung. Auf eine zweite über die Erfassung von Schläfrigkeitsszuständen hinausreichende Entwicklungsperspektive des akustischen Messansatzes weisen zustandsadaptive Unfallrisiko-Management-Systeme. Die Integration von Schläfrigkeits-Warnsystemen in die ganzheitliche Architektur von - umfassend sicherheitskritische Zustände detektierenden und adäquate Reaktionen produzierenden - Systemen, könnte eine lohnenswerte Aufgabe zukünftiger arbeitspsychologischer Forschungsbemühungen darstellen.

8 LITERATURVERZEICHNIS

- Åkerstedt, T., Kecklund, G., Gillberg, M., Lowden, A. & Axelsson, J. (2000). Sleepiness and days of recovery. *Transportation Research Part F3*, 251-261.
- Åkerstedt, T. & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52, 29-37.
- Åkerstedt, T. (1988). Sleepiness as a consequence of shift work. *Sleep*, 11, 17-34.
- Åkerstedt, T., Kecklund, G. & Gillberg, M. (2007). Sleep and sleepiness in relation to stress and displaced work hours. *Psychology and Behavior*, 92, 250-255.
- Åkerstedt, T., Kecklund, G. & Hörte, L.-G. (2001). Night driving, season, and the risk of highway accidents. *Sleep*, 24, 401-406.
- Aldrich, M. S. (1989). Automobile accidents in patients with sleep disorders. *Sleep*, 12, 487-494.
- Alloway, C.E., Ogilvie, R.D., Shapiro, C.M. (1997). The alpha attenuation test: assessing excessive daytime sleepiness in narcolepsy-cataplexy. *Sleep*, 20, 258-266.
- Amundsen, F.H. & Christensen, P. (1986) "Relationship between geometric design and driving speed in residential areas". Oslo: Transportøkonomisk Institutt, Norway.
- Andreeva, E. O., Aarabi, P., Philiastides, M. G., Mohajer, K. & Emami, M. (2004). Driver drowsiness detection using multi-modal sensor fusion. *Proceedings of the International Society for Optical Engineering*, 380-390.
- Apple, W., Streeter, L.A. & Krauss, R.M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, 37, 715-727.
- Arnedt, J.T., Wilde, G.J., Munt, P.W. & MacLean, A.W. (2000). Simulated driving performance following prolonged wakefulness and alcohol consumption: separate and combined contributions to impairment. *Journal of Sleep Research*, 9, 233-241.
- Arnold, P.S., Hartley, L.R., Corry, A., Hochstadt, D., Penna, F. & Feyer, A.M., (1997). Hours of work, and perceptions of fatigue among truck drivers. *Accident Analysis and Prevention* 29, 471-477.
- Bachorowski, J.-A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science*, 8, 53-57.
- Baddeley, A.D. & Salamé, P. (1986). The unattended speech effect: Perception or memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 525-529.

- Baenninger, R. (1997). On yawning and its functions. *Psychonomic Bulletin & Review*, 4, 198-207.
- Baenninger, R., Binkley, S. & Baenninger, M. (1996). Field observations of yawning and activity in humans. *Physiology & Behavior*, 59, 421-425.
- Balkin, T.J., Bliese, P.D., Belenky, G., Sing, H., Thorne, D.R. & Thomas, M. (2004). Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. *Journal of Sleep Research*, 13, 219-227.
- Ball, G. & Breese, M.J. (1999). Modeling the emotional state of computer users. *Online Proceedings of the Workshop on Attitude, Personality and Emotions in User-Adapted Interaction*.
- Barbé, F., Pericás, J., Munoz, A., Findley, L.J., Antó, J.M., Agusti, A. & de Lluc, J.M. (1998). Automobile Accidents in Patients with Sleep Apnea Syndrome. *American Journal of Respiratory and Critical Care Medicine*, 158, 18-22.
- Batliner, A., Huber, R., Niemann, H., Nöth, E., Spilker, J. & Fischer, K. (2000). The recognition of emotion. In W. Wahlster (Ed.), *VerbMobil: Foundations of Speech-to-Speech Translation*, (pp. 122-130). Berlin: Springer.
- Batliner, A., Hacker, C., Steidl, S., Noeth, E.; D'Arcy, S, Rusell, M.; Wong, M. (2004). "You stupid tin box" – Children interacting with the AIBO robot: A crosslinguistic emotional speech corpus. *Proceedings of the 4th International Conference of Language Resources and Evaluation LREC 2004* (LREC Lisbon 2004), 171-174.
- Batliner, A., Steidl, S., Schuller, B, Seppi, D., Laskowski, K, Vogt, T., Devillers, L., Vidrascu, L., Amir, N., Kessous, L & Aharonson, V. (2006). Combining efforts for improving automatic classification of emotional user states. In T. Erjavec & J.Z. Gros (Eds.), *Language Technologies, IS-LTC 2006*, (pp. 240-245). Ljubljana, Slovenia: Informatična Družba (Information Society).
- Bedard, M.A., Montplaisir, J., Richer, F. & Malo, J. (1991). Nocturnal hypoxemia as a determinant of vigilance impairment in sleep apnea syndrome. *Chest*, 100, 367-370.
- Berg, T. (1988). *Die Abbildung des Sprachproduktionsprozesses in einem Aktivationsflußmodell: Untersuchungen an deutschen und englischen Versprechern*. Tübingen: Niemeyer.
- Berrichi, H., Zannad, F., Hannharth, B. & Tiberge, M. (1999). Drowsiness driver: a reliable neurophysiological parameter. *Sleep*, 22, 168.
- Berton, A. (2004). *Konfidenzmaße und deren Anwendungen in der automatischen Sprachverarbeitung*. Dresden: Universitätsverlag.

- Bindreiter, S. (2006). *Audio gesteuerte Animationen*. Unveröffentlichte Diplomarbeit. Fachhochschule Hagenberg, Österreich.
- Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.
- Bittner, R., Smrcka, P., Vysoky, P., Hana, K., Pousek, L. & Schreib, P. 2000. Detecting of fatigue states of a car driver. *Proceedings of the Medical Data Analysis. First International Symposium*, 260-273.
- Blanken, G. (1988). Zur Ausgrenzbarkeit der linguistischen Formulierungsprozesse. Neuro-linguistische Evidenzen. In G. Blanken, J., Dittmann & C.-W. Wallesch (Hrsg.), *Sprachproduktionsmodelle: Neuro- und psycholinguistische Theorien der menschlichen Sprach-erzeugung*, (S. 83-110). Freiburg: Hochschulverlag.
- Boersma, P. & Weenik, D. (1996). *PRAAT, a system for doing phonetics by computer, version 3.4*. Institute of Phonetic Sciences of the University of Amsterdam, Report 132.
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, 17, 97-110.
- Bonnet, M.H. & Arand, D.L. (1998). Sleepiness as measured by modified multiple sleep latency testing varies as a function of preceding activity. *Sleep*, 21, 477- 483.
- Bonnet, M.H. (2000). Sleep Deprivation. In M.H. Kryger, T. Roth & W.C. Dement (Hrsg.), *Principles and Practice of Sleep Medicine*, (S. 53-71). Philadelphia: Saunders Company.
- Bose, I. (1994). *Zur temporalen Struktur frei gesprochener Texte*. Frankfurt/Main: Hector.
- Bose, I. (2001). Methoden der Sprechausdrucksbeschreibung am Beispiel kindlicher Spielkommunikation. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*, 2, 262-303.
- Boucsein, W. (1992). *Electrodermal Activity*. New York: Plenum Press.
- Boucsein, W. (2005). Electrodermal measurement. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. Hendrick. *Handbook of Human Factors and Ergonomic Methods*, 1-18. London: CRC Press.
- Boucsein W. & Ottman W. (1996). Psychophysiological stress effects from the combination of night- shift work and noise. *Biological Psychology*, 42 (3), 301-22.
- Boverie, S. (2004). Driver fatigue monitoring technologies and future ideas. *Proceedings of the 2004 AWAKE Road Safety Workshop*. Balocco, Italy. 29 Sep 2004.
- Boverie, S., Daurenjou, D., Esteve, D., Poulard, H. & Thomas, J. (2002). Driver vigilance monitoring. New Developments. IFAC, 15th Triennial World Congress. Barcelona, Spain. 21-26 July 2002.

- Brandstätter, E. (1999). Confidence intervals as an alternative to significance testing. *Methods of Psychological Research*, 4, 33-46.
- Breazeal, C. & Aryananda, L. (2000). Recognition of Affective Communicative Intent in Robot-Directed Speech. *Autonomous Robots*, 12(1), 83-104.
- Breitenstein, C., Van Lancker, D. & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cognition and Emotion*, 15, 57-79.
- Brenner, M. & Cash, J.R. (1991). Speech Analysis as an Index of Alcohol Intoxication – The Exxon Valdez Accident. *Aviation Space and Environmental Medicine*, 62, 893-898.
- Browman, C.P. & Goldstein, L. (1992). Articulatory phonology: A review. *Phonetica*, 49, 155-180.
- Brügge, W. & Mohs, K. (1996). *Therapie funktioneller Stimmstörungen*. Basel: Ernst Reinhard.
- Bunn, T. L., Slavova, S., Struttman, T.W., & Browning, S. R. (2005). Sleepiness/fatigue and distraction/inattention as factors for fatal versus nonfatal commercial motor vehicle driver injuries. *Accident Analysis and Prevention*, 37, 862-869.
- Burges, C.J.C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2(2), 121-167.
- Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. & Weiss, B. (2005). A database of German emotional speech. *Proceedings of Interspeech 2005* (Lisboa, Portugal) (2005) 1517-1520).
- Butterworth, B. (1989). Lexical access in speech production. In W. Marslen-Wilson (Ed.), *Lexical Representation and Process* (pp. 108-135). Cambridge, MA: MIT Press.
- Buysse, D.J., Reynolds, C.F., Monk, T.H., Berman, S.R. & Kupfer, D.J. (1989). The Pittsburgh Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Research*, 28, 193-213.
- Cacioppo, J.T., Berntson, G.G., Larsen, J.T., Poehlmann, K.M. & Ito, T.A. (2000). The psychophysiology of emotion. In M. Lewis, J.M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 173-191). New York: Guilford Press.
- Cajochen, C. (2007). Schlafregulation. In H. Peter, T. Penzel & J.H. Peter (Hrsg.), *Enzyklopädie der Schlafmedizin*, (S. 1102-1111). Heidelberg: Springer.
- Carskadon, M.A. & Dement, W.C. (1977). Sleepiness and sleep state on a 90-min schedule. *Psychophysiology*, 14, 127-133.

- Carskadon, M.A. & Dement, W.C. (1982). The multiple sleep latency test: what does it measure? *Sleep*, 5, 67-72.
- Cassel, W., Stephan, S., Ploch, T. & Peter, J.H. (1989). Psychologische Aspekte schlafbezogener Atemregulationsstörungen. *Pneumologie*, 43, 625-629.
- Chou W. & Juang B.H. (2003). *Pattern Recognition in Speech and Language Processing*. London: CRC Press.
- Chow, S.L. (1988). Significance test or effect size? *Psychological Bulletin*, 103, 105-110.
- Chow, S.L. (1989). Significance tests and deduction: Reply to Folger (1989). *Psychological Bulletin*, 106, 161-165.
- Chow, S.L. (1991a). Conceptual rigor versus practical impact. *Theory & Psychology*, 1, 337-360.
- Chow, S.L. (1991b). Rigor and logic. A response to comments on .Conceptual rigor. *Theory & Psychology*, 1, 389-400.
- Chow, S.L. (1991c). Some reservations about power analysis. *American Psychologist*, 46, 1088-1089.
- Chow, S.L. (1996). *Statistical significance: Rationale, validity and utility*. London, UK: Sage.
- Chow, S.L. (1998). Précis of Statistical significance: Rationale, validity and utility. *Behavioral and Brain Sciences*, 21, 169-239.
- Chuang, Z. & Wu, C. (2004). Emotion recognition using acoustic features and textual content. *Proceedings of IEEE International Conference on Multimedia and Expo*, 53-56.
- Clark H. & Tree, J. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84, 73-111.
- Cluydts, R., De Valck, E., Verstraeten, E. & Theys, P. (2002). Daytime sleepiness and its evaluation. *Sleep Medicine Review*, 6, 83-96.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003.
- Colt, H.G., Haas, H. & Rich, G.B. (1991). Hypoxemia vs. sleep fragmentation as cause of excessive daytime sleepiness in obstructive sleep apnea. *Chest*, 100, 1542-1548.
- Connor, J., Norton, R., Ameratunga, S., Robinson, E., Civil, I., Dunn, R., Bailey J. & Jackson, R.. (2002). Driver sleepiness and risk of serious injury to car occupants: Population based case control study. *British Medical Journal*, 324(7346), 1125-1129.
- Connor, J., Whitlock, G., Norton, R. & Jackson, R. (2002). The role of driver sleepiness in car crashes: a systematic review of epidemiological studies. *Accident Analysis and Prevention*, 33, 31-41.

- Cook, Y., Schmitt, F. & Berry, D. (1988). The effects of nocturnal sleep, sleep disordered breathing and periodic movements of sleep on the objective and subjective assessment of daytime somnolence in healthy aged adults. *Journal of Sleep Research*, 17, 95.
- Cortina, J. & Dunlap, W.P. (1997). On the logic and purpose of significance testing. *Psychological Methods*, 2, 161-172.
- Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W. & Taylor, J.G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18, 32-80.
- Cowles, M. (1989). *Statistics in psychology: An historical perspective*. Hillsdale, NJ: Erlbaum.
- Cryer, J. D. (2008). *Time series analysis*. 2. Auflage, Springer
- Cristianini, N. & Shawe-Taylor, J.(2000). *An Introduction to Support Vector Machines and other Kernel-based Learning Methods*. Cambridge: Cambridge University Press.
- Crum, M. R., & Morrow, P. C. (2002). The influence of carrier scheduling practices on truck driver fatigue. *Transportation Journal*, 42(1), 20-40.
- Cummings, K., Chin, S.B. & Pisoni, D.B. (1996). Analysis of the glottal excitation of intoxicated versus sober speech: A first report. *The Journal of the Acoustical Society of America*, 99, 2549-2574.
- Cutting, J.C. & Ferreira, V.S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25, 318-344.
- Damousis, I.G. & Cester, I. (2007). *Online EEG analysis and classification for the prediction of hypovigilance related driving accidents*. "Monitoring sleep and sleepiness with new sensors within medical and industrial applications", SENSATION IP 2nd International Conference, 4-5 June 2007, Chania, Greece.
- Davitz, J.R. (1964). Auditory correlates of vocal expression of emotional feeling. In J.R. Davitz (Ed.), *The communication of emotional meaning*, (pp. 101-112). New York: McGraw-Hill.
- Dell, G.S. & O'Seaghdha, P.G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review*, 98, 604-614.
- Dell, G.S. (1986). A Spreading-Activation Theory of Retrieval in Sentence Production. *Psychological Review*, 93, 283-321.

- Desmond, P. A., & Hancock, P. A. (2001). Active and passive fatigue states. In P. A. Hancock & P. A. Desmond (Eds), *Stress, workload, and fatigue* (pp455-465). New Jersey: Lawrence Erlbaum Associates.
- Dinges, D.F. & Kribbs, N. (1991). Performing while sleepy: effects of experimentally induced sleepiness. In T.H.Monk (Ed.), *Sleep, Sleepiness and Performance*, (pp. 97-128). Chichester: J. Wiley & Sons.
- Dinges, D.F. (1995). An overview of sleepiness and accidents. *Journal of Sleep Research*. Vol 4. Suppl.2. 4-14.
- Dinges, D.F., Pack, F., Williams, K., Gillen, K.A., Powell, J.W. & Ott, G.E. (1997). Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. *Sleep*, 20, 267-277.
- Doghramji, K., Mitler, M.M. & Sangal, R.B. (1997). A normative study of the maintenance of wakefulness test (MWT). *Electroencephalography and Clinical Neurophysiology*, 103, 554- 562.
- Dorrian, J., Rogers, N.L. & Dinges, D.F. (2005). Psychomotor vigilance performance: A neurocognitive assay sensitive to sleep loss. In C.A. Kushida (Ed.), *Sleep Deprivation: Clinical Issues, Pharmacology and Sleep Loss Effects*, (pp. 39-70). New York: Marcel Dekker, Inc.
- Duda, R.O., Hart, P.E. & Stork, D.G. (2001). *Pattern Classification*. New York: Wiley.
- Durmer, J.S. & Dinges, D.F. (2005). Neurocognitive consequences of sleep deprivation. *Seminars in Neurology*, 25, 117-129.
- Erdmann, U., Hagenmeyer, L., Muttray, A. & Geißler, B. (2006). Lidschlagparameter als Indikator für Müdigkeit. In Gesellschaft für Arbeitswissenschaft (GfA), editor, *Innovationen für Arbeit und Organisation*, (pp. 521–524), Dortmund: GfA Press.
- European Transport Safety Council (2001) *Driver Fatigue Working Party* - Chair. Dr N Macdonald, Brussels.
- Fairclough, S. H. (2001). Mental effort regulation and the functional impairment of the driver. In P. A. Hancock & P. A. Desmond (Eds), *Stress, workload, and fatigue* (pp479-502). New Jersey: Lawrence Erlbaum Associates.
- Falk, R. & Greenbaum, C.W. (1995). Significance tests die hard. The amazing persistence of a probabilistic misconception. *Theory & Psychology*, 5, 75-98.
- Falleti, M. G., Maruff, P., Darby, D. G., & McStephen, M. (2003). Qualitative similarities in cognitive impairment associated with 24h of sustained wakefulness and a blood alcohol concentration of 0.05%. *Journal of Sleep Research*, 12, 265-274.

- Falou, W., Duchene, J., Grabisch, M., Hewson, D., Langeron, Y., Lino, F. (2003). Evaluation of driver discomfort during long-duration car driving. *Applied Ergonomics*, 34, 249-255.
- Fant, G. (1960). *The Acoustic theory of speech production*. Den Haag, Niederlande: Mouton & Co.
- Fell, D. (1994). *Safety Update: Problem Definition and Countermeasure Summary: Fatigue*. New South Wales Road Safety Bureau RUS No 5
- Fellbaum, K. (1984). *Sprachverarbeitung und Sprachübertragung*. Berlin: Springer-Verlag.
- Fernandez, R. & Picard, R.W. (2003). Modeling drivers' speech under stress. *Speech Communication*, 40, 145-159.
- Fietze, I. (2007). Thermoregulation. In H. Peter, T. Penzel & J.H. Peter (Hrsg.), *Enzyklopädie der Schlafmedizin*, (S. 1224-1229). Heidelberg: Springer.
- Findley, L. J., Suratt, P. M., & Dinges, D. F. (1999). Time-on-task decrements in "steer clear" performance of patients with sleep apnea and narcolepsy. *Sleep*, 22, 804-809.
- Findley, L., Barth, J., Powers, D., Wilhait, S., Boyd, D. & Suratt, P. (1986). Cognitive Impairment in Patients with Obstructive Sleep Apnea and Associated Hypoxemia. *Chest*, 90, 686-690.
- Findley, L., Unverzagt, M., Guchu, R., Fabrizio, M., Buckner, J. & Suratt, P. (1995). Vigilance and automobile accidents in patients with sleep apnea or narcolepsy. *Chest*, 108, 619-624.
- Findley, L.J., Unverzagt, M.E. & Suratt, P.M. (1988). Automobile accidents involving patients with obstructive sleep apnea. *American Review of Respiratory Disease*, 138, 337-340.
- Fletcher, A., Lamond, N., van den Heuvel, C.J., & Dawson, D. (2003). Prediction of performance during sleep deprivation and alcohol intoxication using a quantitative model of work related fatigue. *Sleep Research Online*, 5 (2), 67-75.
- Folger, R. (1989). Significance tests and the duplicity of binary decisions. *Psychological Bulletin*, 106, 155-160.
- Fowler, C. & Housum, J. (1987). Talkers signaling of new and old words in speech and listeners' Perception and Use of This Distinction. *Journal of Memory and Language*, 26, 45-49.
- Friedrich, G., Bigenzahn, W. & Zorowka, P. (2000). *Phoniatrie und Pädaudiologie*. Bern: Hans Huber.
- Frith, C.D., Friston, K.J., Liddle, P.F. & Frackowiak, P.F. (1991). A PET study of word finding. *Neuropsychologia*, 29, 1137-1148.

- Fromkin, V.A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47, 27-52.
- Fulda, S. & Schulz, H. (2003). Cognitive dysfunction in sleep-related breathing disorders: A metaanalysis. *Sleep Research Online*, 5, 13-43.
- Fuster, J.M. (1989). *The prefrontal cortex*. New York: Raven Press.
- Galley, N. & Schleicher, R. (2002). Fatigue indicators from the electrooculogram -a research report. *AWAKE consortium internal report*.
- Galley, N. (2007). On the way to a sleep warner. "Monitoring sleep and sleepiness with new sensors within medical and industrial applications", SENSATION IP 2nd International Conference, 4-5 June 2007, Chania, Greece.
- Gander, P. H., Marshall, N. S., James, I., & Le Quesne, L. (2006). Investigating driver fatigue in truck crashes: Trial of a systematic methodology. *Transportation Research Part F*, 9, 65-76.
- Garbarino, S., Nobili, L., Beelke, M., De Carli, F. & Ferrillo, F. (2001). The contributing role of sleepiness in highway vehicle accidents. *Sleep*, 24, 203-206.
- Garrett, M.F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language Production*, (pp. 177-220). London: Academic Press.
- Garrett, M.F. (1982). Production of speech: Observations from normal and pathological language use. In A. Ellis (Ed.), *Normality and Pathology in Cognitive Functions*, (pp. 19-76). London: Academic Press.
- George, C.F., Boudreau, A.C. & Smiley, A. (1996). Comparison of simulated driving performance in narcolepsy and sleep apnoea patients. *Sleep*, 19, 711-717.
- George, C.F., Boudreau, A.C. & Smiley, A. (1997). Effects of nasal CPAP on simulated driving performance in patients with obstructive sleep apnoea. *Thorax*, 52, 648-653.
- George, C.F., Nickerson, P.W., Hanly, P.J., Millar, T.W. & Kryger, M.H. (1987). Sleep apnoea patients have more automobile accidents. *Lancet*, 2, 447.
- Gigerenzer, G. & Murray, D.J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.
- Hagen, R.L. (1997). In praise of the null hypothesis statistical test. *American Psychologist*, 52, 15-24.
- Gigerenzer, G. (1993). The Superego, the Ego, and the Id in statistical reasoning. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues*, 311-339. Hillsdale, NJ: Erlbaum.

- Gillberg, M., Kecklund, G. & Åkerstedt, T. (1996). Sleep and performance of professional drivers in a truck simulator – comparison between day and night driving. *Journal of Sleep Research*, 5, 12-15.
- Glenville M. & Broughton, R.J. (1979). Reliability of the Stanford sleepiness scale compared to short duration performance tests and the Wilkinson auditory vigilance task. In P. Passouant and I. Oswald (eds), *Pharmacology of the states of alertness* (pp. 235-244). Oxford: Pergamon.
- Golz, M., Sommer, D., Schnupp, T., Holzbrecher, M. & Mandie, D.P. (2007). *Detection of microsleep events data reduction or data fusion? "Monitoring sleep and sleepiness with new sensors within medical and industrial applications"*, SENSATION IP 2nd International Conference, 4-5 June 2007, Chania, Greece.
- Gottlieb, W., Galley, L., Schleicher, R., Galley, N., & Churan, J. (2004). *EEG and Ocular Parameters while Driving in a Simulation Study*
- Grace, R., Guzman, A., Staszewski, J., Dinges, D. F., Mallis, M., & Peters, B. A. (1998). The Carnegie Mellon Truck Simulator, A Tool to Improve Driving Safety. *SAE Technical Paper Series* (982845), 1-6.
- Greely, H.P, Berg, J., Friets, E., Wilson, J.P., Raghavan, S. & Picone, J. (2006). Detecting Fatigue from Voice using Speech Recognition. *Proceedings in International Symposium on Signal Processing and Information Technology, IEEE 2006*, 567-571.
- Greenberg, G., Watson, R. & Deptula, D. (1987). Neuropsychological Dysfunction in Sleep Apnea. *Sleep*, 10, 254-262.
- Greisbach, R (2003). Phonetische Analysen. In T. Hermann & J. Grabowski (Hrsg.), *Sprachproduktion* (= Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie 3: Sprache, Band 1, S. 51-176). Göttingen: Hogrefe.
- Griefahn, B. (2007). Melatonin und zirkadianer Rhythmus. In H. Peter, T. Penzel & J.H. Peter (Hrsg.), *Enzyklopädie der Schlafmedizin*, (S. 730-734). Heidelberg: Springer.
- Griefahn, B., Künemund, C., Bröde, P. & Mehnert, P. (2001). Zur Validität der deutschen Übersetzung des Morningness-Eveningness-Questionnaires von Horne und Östberg. *Somnologie*, 5, 71-80.
- Guilleminault, C., Partinen, M., Quera-Salva, M.A., Hayes, B., Dement, W.C. & Nino-Murcia, G. (1988). Determinants of Daytime Sleepiness in Obstructive Sleep Apnea. *Chest*, 94, 32-37.
- Gussenhoven, C. & Jacobs, H. (2003). *Understanding phonology*. Understanding Language Series. London: Arnold.

- Hagenmeyer, L. (2007). Development of a Multimodal, Universal Human-Machine-Interface for Hypovigilance-Management-Systems. Heimsheim: Jost-Jetter-Verlag.
- Hager, W. (2000). About some misconceptions and the discontent with statistical tests in psychology *Methods of Psychological Research Online* , Vol.5, No.1 Institute for Science Education.
- Haraldsson, P.O., Carenfelt, C., Laurell, H. & Tornros, J. (1990). Driving vigilance simulator test. *Acta Oto-Laryngologica*, 110, 136-140.
- Hargutt, V. (2003). *Das Lidschlussverhalten als Indikator für Aufmerksamkeits- und Müdigkeitsprozesse bei Arbeitshandlungen.*, Universität Würzburg, Würzburg.
- Harley, T.A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science*, 8, 191-219.
- Harley, T.A. (1993). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes*, 8, 291-309.
- Harrison, W. (2006). *Fatigue and young drivers*. Research report PP05/04. Melbourne: Royal Automobile Club of Victoria
- Harrison, Y. & Horne, J.A. (1997). Sleep deprivation affects speech. *Sleep*, 20, 871-877.
- Harrison, Y. & Horne, J.A. (2000). The impact of sleep deprivation on decision making: a review. *Journal of Experimental Psychology: Applied*, 6, 236-249.
- Hartley, L. R. & Arnold, P. (2001). Managing fatigue in the road transport industry: An occupational safety and health solution. In P. A. Hancock & P. A. Desmond (Eds), *Stress, workload, and fatigue* (pp531-549). New Jersey: Lawrence Erlbaum Associates.
- Harwood, K., Barnett, B., & Wickens, C.D. (1988). Situational awareness: A conceptual and methodological framework. In F.E. McIntire (Ed.), *Proceedings of the 11th Biennial Psychology in the Department of Defense Symposium* (pp. 23-27). Colorado Springs, CO: U.S. Air Force Academy.
- Haslam, D.R. (1981). The military performance of soldiers in continuous operations. In L.C. Johnson, D.J. Tepas, W.P. Colquhoun, M.J. Colligan (Eds.), *Biological rhythms, sleep and shiftwork*, (pp. 435-458). New York: Spectrum.
- Hastie, T., Tibshirani, R. & Friedman. J. (2001). *The Elements of Statistical Learning*. New York: Springer Verlag.
- Haworth, N. L., Heffernan, C. J. & Horne, E. J. (1989). *Fatigue in truck accidents* (Report No. 3). Melbourne: Monash University Accident Research Centre.
- Haykin, S. (1999). *Neural Networks – a Comprehensive Foundation*. Upper Saddle River: Prentice-Hall.

- Hein, H. (2002). Schlaf und Unfallrisiko. *Pneumologie*, 56, 11-12.
- Heitmann, A., Guttkuhn, R., Aguirre, A., Trutschel, U. & Moore-Ede, M. (2001). *Technologies for the monitoring and prevention of driver fatigue*. Proceedings of the First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.
- Herrmann, B.L. (2007). Hypophyse und Hypothalamus-Hypophysen-Nebennierenachse. In H. Peter, T. Penzel & J.H. Peter (Hrsg.), *Enzyklopädie der Schlafmedizin*, (S. 538-541). Heidelberg: Springer.
- Hess, W. (2002a). *Lehrskript Grundlagen der Sprachsignalverarbeitung*, Institut für Kommunikationswissenschaften, Universität Bonn.
- Hess, W. (2002b). *Lehrskript Grundlagen der Phonetik2. Deskriptive Phonetik*. Institut für Kommunikationswissenschaften, Universität
- Hess, W. (2002c). *Lehrskript Einführung in die musikalische Akustik*. Institut für Kommunikationswissenschaften, Universität Bonn.
- Hess, W. (2002d). *Lehrskript Einführung in die musikalische Akustik*. Institut für Kommunikationswissenschaften, Universität Bonn.
- Hockey, G.R. (1986). Changes in participant efficiency as a function of environmental stress, fatigue and circadian rhythms. In K.R. Boff, K.B.L. Kaufman & J.P. Thomas (Eds.), *Handbook of Perception and Human Performance*, (pp. 441-449). New York: Wiley.
- Hoddes, E., Zarcone, V., Smythe, H., Phillips, R. & Dement, W.C. (1973). Quantification of sleepiness: a new approach. *Psychophysiology*, 10, 431- 436.
- Högl, B. & Poewe, W. (2001). Exzessive Tagesschläfrigkeit: Ursachen, Differentialdiagnosen und Untersuchungsmethoden. *Wiener Klinische Wochenschrift*, 113, 278-284.
- Hollien, H., De Jong, G., Martin, C.A., Scharz, R. & Liljegren, K. (2001). Effects of ethanol intoxication on speech suprasegmentals. *The Journal of the Acoustical Society of America*, 110, 3198-3206.
- Horne, J.A. & Reyner, L. A. (1995). Sleep related vehicle accidents. *British Medical Journal*, 310, 565-567.
- Horne, J.A. (1988). *Why we sleep*. Oxford: Oxford Press.
- Horne, J.A. (1993). Human sleep, sleep loss and behaviour: implications for the prefrontal cortex and psychiatric disorder. *The British Journal of Psychiatry*, 162, 413-419.
- Horne, J.A. (1998). Sleep loss and 'divergent' thinking ability. *Sleep*, 11, 528-536.
- Horne, J.A., & Baulk, S.D. (2004). Awareness of sleepiness when driving. *Psychophysiology*, 41, 161-165.

- Horstmann, S., Hess, C.W., Bassetti, C., Gugger, M. & Mathis, J. (2000). Sleepiness-related accidents in sleep apnea patients. *Sleep*, 23, 383-389.
- Huber, R. (2002). *Prosodisch-linguistische Klassifikation von Emotion*. Berlin: Logos.
- Hulbert, S. (1972). Effects of driver fatigue. In T. W. Forbes (Ed.), *Human factors in highway safety research* (pp.110-132). New York: Wiley-Interscience.
- Hursh, S.R., Redmond, D.P., Johson, M.L., Thorne, D.R. & Belenky, G. (2004). Fatigue models for applied research in Warfighting. *Aviation, Space and Environmental Medicine*, 75, 44-53.
- Huttar, G.L. (1968). Relations between prosodic variables and emotions in normal American English utterances. *Journal of Speech and Hearing Research*, 11, 481-487.
- IPA (Hrsg.) (1989). The international phonetic alphabet (as revised to 1989). *Journal of the International Phonetic Organisation*, 19.
- Jescheniak, J.D. & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1256-1274.
- Jescheniak, J.D. (2002). *Sprachproduktion: Der Zugriff auf das lexikale Gedächtnis beim Sprechen*. Göttingen: Hogrefe.
- Jewett, M.E., Dijk, D.J., Kronauer, R.E. & Dinges, D.F. (1999). Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. *Sleep*, 22, 171-179.
- Jiang, J., Lin, E. & Hanson, D.G. (2000). Acoustic and airflow spectral analysis of voice tremor. *Journal of Speech, Language and Hearing Research*, 43, 191-204.
- Johns, M.W. (1991). A new method for measuring daytime sleepiness: the epworth sleepiness Skala. *Sleep*, 14, 540-545.
- Johns, M.W. (1994). Sleepiness in different situations measured by the Epworth sleepiness Scale. *Sleep*, 17, 703-710.
- Johns, M.W. (1998). Rethinking the assessment of sleepiness. *Sleep Medicine Review*, 2, 3-15.
- Johns, M.W. (2002). Sleep propensity varies with behavior and the situation in which it is measured: the concept of somnificity. *Journal of Sleep Research*, 11, 61-67.
- Juniper, M., Hack, M.A., George, C.F., Davies, R.J. & Stradling, J.R. (2000). Steering simulation performance in patients with obstructive sleep apnoea and matched control subjects. *European Respiratory Journal*, 15, 590-595.

- Jurado, J.L., Luna-Villegas, G. & Buela-Casal, G. (1989). Normal human subjects with slow reaction times and larger time estimations after waking have diminished delta sleep. *Electroencephalography and Clinical Neurophysiology*, 73, 124-128.
- Juslin, P.N. & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129, 770–814.
- Kaida, K., Takahashi, M., Akerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology*, 117, 1574-1581.
- Kales, A., Caldwell, A.B., Cadieux, R.J., Vela -Bueno, A., Ruch, L.G. & Mayes, S.D. (1985). Severe Obstructive Sleep Apnea – II: Associated Psychopathology and Psychological Consequences. *Journal of Chronic Diseases*, 38, 427-434.
- Kapoor, A., Qi, Y., & Picard, R. (2003). *Fully automatic upper facial action recognition*. IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2003) held in conjunction with ICCV 2003.
- Kehrein, R. (2002). The prosody of authentic emotions. In B. Bel & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 Conference*, (pp. 423-426). Aix-en-Provence, France: Université de Provence.
- Kienast, M. & Sendlmeier, W.F. (2000). Acoustical analysis of Spectral and Temporal Changes in Emotional Speech. In Sendlmeier, W. (Ed.), *Speech and Signals: Aspects of Speech Synthesis and Automatic Speech Recognition*, (pp. 157-168). Frankfurt a.M.: Forum Phonicum, Hector.
- Klasmeyer, G. & Sendlmeier, W.F. (2000). Voice and Emotional States. In R.D. Kent & M.J. Ball (Eds.), *Voice Quality Measurement*, (pp. 339-357). San Diego: Singular Publishing Group.
- Knipling, R.R. & Wang, J.-S. (1995) Revised estimates of the US drowsy driver crash problem size based on general estimates system case reviews. *39th Annual Proceedings*. Association for the Advancement of Automotive Medicine.
- Kohler, K. (1977). *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt.
- Kollias, S., Amir, N., Kim, J. & Grandjean, D. (2004). *Description of potential exemplars: Signals and Signs of Emotion*. HUMAINE Human-Machine Interaction Network on Emotions.
- Kotterba, S., Rasche, K., Widdig, W., Duscha, C., Blombach, S., Schultze-Werninghaus, G. & Malin, J.P. (1998). Neuropsychological investigations and event-related potentials in ob-

- structive sleep apnea syndrome before and during CPAP-therapy. *Journal of the Neurological Sciences*, 159, 45-50.
- Kraiss, K.-F. (2003). *Skript zur Vorlesung Mensch-Maschine-Systeme II*. Lehrstuhl für Technische Informatik, RWTH Aachen.
- Krajewski, J. & Wieland R. (2004a). „Silent-Rooms“ als infrastrukturelle Voraussetzungen eines betrieblichen Regenerationsmanagements. In W. Bungard, B. Koop & C. Liebig (Hrsg.), *Psychologie und Wirtschaft leben*. Aktuelle Themen der Wirtschaftspsychologie in Forschung und Praxis. München: Hampp.
- Krajewski, J. & Wieland, R. (2003) Silent Room- die andere Mittagspause: Ein Beitrag zum Human Resource Management im Call Center. *Wirtschaftspsychologie*, 4, 50-53.
- Krajewski, J. & Wieland, R. (2004b). “Endlich allein”- Privatheitsregulation als Determinante erfolgreicher Erholungsprozesse. In Gesellschaft für Arbeitswissenschaft (Hrsg.), *Arbeit + Gesundheit in effizienten Arbeitssystemen*. Dortmund: GfA-Press.
- Krajewski, J. & Wieland, R. (2004c). Einfluss von Pausenorganisation auf unmittelbare und arbeitszeitüberdauernde Beanspruchungsfolgen im Arbeits-Erholungs-Zyklus. In R. Stoll & S. Kreuzfeld (Hrsg.), *Arbeitsphysiologie für Nachwuchswissenschaftler*. Rostock: IMS Press.
- Krajewski, J. & Wieland, R. (2006).Tiefenentspannungsbasierte Restauration von Beanspruchungszuständen im Arbeits-Erholungs-Zyklus. *Wuppertaler Beiträge zur Arbeits- und Organisationspsychologie*, 1, 81-84.
- Krajewski, J. (2005). Cortisolbasierter Nachweis des differentiellen Erholungswertes von Mittagspausen. In M. Rieger (Hrsg.). *Forum Arbeitsphysiologie*. Wuppertal: Institutsdruck ASER.
- Kröger, B.J. (1998). *Ein phonetisches Modell der Sprachproduktion*. Niemeyer-Verlag: Tübingen.
- Kröger, B.J. (2003). Ein visuelles Modell der Artikulation. *Laryngo-Rhino-Otologie* 82, 402-407.
- Kröger B.J. (2007). Artikulatorische und Akustische Phonetik: Ein Kurzüberblick. Online Vorlesungsskript. www.speechtrainer.eu.
- Kröger, B.J., Birkholz, P., Neuschaefer-Rube, C. (2007). Ein neuronales Modell zur sensorischen Entwicklung des Sprechens. *Laryngo-Rhino-Otologie*, 86, 365-370.
- Krüger, H.-P. & Vollrath, M. (1996). Temporal analysis of speech patterns in the real world using the LOGOPORT. In J. Fahrenberg & M. Myrtek (Eds.), *Ambulatory Assessment*, (pp. 101-113). Seattle: Hogrefe & Huber.

- Krupp, L.B., LaRocca, N.G., Muir-Nash, J. & Steinberg, A.D. (1989). The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. *Archives of Neurology*, 46, 1121-1123.
- Kwon, O.-W., Chan, K., Hao, J. & Lee, T.-W. (2003). Emotion recognition by speech signals. *Proceedings of the Eurospeech*, 125-128.
- Ladefoged, P. (1967). *Three areas of experimental phonetics*. Oxford: Oxford University Press.
- Ladefoged, P. (1975). *A Course in Phonetics*. New York: Harcourt Brace.
- Lam, L. T. (2003). Factors associated with young drivers' car crash injury: Comparisons among learner, provisional, and full licenses. *Accident Analysis and Prevention*, 35, 913-920.
- Lang, P.J., Bradley, M.M. & Cuthbert, B.N. (1998). Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology. *Biological Psychiatry*, 44, 1248-1263.
- Langwieder, K., Spörner, A. & Hell, W. (1994). *Struktur der Unfälle mit Getöteten auf Autobahnen im Freistaat Bayern im Jahr 1991*. HUK-Verband 1994.
- Laube, I., Seeger, R., Russi, E.W., & Bloch, K.E. (1998). Accidents related to sleepiness: review of medical causes and prevention with special reference to Switzerland. *Schweiz Med Wochenschr*, 128(40), 1487-1499.
- Laukka, P., Juslin P.N. & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition and Emotion*, 19, 633-653.
- Laver, J. (1980). *The phonetic description of voice quality*. Cambridge, England: Cambridge University press.
- Lecluse, F.L.E. (1977). *Electroglottografie*. Dissertation, Universität von Rotterdam.
- Lee, C.M., Narayanan, R. & Pieraccini, R. (2002). Classifying emotion in human-machine spoken dialogs. In *Tagungsband ICME 2002, 3rd International Conference on Multimedia and Expo*, 1, 737-740.
- Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W.J.M., Roelofs, A. & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1-75.
- Levelt, W.J.M., Schriefers, H., Vorberg, D., Meyer, A.S., Pechmann, T. & Havinga, J. (1991a). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, 98, 122-142.

- Levelt, W.J.M., Schriefers, H., Vorberg, D., Meyer, A.S., Pechmann, T. & Havinga, J. (1991b). Normal and deviant lexical processing: A reply to Dell & O'Seaghdha (1991). *Psychological Review*, 98, 615-618.
- Levin, H. & Lord, W. (1975). Speech pitch frequency as an emotional state indicator. *IEEE Transactions on Systems, Man, and Cybernetics*, 5, 259-273.
- Levit, M., Huber, R., Batliner, A. & Nöth, E. (2001). Use of Prosodic Speech Characteristics for Automated Detection of Alcohol Intoxication. In Bacchiani, M., Hirschberg, J., Litman & D. (Eds.), *Proceedings of the Workshop on Prosody in Speech Recognition and Understanding 2001*, (pp. 103-106). Red Bank, New Jersey, 22-24 october, 2001.
- Lindberg, E., Carter, N., Gislason, T. & Janson, C. (2001). Role of snoring and daytime sleepiness in occupational accidents. *American Journal of Respiratory and Critical Care Medicine*, 164, 2031-2035.
- Linde, L. & Bergström, M. (1992). The effect of one night without sleep on problem-solving and immediate recall. *Psychological Research*, 54, 127-136.
- Lisper, H.-O., Laurell, H., van Loon, J. (1986). Relation Between Time to Falling Asleep Behind the Wheel on a Closed Track and Changes in Subsidiary Reaction Time During Prolonged Driving on a Motorway; *Ergonomics*; 29, 3, 445-453.
- Litman, D., Forbes, R. & Sillmann, K. (2003). Towards Emotion Prediction in Spoken Tutoring Dialogues. *Proceedings of the Human Language Technology Conference, North American Chapter of the Association for Computational Linguistics*, 25-30.
- Lloberes, P., Levy, G., Descals, C., Sampol, G., Roca, A., Sagales, T. & et al. (2000). Self-reported sleepiness while driving as a risk factor for traffic accidents in patients with obstructive sleep apnoea syndrome and in non-apnoeic snorers. *Respiratory Medicine*, 94, 971-976.
- Lüdtke, H., Wilhelm, B., Adler, M., Schaeffel, F., Wilhelm, H. (1998). Mathematical procedures in data recording and processing of pupillary fatigue waves. *Vision Research* 38, 2889- 2896.
- Lyznicki, J.M., Doege, T.C., Davis, R.M. & Williams, M.A. (1998) Sleepiness, driving, and motor vehicle crashes. *JAMA-Journal of the American Medical Association* 279(23), 1908-1913.
- Ma, T., Williamson, A., & Friswell, R. (2003). *A pilot study of fatigue on motorcycle day trips*. Sydney: NSW Injury Risk Management Research Centre.
- Macdonald, R.R. (1997). On statistical testing in psychology. *British Journal of Psychology*, 88, 333-347.

- MacKay, D.G. (1987). *The Organization of Perception and Action: A Theory for Language and other Cognitive Skills*. New York: Springer.
- Mackworth, N.H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, 1, 6-21.
- MacLean, A.W., Davies, D.R. & Thiele, K. (2003). The hazards and prevention of driving while sleepy. *Sleep Medicine Reviews*, 7, 507-521.
- Marsalek, K., Gerber, R., Golz, M., Richter, S., Sommer, D., Gundel, A. (2006). Subjective Sleepiness and Microsleep in Driving Simulation. *Somnologie*, 10, 57.
- Maruff, P., Falletti, M., Collie, A., Darby, D., & McStephen, M. (2005). Fatigue-related impairment in the speed, accuracy and variability of psychomotor performance: Comparison with blood alcohol levels. *Journal of Sleep Research*, 14, 21-27.
- Masa, J.F., Rubio, M., & Findley, L.J. (2000). *Habitually sleepy drivers have a high frequency of automobile crashes associated with respiratory disorders during sleep*. *American Journal of Respiratory and Critical Care Medicine*, 162, 1407-1412.
- Matthews, G., & Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. *The Quarterly Journal of Experimental Psychology*, 55A(2), 659-686.
- May, J. & Kline, P. (1987). Measuring the effects upon cognitive abilities of sleep loss during continuous operations. *The British Journal of Psychology*, 78, 433-455.
- Maycock, G. (1997). Sleepiness and driving: the experience of U.K. car drivers. *Accident Analysis and Prevention*, 29, 453-462.
- Mayer, G. (2007). Motorik. In H. Peter, T. Penzel & J.H. Peter (Hrsg.), *Enzyklopädie der Schlafmedizin*, (S. 764-768). Heidelberg: Springer.
- McCarthy, G. Blamire, A.M., Rothman, D.L., Gruetter, R. & Schulman, R.G. (1993). Echo planar magnetic resonance imaging studies of frontal cortex activation during word generation in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 90, 4952-4956.
- McCartt, A.T., Hammer, M.C. & Fuller, S.Z. (1998) *Study of fatigue-related driving among long-distance truck drivers in New York State*. New York: Institute for Traffic Safety Management and Research.
- McCulloch, W.S. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.

- McGilloway, S., Cowie, R., Douglas-Cowie, E., Gielen, C., Westerdijk, M. & Stroeve, S. (2000). Approaching automatic recognition of emotion from voice: a rough benchmark. *Proceedings of the ISCA Workshop on Speech and Emotion*, 207-212.
- Melamed, S. & Oksenberg, A. (2002). Excessive daytime sleepiness and risk of occupational injuries in non-shift daytime workers. *Sleep*, 25, 315-322.
- Mensík, R. (1999). Recognition of Alcohol Influence on Speech. In V. Matousek (Ed.), *Proceedings in Text, Speech and Dialogue: Second International Workshop, TSD '99*, (pp. 384-387). Heidelberg: Springer.
- Mierswa, I. (2004). *Automatisierte Merkmalsextraktion aus Audiodaten*. Unveröffentlichte Diplomarbeit. Universität Dortmund, Fachbereich Informatik.
- Mills, K. C., Spruill, S. E., Kanne, R. W., Parkman, K. M., & Zhang, Y. (2001). The influence of stimulants, sedatives, and fatigue on tunnel vision: Risk factors for driving and piloting. *Human Factors*, 43(2), 310-327.
- Mitler, M.M., Gujavarty, K.S. & Browman, C.P. (1982). Maintenance of wakefulness test: A polysomnographic technique for evaluation treatment efficacy in patients with excessive somnolence. *Electroencephalography and Clinical neurophysiology*, 53, 658-661.
- Mitler, M.M., Hajdukovic, R. & Erman, M.K. (1993). Treatment of narcolepsy with methamphetamine. *Sleep*, 16, 306-317.
- Moldofsky, H. (1992). Evaluation of daytime sleepiness. *Clinics in Chest Medicine*, 13, 417-425.
- Monien, K. (2006). *Support Vektor Maschinen als Analyseinstrument im Marketing*. Dissertation, Universität Bielefeld.
- Morris, G.O., Williams, H.L. & Lubin, A. (1960). Misperception and disorientation during sleep deprivation. *Archives of General Psychiatry*, 2, 247-254.
- Morris, R.G, Ahmed, S., Syed, G.M. & Toone, B.K. (1993). Neural correlates of planning ability: frontal lobe activation during the Tower of London Test. *Neuropsychologia*, 31, 1367-1378.
- Morrow, P. C., & Crum, M. R. (2004). Antecedents of fatigue, close calls, and crashes among commercial motor- vehicle drivers. *Journal of Safety Research*, 35, 59-69.
- Motley, M.T., Baars, B.J. & Camden, C.T. (1983). Experimental verbal slip studies: A review and an editing model of language encoding. *Communication Monographs*, 50, 79-101.
- Mousset, E. Ainsworth, W.A. Fonollosa, J.A.R. (1996). A comparison of several recent methods of fundamental frequency and voicing decision estimation. *Proceedings ICSLP*, 1273-1276

- Muttray, A., Hagenmeyer, L., Unold, B., du Prel, J.B. & Geißler, B. (2007). Videoanalyse der Schläfrigkeit von Fahrern – eine Pilotstudie. *Arbeitsmed Sozialmed Umweltmed*, 42, 184–185.
- Naègele, B., Thouvard, V., Pèpin, J.L., Lèvy, P., Bonnet, C., Perret, J.E., Pellat, J. & Feuerstein, C. (1995). Deficits of Cognitive Executive Functions in Patients with Sleep Apnea Syndrome. *Sleep*, 18, 43-52.
- Nass, C., Jonsson, I. M., Harris, H., Reaves, B., Endo, J., Brave, S. (2005). Improving automotive safety by pairing driver emotion and care voice emotion. *Conference on human factors in computing systems, 1973–1976*.
- National Transport Commission (2001). *Options for a regulatory approach to fatigue in drivers of heavy vehicles in Australia and New Zealand*. Retrieved 1/3/06 from www.ntc.gov.au
- Neppert, J.M.H. (1999). *Elemente einer Akustischen Phonetik*. Hamburg: Helmut Buske Verlag.
- Niemann, H. (1983). *Klassifikation von Mustern*. Berlin: Springer.
- Niemann, H. (1990). *Pattern Analysis and Understanding*. Heidelberg: Springer.
- Nilsson, J.P., Soderstrom, M., Karlsson, A.U., Lekander, M., Åkerstedt, T., Lindroth, N.E. & Axelsson, J. (2005). Less effective executive functioning after one night's sleep deprivation. *Journal of Sleep Research*, 14, 1–6.
- Nogueiras, A., Moreno, A., Bonafonte, A. & Mariño, J. (2001). Speech Emotion Recognition Using Hidden Markov Models. *Proceedings of the Eurospeech 2001*, 2679-2682.
- O'Shaughnessy, D. (2000). *Speech Communications: Human and Machine*, (2te Auflage). New York, USA: IEEE Press.
- Oakes, M. (1987). *Statistical inference: A commentary for the social and behavioral sciences* (reprint). Chichester, UK: Wiley.
- Oran-Gilad, T., & Hancock, P. A. (2005). *Road environment and driver fatigue*. In proceedings of the Third International Driving Symposium on Human Factors in Driving Assessment, Training and Vehicle Design.
- Paeschke, A. (2003). *Prosodische Analyse emotionaler Sprechweise*. Reihe Mündliche Kommunikation, Bd.1. Berlin: Logos Verlag.
- Pahn, J. (1988). Das Registerproblem. In D. Gless (Hrsg.), *Singen: Stimmbildung und Liedgestaltung*, (S. 29-56). Berlin: Volk und Wissen

- Pereira, C. (2000). Dimensions of emotional meaning in speech. In R. Cowie, E. Douglas-Cowie & M. Schröder (Eds.), *Proceedings of the ISCA Workshop on Speech and Emotion*, (pp. 25-28). Belfast, Northern Ireland.
- Peter, J.H. & Peter, H. (2007). Arousal. In H. Peter, T. Penzel & J.H. Peter (Hrsg.), *Enzyklopädie der Schlafmedizin*, (S. 67-70). Heidelberg: Springer.
- Peters, B., Anund, A., & Lowden, A. (2005). *Age related differences in driving and driver behaviour among fatigue drivers: A driving simulator experiment*. In abstracts from the Conference for Fatigue Management in Transportation Operations.
- Petrushin, V. (1999). Emotion in Speech: Recognition and Application to Call Centers. *Proceedings of the Conference on Artificial Neural Networks in Engineering, ANNIE '99*, 7-10.
- Pétursson, M. & Neppert, J.M.H. (1996). *Elementarbuch der Phonetik*. Hamburg: Buske.
- Pétursson, M. & Neppert, M.H. (2002). *Elementarbuch der Phonetik*. Hamburg: Buske.
- Philip, P., Sagaspe, P., Moore, N., Taillard, J., Charles, A., Guilleminault, C., & Bioulac, B. (2005). Fatigue, sleep restriction and driving performance. *Accident Analysis and Prevention*, 37, 473-478.
- Philip, P., Vervialle, F., Le Breton, P., Taillard, J. & Horne, J. A. (2001). Fatigue, alcohol, and serious road crashes in France: factorial study of national data. *British Medical Journal*, 322, 829-830.
- Pilcher, J.J. & Huffcutt, A.I. (1996). Effects of sleep deprivation on performance: a meta-analysis. *Sleep*, 19, 318-326.
- Pilutti, T. & Ulsoy, G. 1995. On-line identification of driver state for lane-keeping tasks. *Proceedings of 1995 American Control Conference*, 1, 678-681.
- Pittam, J., Gallois, C. & Callan, V. (1990). The long-term spectrum and perceived emotion. *Speech Communication*, 9, 177-187.
- Podszus, T. (2007). Atmung. In H. Peter, T. Penzel & J.H. Peter (Hrsg), *Enzyklopädie der Schlafmedizin*, (S. 80-85). Heidelberg: Springer.
- Pollard, P. (1993). How significant is significance? In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues*, 449-450. Hillsdale, NJ: Erlbaum.
- Pompino-Marschall, B. (1995). *Einführung in die Phonetik*. Berlin, New York: De Gruyter.
- Porcu, S., Ferrara, M., Urbani, L., Bellatreccia, A. & Casagrande, M. (1998). Smooth pursuit and saccadic eye movements as possible indicators of nighttime sleepiness. *Physiology & Behavior*, 65, 437-443.

- Queensland Parliamentary Travelsafe Committee (2005). *Driving on empty: Fatigue driving in Queensland – Report No. 43*. Brisbane: Legislative assembly of Queensland.
- Rabiner, C. & Schafer, R. (1978). *Digital Processing of speech signals*. Englewood Cliffs, New Jersey: Prentice-Hall International, Inc.
- Raghavan, S. (2006). *Applications of large vocabulary continuous speech recognition to fatigue detection*. Masterarbeit, Department of Electrical and Computer Engineering, Mississippi State University.
- Rechtschaffen, A. & Kales, A. (1968). *A manual of Standardized Technology, Techniques, and Scoring Systems for Sleep Stages of Human Subjects*. Technical Report, Brain Research Institute UCLA, Los Angeles.
- Reid, K. & Dawson, D. (1999). Correlation between wrist activity monitor and electrophysiological measures of sleep in a simulated shiftwork environment for younger and older subjects. *Sleep*, 22, 378-385.
- Reyner, L.A. & Horne, J.A. (1998). Falling asleep whilst driving: are drivers aware of prior sleepiness? *International Journal of Legal Medicine*, 111, 120-123.
- Richardson, G.S., Carskadon, M.A. & Flagg, W. (1978). Excessive daytime sleepiness in man: Multiple sleep latency measurement in narcoleptic and control subjects. *Electroencephalography and Clinical Neurophysiology*, 45, 621- 627.
- Richter, P. & Hacker, W. (1998). *Belastung und Beanspruchung. Stress, Ermüdung und Burnout im Arbeitsleben*. Heidelberg: Asanger.
- Riedel, T. (2003). *Klassifikation von Klangdaten*. Unveröffentlichte Diplomarbeit. Technische Universität Chemnitz, Fakultät für Informatik.
- Rigoll, G. (2004). *Pattern Recognition*. Lehrskript, Lehrstuhl für Mensch-Maschine-Kommunikation, Technische Universität München.
- Rimini-Doering, M., Manstetten, D., Altmueller, T., Ladstaetter, U. & Mahler, M. (2001). Monitoring driver drowsiness and stress in a driving simulator. *First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. 1st. Proceedings*. 58–63.
- Risser, M.R., Ware, J.C. & Freeman, F.G. (2000). Driving simulation with EEG monitoring in normal and obstructive sleep apnea patients. *Sleep*, 23, 393-398.
- Rodenbeck, A. (2007). Chronobiologie. In H. Peter, T. Penzel & J.H. Peter (Hrsg.), *Enzyklopädie der Schlafmedizin*, (S. 212-217). Heidelberg: Springer.

- Roehrs, T., Merrion, M., Pedrosi, B., Stepanski, E., Zorick, F. & Roth, T. (1995). Neuropsychological Function in Obstructive Sleep Apnea Syndrome (OSAS) Compared to Chronic Pulmonary Disease (COPD). *Sleep*, 18, 382-388.
- Rojas, R. (1993). *Theorie der neuronalen Netze*. Berlin: Springer.
- Rosa, R. R. (2001). Examining work schedules for fatigue: It's not just hours of work. In P.A. Hancock & P. A. Desmond (Eds), *Stress, workload, and fatigue* (pp513-528). New Jersey: Lawrence Erlbaum Associates.
- Rosenblatt, F. (1962). *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. New York: Spartan.
- Rosenthal, L., Roehrs, T.A. & Roth, T. (1993). The Sleep-Wake Activity Inventory: a self-report measure of daytime sleepiness. *Biological Psychiatry*, 34, 810- 820.
- Roske-Hofstrand, R. (1995). Raising awareness for fatigue among air traffic controllers. In *Proceedings of the Eighth International Symposium on Aviation Psychology*, 2, 985-991, Ohio State University, Columbus, OH, April 24-27, 1995.
- Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986). Learning internal representations by error propagation. In D.E. Rumelhart, J.L. McClelland & the PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, (pp. 318-362). Cambridge, MA, USA: MIT Press. Reprinted in Anderson and Rosenfeld (1988).
- Safford, R.R., and Rockwell, T.H. (1967). Performance decrement in twenty-four hour driving. *Highway Research Record*, 163, 68-79.
- Sagberg, F. (1999). Road accidents caused by drivers falling asleep. *Accident Analysis and Prevention*, 31, 639-649.
- Sagberg, F., Jackson, P, Krueger, H.-P., Muzet, A., Williams, A. (2004). Fatigue, sleepiness and reduced alertness as risk factors in driving. TOI Report 739/2004. Oslo.
- Saito, I, Fufiware, O., Utsuki, N., Mizumoto, C. & Arimoi, T. (1980). Hypoxia-Induced Fatal Aircraft Accident Revealed by Voice Analysis. *Aviation, Space, and Environmental Medicine*, 51, 402-406.
- Samn, S.W. & Perelli, L. (1982). *Estimation Aircraft Fatigue: A technique with Application to Airline Operations*. Brooks AFB, Texas: USAF School of Medicine. Technical report no. SAM-TR-82-21.
- Sanatamaria, J. & Chiappa K.H. (1987). The EEG of Drowsiness in Normal Adults. *Journal of Clinical Neurophysiologie*, 4, 327-382.

- Santana Diaz, A., Jammes, B., Esteve, D. & Gonzalez Mendoza, M. 2002. Driver hypovigilance diagnosis using wavelets and statistical analysis. *Proceedings on IEEE 5th International Conference on Intelligent Transportation Systems*, 162-167.
- Schade, U. & Berg, T. (1992). The role of inhibition in a spreading-activation model of language production: II. The simulational perspective. *Journal of Psycholinguistic Research*, 21, 435-462.
- Schade, U. (1988). Ein konnektionistisches Modell für die Sprachproduktion. In J. Kindermann & C. Lischka (Hrsg.), *Workshop Konnektionismus, Arbeitspapiere der GMD*, 329, 207-220. St. Augustin: GMD.
- Schäfer, T. (2007). Neurotransmitter. In H. Peter, T. Penzel & J.H. Peter (Hrsg.), *Enzyklopädie der Schlafmedizin*, (S. 820-825). Heidelberg: Springer.
- Schauten, D., Kiendel, H., Meyer, J. & Mache, D. H. (2003). Rekonstruktionsbasierte Selektion relevanter Einflussgrößen. In *Computational Intelligence, Sonderforschungsbereich „Design und Management komplexer technischer Prozesse & Systeme mit Methoden der Computational Intelligence“*, (S. 531).
- Scherer, K.R. & Oshinsky, J.S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, 1, 331-346.
- Scherer, K.R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 143-165.
- Scherer, K.R. (2001). Appraisal considered as a process of multi-level sequential checking. In K.R. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research*, (pp. 137-162). New York: Oxford University Press.
- Schleicher, R., Galley, N., Briest, S., Galley, L. (submitted). Looking tired? Blinks and saccades as indicators of fatigue in sleepiness warners. *Ergonomics*.
- Schmidt, F.L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, 1, 115- 129.
- Schölkopf, B., Müller, K.-R. & Smola, A.J. (1999). Lernen mit Kernen: Support-Vektor-Methoden zur Analyse hochdimensionaler Daten. *Informatik Forschung und Entwicklung*, 14, 154-163.
- Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M. & Gielen, S. (2001). Acoustic correlates of emotion dimensions in view of speech synthesis. *Proceedings of Eurospeech 2001*, 1, 87-90.
- Schukat-Talamazzini, E.G. (1995). *Automatische Spracherkennung – Grundlagen, statistische Modelle und effiziente Algorithmen*. Braunschweig: Vieweg.

- Schuller, B. (2006). *Automatische Emotionserkennung aus sprachlicher und manueller Interaktion*. Dissertation, Technische Universität München.
- Schuller, B., Lang, M. & Rigoll, G. (2002). Automatic Emotion Recognition by the Speech Signal. *Tagungsband SCI 2002, 6th World Multiconference on Systemics, Cybernetics and Informatics, IX*, 367-372.
- Schuller, B., Batliner, A., Seppi, D., Steidl, S., Vogt, T., Wagner, J., Devillers, L., Vidrascu, L., Amir, N., Kessous, L., Aharonson, V. (2007). The Relevance of Feature Type for the Automatic Classification of Emotional User States: Low Level Descriptors and Functionals. *Proceedings INTERSPEECH 2007*, 2253-2256.
- Sedlmeier, P. (1996). Jenseits des Signifikanztest-Rituals: Ergänzungen und Alternativen [Beyond the significance testing ritual]. *Methods of Psychological Research*, 1, 45-68.
- Seidner, W. & Wendler, J. (1982). *Die Sängerstimme. Phoniatische Grundlagen der Gesangsausbildung*. Berlin: Henschel.
- Serlin, R.C. (1987). Hypothesis testing, theory building, and the philosophy of science. *Journal of Counseling Psychology*, 34, 365-371.
- Shaver, J.P. (1993). What statistical significance testing is, and what it is not. *Journal of Experimental Education*, 61, 293-316.
- Shen, J., Barbera, J. & Shapiro C.M. (2006). Distinguishing sleepiness and fatigue: focus on definition and measurement. *Sleep Medicine Reviews*, 10, 63-76.
- Shriberg, E. (2001) To ,errrr‘ is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, 31, 153-169.
- Smith, S., Carrington, M., & Trinder, J. (2005). Subjective and predicted sleepiness while driving in young adults. *Accident Analysis and Prevention*, 37, 1066-1073.
- Sohn, D. (1993). Psychology of the scientist: LXVI. The idiots savants have taken over the psychology labs! Or why in science using the rejection of the null hypothesis as the basis for affirming the research hypothesis is unwarranted. *Psychological Reports*, 73, 1167-1175.
- Spencer, A. (1996). *Phonology*. Oxford: Blackwell.
- Stackhouse J. & Wells, B. (1997). *Children’s Speech and Literacy Difficulties – A Psycholinguistic Framework*. London: Whurr Publishers Ltd.
- Stampi, C., Stone, P. & Michimori, A. (1995). A new quantitative method for assessing sleepiness: the alpha attenuation test. *Work & Stress*, 9, 369 –376.

- Steidl, S., Hacker, C., Ruff, C., Batliner, A., Noeth, E., Haas, J. (2004). Looking at the Last Two Turns, I'd Say This Dialogue is Doomed – Measuring Dialogue Success., *Proceedings TSD (Text, Speech and Dialog)*, 629-636.
- Stemberger, J.P. (1985). An interactive activation model of language production. In A. Ellis (Ed.), *Progress in psychology of language*, (pp. 143-186). London: Erlbaum.
- Stoohs, R. A., Guilleminault, C., Itoi, A., & Dement, W. C. (1994). Traffic accidents in commercial long-haul truck drivers: the influence of sleep-disordered breathing and obesity. *Sleep*, 17, 619-623.
- Sturm, W., Herrmann, M. & Wallesch, C. (2000). *Lehrbuch der Klinischen Neuropsychologie*. Lisse: Swets & Zeitlinger Publishers.
- Summala, H. & Mikkola, T. (1994). Fatal accidents among car and truck drivers: Effects of fatigue, age, and alcohol consumption; *Hum Factors* 36:315; 106. Valtier S, Cody JT: Evaluation of internal.
- Sussman, E.D. & Morris, D.F. (1970). *An investigation of factors affecting driver alertness (Report No. VJ-11-7313)*. Washington, DC: Department of Transportation, National Highway Safety Program.
- Swann, P. (2002). *The major issue of drugs, alcohol and fatigue in heavy vehicle safety*. In proceedings from the National Heavy Vehicle Seminar, Melbourne.
- Tassi, P., Bonnefond, A., Hoeft, A., Eschenlauer, R. & Muzetand, A. (2003). Arousal and vigilance: do they differ? Study in a Sleep Inertia paradigm. *Sleep research Online*,5, 83-87.
- Teran-Santos, J., Jimenez-Gomez, A. & Cordero-Guevara, J. (1999). The association between sleep apnea and the risk of traffic accidents. *The New England Journal of Medicine*, 340, 847-851.
- Ternes, E. (1999). *Einführung in die Phonologie*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Thiffault, P., & Bergeron, J. (2003). Monotony of road environment and driver fatigue: A simulator study. *Accident Analysis and Prevention*, 35, 381-391.
- Thorpy, M. J. (1992). The clinical use of the Multiple Sleep Latency Test. The Standards of Practice Committee of the American Sleep Disorders Association. *Sleep*, 15, 268-276.
- Tornros, J. (1998). Driving behavior in a real and a simulated road tunnel - a validation study. *Accident Analysis and Prevention*, 30, 497-503.

- Torsvall, L. & Akerstedt, T., (1987). Sleepiness on the job: continuously measured EEG changes in train drivers. *Electroencephalography Clinical Neurophysiology*, 66, 502–511
- Trojan, F. (1975). *Biophonetik*. Mannheim: Bibliographisches Institut.
- Tse, J. L. M., Flin, R., & Mearns, K. (2006). Bus driver well-being: 50 years of research. *Transportation Research Part F*, 9, 89-114.
- Tzamalouka, G., Papadakaki, M., & Chliaoutakis, J. (2005). Freight transport and nondriving work duties as predictors of falling asleep at the wheel in urban areas of Crete. *Journal of Safety Research*, 36, 75-84.
- Uhmann, S. (2001). Some Arguments for the Relevance of Syntax to Same-Sentence-Self-Repair in Everyday German conversation. In M. Selting & E. Couper-Kuhlen (eds.), *Studies in Interactional Linguistics*. (pp. 373 – 404). Amsterdam: John Benjamins.
- Uldall, E. (1960). Attitudinal meanings conveyed by intonation contours. *Language and speech*, 3, 223-234.
- v. d. Ruhren, S. (2006). *Kurzfristprognosen von Verkehrszuständen auf Basis von Verfahren der Mustererkennung und von dynamischen Routensuch- und Umlegungsverfahren*. Dissertation, RWTH Aachen.
- Van der Hulst, M., Meijman, T., & Rothengatter, T. (2001). Maintaining task set under fatigue: A study of time-on-task effects in simulated driving. *Transportation research Part F*, 4, 103-118.
- Van der Merwe, A. (1997). A theoretical framework for the characterization of pathological speech sensorimotor control. In M.R. McNeil (ed.), *Clinical management of sensorimotor speech disorders*. New York: Thieme
- Van den Berg, J. (2006). Indicators and Predictors of Sleepiness. PhD Thesis, Umea University Schweden.
- Van Dongen, H. P. A., Maislin, G., Mullington, J. M & Dinges, D. F. (2003). The cumulative cost of additional wakefulness: Dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*, 26, 117-126.
- Van Dongen, H.P., Maislin, G., Mullington, J.M. & Dinges, D. F. (2003). The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*, 26, 117-126.

- Van Son, R., Pols, L. (1999). An acoustic description of consonant reduction. *Speech Communication*, 28, 125-140.
- Vapnik, V. & Cortes, C. (1995). Support Vector Networks. *Machine Learning*, 20, 273-297.
- Vapnik, V. (1995). *The Nature of Statistical Learning*. New York: Springer.
- Verdolini, K., Titze, I.R. & Rennell, A. (1994). Dependence of phonatory effort on hydration level. *Journal of Speech and Hearing Research*, 37, 1001-1007.
- Ververidis, D., Kotropoulos, C. & Pitas, L. (2004). Automatic emotional speech classification. *Proceedings of the ICASSP 2004*, International Conference on Acoustics, Speech, and Signal Processing, IEEE, 593-596.
- Vieregge, W.H. (1989). *Phonetische Transkription. Theorie und Praxis der Symbolphonetik*. Stuttgart: Steiner.
- Vieregge, W.H. (1996). *Patho-Symbolphonetik: auditive Deskription pathologischer Sprache*. Stuttgart: Steiner.
- Vlasenko, B., Schuller, B., Wendemuth, A., Rigoll, G. (2007). Combining frame and turn-level information for robust recognition of emotions within speech. *Proceedings of Interspeech*, 2249-2252.
- Vöhringer-Kuhnt, T., Baumgarten, T. Karrer, K. & Briest, S. (2004). Wierwille's method of driver drowsiness evaluation revisited. *Proceeding of International Conference on Traffic & Transport Psychology*.
- Vollrath, M. & Krüger, H.-P. (2002). Gefährdung der Verkehrssicherheit durch Alkohol, Drogen und Benzodiazepine - ein Überblick. *Suchtsmedizin in Forschung und Praxis*, 4, 175 - 183.
- Volow, M.R., & Erwin, C.W. (1973). The heart rate variability correlates of spontaneous drowsiness onset. In *International automotive engineering congress*. New York: ASME.
- Walter, P. (2000) *Verfahren der sequentiellen Merkmalsanalyse für die Mustererkennung*. Aachen: Shaker.
- Wampold, B.E., Davis, B., & Good, R.H., III (1990). Hypothesis validity of clinical research. *Journal of Consulting and Clinical Psychology*, 58, 360-367.
- Weeß, H.G. (1996). *Leistungserfassung beim Obstruktiven Schlaf-Apnoe Syndrom: aufmerksamkeitsbezogene Einschränkungen und deren Reversibilität*. Regensburg: Roderer.
- Weeß, H.G., Sauter, C., Geisler, P., Böhning, W., Wilhelm, B. & Rotte, M. (2000). Vigilanz, Einschlafneigung, Daueraufmerksamkeit, Müdigkeit, Schläfrigkeit – Diagnostische Instrumentarien zur Messung müdigkeits- und schläfrigkeitsbezogener Prozesse und deren Gütekriterien. *Somnologie*, 4, 20-38.

- Wesensten, N.J., Belenky, G., Thorne, D.R., Kautz, M.A. & Balkin, T.J. (2004). Modafinil vs. caffeine: effects on fatigue during sleep deprivation. *Aviation, Space, and Environmental Medicine*, 75, 520–525.
- Whitmore, J. & Fisher, S. (1996). Speech During Sustained Operations. *Speech Communication*, 20, 55–70.
- Widrow, B. & Hoff, M.E. (1960). Adaptive switching circuits. *IRE WESCON Convention Record*, 4, 96-104.
- Wieland, R. & Kappes, O. (2001). *Sprachausgabe im Fahrzeug. Kognitive und emotionale Determinanten bei der Nutzung eines Fahrerassistenzsystems*. Forschungsbericht für DaimlerChrysler, Stuttgart.
- Wieland, R. (2002). Emotional ergonomics: A method to evaluate vehicle`s design features. In H. Luczak, A.E. Cakir & G. Cakir (Ed.) *Proceedings of the 6th International Scientific Conference on Work With Display Units*. pp. 285-286. Berlin: ERONOMIC Institut für Arbeits- und Sozialforschung Forschungsgesellschaft mbH.
- Wieland, R., Braun, V., Hüttges, M., Timm, E. & Willamowski, A. (2002). Instrumente zur Personalauswahl und –entwicklung im Call Center [CCall Report 13]. Hamburg: Verwaltungs-Berufsgenossenschaft.
- Wieland-Eckelmann, R. & Baggen, R. (1994). Beanspruchung und Erholung im Arbeits-Erholungs-Zyklus. In R. Wieland-Eckelmann, H. Allmer, K.W. Kallus & J. Otto (Hrsg.). *Erholungsforschung. Beiträge der Emotionspsychologie, Sportpsychologie und Arbeitspsychologie*. Weinheim: Psychologie Verlagsunion.
- Wieland-Eckelmann, R. (1992). *Kognition, Emotion und psychische Beanspruchung*. Theoretische und empirische Studien zu informationsverarbeitenden Tätigkeiten. Göttingen: Hogrefe.
- Wieland-Eckelmann, R. (1996). A cognitive-actional model of self-regulation and coping. In W. Battmann & St. Dutke (Eds.), *Processes of the molar regulation of behavior* (pp. 169-187). Lengerich: Pabst.
- Wierwille, W.W. & Ellsworth, L.A. (1994). Evaluation of driver drowsiness by trained raters. *Accident Analysis and Prevention*, 26, 571-581.
- Wierwille, W.W. & Muto, W.H. (1981). Significant changes in driver vehicle response measures for extended duration simulated driving tasks. *Proceedings of the First European Annual Conference on Human Decision Making and Manual Control*, 298-314.
- Wild, D. (1996). *Die Prognose von Verkehrsstärken anhand klassifizierter Ganglinien*. Aachen: Shaker.

- Wilhelm, B., Giedke H., Lüdtke, H., Bittner, E., Hofmann A., Wilhelm, H. (2001). Daytime variations in central nervous system activation measured by a pupillographic sleepiness test. *J. Sleep. Res.* 10, 1-7.
- Wilhelm, B., Körner, A., Heldmaier, K., Moll, K., Wilhelm, H., Lüdtke, H. (2001). Normwerte des pupillographischen Schläfrigkeitstests für Frauen und Männer zwischen 20 und 60 Jahren. *Somnologie* 5, 115-120
- Williamson, A. M., & Boufous, S. (2005). *A data-matching study of the role of fatigue in work-related crashes*. In abstracts from the Conference for Fatigue Management in Transportation Operations.
- Williamson, A. M., Feyer, A, Mattick, R. P., Friswell, R., & Finlay-Brown, S. (2001). Developing measures of fatigue using an alcohol comparison to validate the effects of fatigue on performance. *Accident Analysis & Prevention*, 33, 313-326.
- Williamson, A. M., Feyer, A. M. & Friswell, R. (1994). The impact of work practices on fatigue in long distance truck drivers. *Accident Analysis and Prevention*, 28, 709-719.
- Wright, N. & McGown, A. (2001). Vigilance on the civil flight deck: incidence of sleepiness and sleep during long-haul flights and associated changes in physiological parameters. *Ergonomics*, 44, 82-106.
- Wu, H. & Yan-Go, F. (1996). Self-reported automobile accidents involving patients with obstructive sleep apnea. *Neurology*, 46, 1254-1257.
- Wylie, C.D., Schultz, T., Miller, J.C., Mitler, M.M., Mackie, R.R. (1996). *Commercial motor vehicle driver fatigue and alertness study: technical summary*. Washington DC: Federal Highway Administration.
- Yacoub, S., Simske, S., Lin, X. & Burns, J. (2003). Recognition of eEmotions in interactive voice response systems. *Proceedings of the Eurospeech 2003*, 729-732.
- Yamaguchi, M., Deguchi, M., Wakasugi, J., Ono, S., Takai, N., Higashi, T., & Mizuno, Y. (2006). Hand-held monitor of sympathetic nervous system using salivary amylase activity and its validation by driver fatigue assessment. *Biosensors and Bioelectronics*, 21, 1007-1014.
- Young, T., Blustein, J., Finn, L., & Palta, M. (1997). Sleep-disordered breathing and motor vehicle accidents in a population-based sample of employed adults. *Sleep*, 20, 608-613.
- Zell, A. (1997). *Simulation neuronaler Netze*. München: Addison Wesley Longmann.
- Zeppelzauer, M. (2005). Discrimination and retrieval of animal sound. Magisterarbeit, TU Wien

- Zhou, G., Hansen, J.H.L. & Kaiser, J.F. (1999). Methods for Stressed Speech Classification: Nonlinear TEO and Linear Speech Based Features. *IEEE International Conference on Acoustics, Speech, Signal Processing*, 4, 2087-2090.
- Zimmermann, P., Guttormsen, S., Danuser, B. & Gomez, P. (2003). Affective Computing - A rationale for measuring mood with mouse and keyboard. *International Journal of Occupational Safety and Ergonomics*, 9, 539-551.
- Zölzer, U. (2002). *DAFX - Digital Audio Effects*. John Wiley Sons, Ltd., England.
- Zulley, J., Crönlein, T., Hell, W., & Langwieder, K. (1995). Einschlafen am Steuer: Hauptursache schwerer Verkehrsunfälle. *Wiener Medizinische Wochenzeitschrift*, 17/18, 473.
- Zwicker, E. & Fastl, H. (1990). Psychoacoustics. Facts and Models. In *Springer Series in Information Sciences*, 22. Berlin: Springer-Verlag.

9 ABKÜRZUNGSVERZEICHNIS

Abkürzung	Bedeutung
AAT	Alpha Attenuation Test
ACF	Autocorrelation Function
ANN	Artificial Neural Net
ASA	Acoustic Sleepiness Analysis
ASR	Automatic Speech Recognition
DFT	Discrete Fourier Transformation
EDA	Elektrodermale Aktivität
EEG	Elektroenzephalogramm
EOG	Elektrookulogramm
F0	Fundamental Frequenz
F1-F5	Formant 1 bis Formant 5
HCI	Human Computer Interaction
HNR	Harmonics-to-Noise Ratio
ICA	Independent Component Analysis
KDS	Karolinska Drowsiness Scoring
KSS	Karolinska Sleepiness Scale
KNN	K-Nearest-Neighbour
LAR	Log Area Ratios
LDA	Linear Discriminant Analysis
LPC	Linear Predictive Coding
LPCC	Linear Predictive Cepstral Coefficient
MFCC	Mel Frequency Cepstral Coefficient
MLP	Multilayer Perceptron
MSLT	Multipler Schlaflatenz Test
MWT	Maintenance of Wakefulness Test
LOSO	Leave-One-Speaker-Out

LTAS	Long Term Average Spectrum
PARCOR	Partial Autocorrelation
PCA	Principal Component Analysis
PSD	Power Spectral Density
PST	Pupillographischer Schläfrigkeitstest
PSQI	Pittsburgh Sleep Quality index
PVT	Psychomotor Vigilance Test
SSS	Stanford Sleepiness Scale
SVM	Support Vector Machine

NOMENKLATUR

Die verwendete Nomenklatur ist in Anlehnung an die international jeweils gebräuchlichste Notation gewählt. Auf Grund der Vielzahl diverser behandelter Verfahren kommt es dabei teilweise zur mehrfachen Besetzung des gleichen Symbols. Dieser Umstand wurde dann zu Gunsten eines vereinfachten Vergleichs mit der Literatur in Kauf genommen.

Nomenklatur	Bedeutung
Signalverarbeitung	
x_i	Element i des Vektors x mit $i= 1,\dots,N$
f	Frequenz
T	Periodendauer eines periodischen Signals
$H(f)$	Allgemeine Übertragungsfunktion
E_t	Energie im Rahmen mit Zeitindex t
$E_t(f)$	Spektrale Bandenergie bei der Mittenfrequenz f im Rahmen mit Zeitindex t
$w[n]$	Allgemeine Fensterfunktion
$w_{\text{Ham}} [n]$	Hamming Fensterfunktion
N	Gesamtanzahl der Abtastwerte je Rahmen
n	Diskrete Zeitvariable innerhalb eines Rahmens
t	Diskrete Zeitindex auf Rahmenebene
$s(t)$	Zeitsignal
$s[n]$	Zeitsignal mit diskretisiertes Zeitachse
$s_t[n]$	Signalwert zum Abtastwert n im Rahmen mit Zeitindex t
Mustererkennung	
Ω	Menge der Klassen
Ω_k	Musterklasse mit Ω_k Element von Ω und $k = 1,\dots, K$
j	Partitionskoeffizient der j -fachen Kreuzvalidierung
$H(w,b)$	Hypereben mit Normalenvektor w und Bias b (in SVM)
K^Φ	Kernel-Funktion (in SVM)
x^{sv}	Stützvektor (Support Vector) (in SVM)
$r(x)$	Punktabstand von x zur Hypereben H (in SVM)

ξ	Schlupfvariable (in SVM)
G	Fehlergewichtungsfaktor (in SVM)
k	Verstärkungsfaktor des Sigmoid-Kernels (in SVM)
Θ	Offset des Sigmoid-Kernels (in SVM)
$w_{i,j}$	Gewicht $w_{i,j}$ (in MLP)
W	Gewichtmatrix
$T(u)$	Transferfunktion
α	Steigungsparameter der Sigmoid-Funktion
$F(x, W)$	Fehlerfunktion für das Lernen in Abhängigkeit von x und W
β	Schrittweite beim Gradientenabstieg
Sprachverarbeitung	
E_k	Kuzzeitenergie
$F_{0,t}$	Fundamentalfrequenz im Rahmen mit Zeitindex t
FX_t	Formantfrequenz im Rahmen mit Zeitindex t
HNR_t	Harmonische Ausprägtheit im Rahmen mit Zeitindex t
$MFCCX_t$	MFCC Koeffizient X im Rahmen mit Zeitindex t
$S_{flux,t}$	Spektraler Fluss im Rahmen mit Zeitindex t

**ANHANG A:
SPEECH ASSESSMENT METHODS PHONETIC ALPHABET**

Tabelle A-1: SAMPA – deutsch: ein ASCII-basiertes phonetisches Alphabet

Plosive: [p] [b] [t] [d] [k] [g]		
Symbol	Beispiel	Transkription
[p]	Pein	[paIn]
[b]	Bein	[baIn]
[t]	Teich	[taIC]
[d]	Deich	[daIC]
[k]	Kunst	[kUnst]
[g]	Gunst	[gUnst]
Affrikaten: [pf] [ts] [tS]		
Symbol	Beispiel	Transkription
[pf]	Pfahl	[pfa:l]
[ts]	Zahl	[tsa:l]
[tS]	deutsch	[dOI tS]
Frikative: [f] [v] [s] [z] [S] [Z] [C] [x] [r] [h]		
Symbol	Beispiel	Transkription
[f]	fast	[fast]
[v]	was	[vas]
[s]	das	[das]
[z]	sein	[zaIn]
[S]	Schein	[SaIn]
[Z]	Genie	[Ze/ni]
[C]	ich	[IC]
[x]	ach	[ax]
[r]	rein	[raIn]
[h]	Hand	[hant]
Sonoranten (Nasale, Liquide, Glides): [m] [n] [N] [l] [j]		
Symbol	Beispiel	Transkription
[m]	mein	[maIn]
[n]	nein	[naIn]
[N]	lang	[laN]
[l]	Laut	[laUt]
[j]	ja	[ja:]
Ungespannte Vokale: [I] [Y] [9] [E] [U] [O] [a]		
Symbol	Beispiel	Transkription
[I]	Sitz	[zIts]
[Y]	hübsch	[hYpS]
[9]	plötzlich	[/pl9tslIC]
[E]	Bett	[bEt]
[U]	Schutz	[SÜts]
[O]	Trotz	[trOtS]
[a]	Satz	[zats]

Table A-2: SAMPA – deutsch: ein ASCII-basiertes phonetisches Alphabet

Gespannte Vokale und Diphthonge: [i] [y] [e] [ɛ] [ɛ:] [u] [o] [a:] [aI] [aU] [OI]		
Symbol	Beispiel	Transkription
[i]	Lied	[lit]
[y]	süß	[zys]
[e]	Beet	[bet]
[ɛ]	blöd	[blɛt]
[ɛ:]	spät	[Spɛ:t]
[u]	Blut	[blut]
[o]	rot	[rot]
[a:]	Tat	[ta:t]
[aI]	Eis	[aIs]
[aU]	Haus	[haUs]
[OI]	Kreuz	[krOIts]
Schwa: [ə]		
Symbol	Beispiel	Transkription
[ə]	bitte	["bitə]
[ə6, 6]	bitter	[bitə6, bit6]
[I6]	Wirt	[vI6t]
[Y6]	Türke	[/tY6kə]
[96]	Wörter	[/v96t6]
[E6]	Berg	[bE6k]
[U6]	kurz	[kU6ts]
[O6]	dort	[dO6t]
[a6]	Art	[a6t]
[i6]	Tier	[ti6]
[y6]	Tür	[ty6]
[e6]	Meer	[me6]
[ɛ6]	Nadelöhr	[/nadə126]
[ɛ:6]	Bär	[bɛ:6]
[u6]	Kur	[ku6]
[o6]	Ohr	[o6]

ANHANG B:

**Vokalspezifische Ergänzungen: Zusammenhänge zwischen Schläfrigkeit
und akustischen Kennzahlen der Vokale [ə], [a:], [u:], [e:], [o:] und [i:]**

Tabelle B-1: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rU-fen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE

f0_min	f0_timemin	f0_max	f0_timemax	f0_mean	f0_sd	
	-.07	.18	-.06	.09	-.07	.05

Satz04_tE_z

f0_min	f0_timemin	f0_max	f0_timemax	f0_mean	f0_sd	
	-.06	.10	-.06	.05	-.06	.05

Satz06_nA

f0_min	f0_timemin	f0_max	f0_timemax	f0_mean	f0_sd	
	.00	.09	.02	.04	.00	.01

Satz06_nA_z

f0_min	f0_timemin	f0_max	f0_timemax	f0_mean	f0_sd	
	.07	-.12	.06	.03	.03	-.03

Satz09_rUf

f0_min	f0_timemin	f0_max	f0_timemax	f0_mean	f0_sd	
	-.11	.07	-.09	.07	-.09	.00

Satz09_rUf_z

f0_min	f0_timemin	f0_max	f0_timemax	f0_mean	f0_sd	
	-.24	.00	-.23	.10	-.23	-.09

Satz09_dE

f0_min	f0_timemin	f0_max	f0_timemax	f0_mean	f0_sd	
	-.13	.10	-.14	-.17	-.14	-.13

Satz09_dE_z

f0_min	f0_timemin	f0_max	f0_timemax	f0_mean	f0_sd	
	-.34	.19	-.31	-.07	-.32	.00

Satz09_nOt

f0_min	f0_timemin	f0_max	f0_timemax	f0_mean	f0_sd	
	-.14	-.08	-.24	-.04	-.16	-.25

Satz09_nOt_z

f0_min	f0_timemin	f0_max	f0_timemax	f0_mean	f0_sd	
	-.13	-.18	-.33	-.08	-.16	-.36

Satz09_dIen

f0_min	f0_timemin	f0_max	f0_timemax	f0_mean	f0_sd	
	.04	.23	.03	-.03	.03	-.02

Satz09_dIen_z

f0_min	f0_timemin	f0_max	f0_timemax	f0_mean	f0_sd	
	.06	.10	.05	.11	.05	-.02

Tabelle B-2: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rUfen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE

f0_25quant	f0_75quant	f0_mintime_rel	f0_maxtime_rel	f0_range	f0_slope
-0.07	-0.06	.15	-0.05	.09	-0.01

Satz04_tE_z

f0_25quant	f0_75quant	f0_mintime_rel	f0_maxtime_rel	f0_range	f0_slope
-0.06	-0.06	.09	-0.06	.07	.01

Satz06_nA

f0_25quant	f0_75quant	f0_mintime_rel	f0_maxtime_rel	f0_range	f0_slope
.00	.00	-0.02	.06	.08	-0.10

Satz06_nA_z

f0_25quant	f0_75quant	f0_mintime_rel	f0_maxtime_rel	f0_range	f0_slope
.06	.03	-0.02	.10	.02	-0.08

Satz09_rUf

f0_25quant	f0_75quant	f0_mintime_rel	f0_maxtime_rel	f0_range	f0_slope
-0.10	-0.09	.03	.00	.00	-0.12

Satz09_rUf_z

f0_25quant	f0_75quant	f0_mintime_rel	f0_maxtime_rel	f0_range	f0_slope
-.24	-.23	-0.08	.08	-0.09	-0.19

Satz09_dE

f0_25quant	f0_75quant	f0_mintime_rel	f0_maxtime_rel	f0_range	f0_slope
-0.13	-0.14	.21	-0.16	-0.12	-0.12

Satz09_dE_z

f0_25quant	f0_75quant	f0_mintime_rel	f0_maxtime_rel	f0_range	f0_slope
-.33	-.32	.21	-0.16	.02	-0.02

Satz09_nOt

f0_25quant	f0_75quant	f0_mintime_rel	f0_maxtime_rel	f0_range	f0_slope
-0.15	-.23	-0.05	.01	-.27	-.33

Satz09_nOt_z

f0_25quant	f0_75quant	f0_mintime_rel	f0_maxtime_rel	f0_range	f0_slope
-0.15	-.30	-.24	.02	-.36	-.41

Satz09_dIen

f0_25quant	f0_75quant	f0_mintime_rel	f0_maxtime_rel	f0_range	f0_slope
.04	.04	.22	-0.01	-0.02	.09

Satz09_dIen_z

f0_25quant	f0_75quant	f0_mintime_rel	f0_maxtime_rel	f0_range	f0_slope
.06	.05	.06	.03	-0.03	.06

Tabelle B-3: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rUfen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE						
	f0_slope_octjump	Jitter	Shimmer	HNR	f1_min	f1_mintime_rel
	.08	-0.04	-0.14		.07	.01
Satz04_tE_z						
	f0_slope_octjump	Jitter	Shimmer	HNR	f1_min	f1_mintime_rel
	.05	-0.07	-0.03		.04	-0.03
Satz06_nA						
	f0_slope_octjump	Jitter	Shimmer	HNR	f1_min	f1_mintime_rel
	-0.09	-0.09	-0.14		.13	-0.09
Satz06_nA_z						
	f0_slope_octjump	Jitter	Shimmer	HNR	f1_min	f1_mintime_rel
	-0.06	-0.04	.10		.00	-.21
Satz09_rUf						
	f0_slope_octjump	Jitter	Shimmer	HNR	f1_min	f1_mintime_rel
	-0.14	-0.03	.30		-0.14	.09
Satz09_rUf_z						
	f0_slope_octjump	Jitter	Shimmer	HNR	f1_min	f1_mintime_rel
	-0.17	-0.05	.40		-.25	.19
Satz09_dE						
	f0_slope_octjump	Jitter	Shimmer	HNR	f1_min	f1_mintime_rel
	-0.08	-0.06	-0.01		-0.12	-0.02
Satz09_dE_z						
	f0_slope_octjump	Jitter	Shimmer	HNR	f1_min	f1_mintime_rel
	.06	.05	.14		-.29	-0.07
Satz09_nOt						
	f0_slope_octjump	Jitter	Shimmer	HNR	f1_min	f1_mintime_rel
	-.29	-.33	-0.15		.18	-0.01
Satz09_nOt_z						
	f0_slope_octjump	Jitter	Shimmer	HNR	f1_min	f1_mintime_rel
	-.29	-.23	.00		.06	.10
Satz09_dIen						
	f0_slope_octjump	Jitter	Shimmer	HNR	f1_min	f1_mintime_rel
	.13	.11	.03		-0.05	.29
Satz09_dIen_z						
	f0_slope_octjump	Jitter	Shimmer	HNR	f1_min	f1_mintime_rel
	.13	.08	.22		-0.16	.10

Tabelle B-4: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rU-fen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE						
	f1_max	f1_maxtime_rel	f1_range	f1_mean	f1_sd	f1_bandw_q25
	.01	-.03	.00	.01	-.15	-.07
Satz04_tE_z						
	f1_max	f1_maxtime_rel	f1_range	f1_mean	f1_sd	f1_bandw_q25
	-.02	-.02	.02	-.04	-.02	-.07
Satz06_nA						
	f1_max	f1_maxtime_rel	f1_range	f1_mean	f1_sd	f1_bandw_q25
	-.08	.12	.01	-.09	.00	-.05
Satz06_nA_z						
	f1_max	f1_maxtime_rel	f1_range	f1_mean	f1_sd	f1_bandw_q25
	-.10	.16	-.03	-.11	.08	-.12
Satz09_rUf						
	f1_max	f1_maxtime_rel	f1_range	f1_mean	f1_sd	f1_bandw_q25
	.16	.00	.06	.22	.03	.02
Satz09_rUf_z						
	f1_max	f1_maxtime_rel	f1_range	f1_mean	f1_sd	f1_bandw_q25
	.18	-.24	-.01	.23	-.04	.05
Satz09_dE						
	f1_max	f1_maxtime_rel	f1_range	f1_mean	f1_sd	f1_bandw_q25
	-.04	-.12	-.07	-.02	-.09	-.01
Satz09_dE_z						
	f1_max	f1_maxtime_rel	f1_range	f1_mean	f1_sd	f1_bandw_q25
	-.07	-.12	-.04	-.07	-.05	-.05
Satz09_nOt						
	f1_max	f1_maxtime_rel	f1_range	f1_mean	f1_sd	f1_bandw_q25
	-.02	.08	-.06	.04	.01	.03
Satz09_nOt_z						
	f1_max	f1_maxtime_rel	f1_range	f1_mean	f1_sd	f1_bandw_q25
	.20	.25	-.08	.19	.17	.09
Satz09_dIen						
	f1_max	f1_maxtime_rel	f1_range	f1_mean	f1_sd	f1_bandw_q25
	.08	.08	-.01	.26	.00	.00
Satz09_dIen_z						
	f1_max	f1_maxtime_rel	f1_range	f1_mean	f1_sd	f1_bandw_q25
	.02	-.04	.00	.15	-.01	.18

Tabelle B-5: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rUfen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE						
	f1_bandw_q50	f1_bandw_q75	f2_min	f2_mintime_rel	f2_max	f2_maxtime_rel
	-0.05	-0.07	.28	.21	.34	-.20
Satz04_tE_z						
	f1_bandw_q50	f1_bandw_q75	f2_min	f2_mintime_rel	f2_max	f2_maxtime_rel
	-0.04	-0.06	.06	.05	.26	-.08
Satz06_nA						
	f1_bandw_q50	f1_bandw_q75	f2_min	f2_mintime_rel	f2_max	f2_maxtime_rel
	-0.06	-0.10	.23	.04	.09	-.09
Satz06_nA_z						
	f1_bandw_q50	f1_bandw_q75	f2_min	f2_mintime_rel	f2_max	f2_maxtime_rel
	-.11	-.13	.17	-.04	-.07	-.03
Satz09_rUf						
	f1_bandw_q50	f1_bandw_q75	f2_min	f2_mintime_rel	f2_max	f2_maxtime_rel
	-.09	-.12	.12	.12	.04	.11
Satz09_rUf_z						
	f1_bandw_q50	f1_bandw_q75	f2_min	f2_mintime_rel	f2_max	f2_maxtime_rel
	-.01	-.17	.17	-.07	.15	.16
Satz09_dE						
	f1_bandw_q50	f1_bandw_q75	f2_min	f2_mintime_rel	f2_max	f2_maxtime_rel
	-.09	-.13	.04	.06	-.24	-.19
Satz09_dE_z						
	f1_bandw_q50	f1_bandw_q75	f2_min	f2_mintime_rel	f2_max	f2_maxtime_rel
	-.12	-.12	-.15	.15	-.19	-.24
Satz09_nOt						
	f1_bandw_q50	f1_bandw_q75	f2_min	f2_mintime_rel	f2_max	f2_maxtime_rel
	.03	-.06	.16	.18	.07	-.15
Satz09_nOt_z						
	f1_bandw_q50	f1_bandw_q75	f2_min	f2_mintime_rel	f2_max	f2_maxtime_rel
	.06	.01	.09	.06	.16	.05
Satz09_dIen						
	f1_bandw_q50	f1_bandw_q75	f2_min	f2_mintime_rel	f2_max	f2_maxtime_rel
	-.05	-.04	.08	-.11	.06	.19
Satz09_dIen_z						
	f1_bandw_q50	f1_bandw_q75	f2_min	f2_mintime_rel	f2_max	f2_maxtime_rel
	.18	.14	-.18	-.19	.00	.11

Tabelle B-6: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rUfen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE

f2_range	f2_mean	f2_sd	f2_bandw_q25	f2_bandw_q50	f2_bandw_q75
.01	.33	-.17	-.09	-.15	-.14

Satz04_tE_z

f2_range	f2_mean	f2_sd	f2_bandw_q25	f2_bandw_q50	f2_bandw_q75
.18	.18	.11	-.03	.08	.09

Satz06_nA

f2_range	f2_mean	f2_sd	f2_bandw_q25	f2_bandw_q50	f2_bandw_q75
-.20	.15	-.26	-.08	-.11	-.10

Satz06_nA_z

f2_range	f2_mean	f2_sd	f2_bandw_q25	f2_bandw_q50	f2_bandw_q75
-.26	.04	-.23	-.09	-.13	-.16

Satz09_rUf

f2_range	f2_mean	f2_sd	f2_bandw_q25	f2_bandw_q50	f2_bandw_q75
-.04	.11	-.09	.09	.06	.02

Satz09_rUf_z

f2_range	f2_mean	f2_sd	f2_bandw_q25	f2_bandw_q50	f2_bandw_q75
.01	.25	-.06	.15	.09	.02

Satz09_dE

f2_range	f2_mean	f2_sd	f2_bandw_q25	f2_bandw_q50	f2_bandw_q75
-.15	-.13	-.12	.03	-.07	-.08

Satz09_dE_z

f2_range	f2_mean	f2_sd	f2_bandw_q25	f2_bandw_q50	f2_bandw_q75
.04	-.27	.05	.04	.10	.09

Satz09_nOt

f2_range	f2_mean	f2_sd	f2_bandw_q25	f2_bandw_q50	f2_bandw_q75
.00	.25	-.01	.11	.07	.01

Satz09_nOt_z

f2_range	f2_mean	f2_sd	f2_bandw_q25	f2_bandw_q50	f2_bandw_q75
.10	.26	.04	.08	.11	.04

Satz09_dIen

f2_range	f2_mean	f2_sd	f2_bandw_q25	f2_bandw_q50	f2_bandw_q75
-.05	.16	-.06	.01	-.02	-.07

Satz09_dIen_z

f2_range	f2_mean	f2_sd	f2_bandw_q25	f2_bandw_q50	f2_bandw_q75
.15	-.07	.07	-.01	.10	-.02

Tabelle B-7: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rU-fen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE						
	f3_min	f3_mintime_rel	f3_max	f3_maxtime_rel	f3_range	f3_mean
	.08	-.09	.03	.10	-.07	.06
Satz04_tE_z						
	f3_min	f3_mintime_rel	f3_max	f3_maxtime_rel	f3_range	f3_mean
	-.04	.05	.04	.03	.07	-.01
Satz06_nA						
	f3_min	f3_mintime_rel	f3_max	f3_maxtime_rel	f3_range	f3_mean
	.27	-.09	.12	-.06	-.12	.22
Satz06_nA_z						
	f3_min	f3_mintime_rel	f3_max	f3_maxtime_rel	f3_range	f3_mean
	.21	-.10	.10	-.01	-.08	.18
Satz09_rUf						
	f3_min	f3_mintime_rel	f3_max	f3_maxtime_rel	f3_range	f3_mean
	.11	.07	-.04	.10	-.09	.06
Satz09_rUf_z						
	f3_min	f3_mintime_rel	f3_max	f3_maxtime_rel	f3_range	f3_mean
	-.02	.06	.07	.01	.07	.15
Satz09_dE						
	f3_min	f3_mintime_rel	f3_max	f3_maxtime_rel	f3_range	f3_mean
	.08	.35	-.23	-.27	-.24	-.05
Satz09_dE_z						
	f3_min	f3_mintime_rel	f3_max	f3_maxtime_rel	f3_range	f3_mean
	-.04	.16	-.13	-.12	-.08	-.07
Satz09_nOt						
	f3_min	f3_mintime_rel	f3_max	f3_maxtime_rel	f3_range	f3_mean
	.19	-.05	.08	.03	-.02	.29
Satz09_nOt_z						
	f3_min	f3_mintime_rel	f3_max	f3_maxtime_rel	f3_range	f3_mean
	.08	-.02	.17	.20	.12	.27
Satz09_dIen						
	f3_min	f3_mintime_rel	f3_max	f3_maxtime_rel	f3_range	f3_mean
	.07	-.14	-.03	-.04	-.08	.07
Satz09_dIen_z						
	f3_min	f3_mintime_rel	f3_max	f3_maxtime_rel	f3_range	f3_mean
	-.17	-.11	.06	.06	.16	-.05

Tabelle B-8: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rU-fen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE						
	f3_sd	f3_bandw_q25	f3_bandw_q50	f3_bandw_q75	f4_min	f4_mintime_rel
	-.17	-.10	-.13	-.08	.01	.06
Satz04_tE_z						
	f3_sd	f3_bandw_q25	f3_bandw_q50	f3_bandw_q75	f4_min	f4_mintime_rel
	.07	-.06	-.03	.03	-.04	-.04
Satz06_nA						
	f3_sd	f3_bandw_q25	f3_bandw_q50	f3_bandw_q75	f4_min	f4_mintime_rel
	-.12	.03	-.01	-.11	.10	.05
Satz06_nA_z						
	f3_sd	f3_bandw_q25	f3_bandw_q50	f3_bandw_q75	f4_min	f4_mintime_rel
	-.04	.11	.10	-.01	.09	.06
Satz09_rUf						
	f3_sd	f3_bandw_q25	f3_bandw_q50	f3_bandw_q75	f4_min	f4_mintime_rel
	-.10	-.27	-.30	-.32	.11	.08
Satz09_rUf_z						
	f3_sd	f3_bandw_q25	f3_bandw_q50	f3_bandw_q75	f4_min	f4_mintime_rel
	.07	-.01	-.03	-.02	.12	.11
Satz09_dE						
	f3_sd	f3_bandw_q25	f3_bandw_q50	f3_bandw_q75	f4_min	f4_mintime_rel
	-.25	-.11	-.16	-.28	-.08	.11
Satz09_dE_z						
	f3_sd	f3_bandw_q25	f3_bandw_q50	f3_bandw_q75	f4_min	f4_mintime_rel
	-.11	-.12	-.14	-.11	-.24	-.06
Satz09_nOt						
	f3_sd	f3_bandw_q25	f3_bandw_q50	f3_bandw_q75	f4_min	f4_mintime_rel
	.03	.10	.02	-.05	.03	.06
Satz09_nOt_z						
	f3_sd	f3_bandw_q25	f3_bandw_q50	f3_bandw_q75	f4_min	f4_mintime_rel
	.13	.01	-.03	-.05	.12	.02
Satz09_dIen						
	f3_sd	f3_bandw_q25	f3_bandw_q50	f3_bandw_q75	f4_min	f4_mintime_rel
	-.08	-.04	-.18	-.11	-.03	.11
Satz09_dIen_z						
	f3_sd	f3_bandw_q25	f3_bandw_q50	f3_bandw_q75	f4_min	f4_mintime_rel
	.15	-.05	-.24	-.08	-.18	.02

Tabelle B-9: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rUfen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE						
	f4_max	f4_maxtime_rel	f4_range	f4_mean	f4_sd	f4_bandw_q25
	.10	.01	.10	.01	-.02	-.03
Satz04_tE_z						
	f4_max	f4_maxtime_rel	f4_range	f4_mean	f4_sd	f4_bandw_q25
	.14	-.01	.18	-.04	.15	-.13
Satz06_nA						
	f4_max	f4_maxtime_rel	f4_range	f4_mean	f4_sd	f4_bandw_q25
	.00	-.36	-.10	-.09	-.15	-.01
Satz06_nA_z						
	f4_max	f4_maxtime_rel	f4_range	f4_mean	f4_sd	f4_bandw_q25
	-.17	-.27	-.26	-.11	-.26	.13
Satz09_rUf						
	f4_max	f4_maxtime_rel	f4_range	f4_mean	f4_sd	f4_bandw_q25
	-.07	.01	-.14	.22	-.16	-.10
Satz09_rUf_z						
	f4_max	f4_maxtime_rel	f4_range	f4_mean	f4_sd	f4_bandw_q25
	-.08	.02	-.15	.23	-.20	.02
Satz09_dE						
	f4_max	f4_maxtime_rel	f4_range	f4_mean	f4_sd	f4_bandw_q25
	-.25	.01	-.18	-.02	-.17	.10
Satz09_dE_z						
	f4_max	f4_maxtime_rel	f4_range	f4_mean	f4_sd	f4_bandw_q25
	-.17	-.05	-.03	-.07	-.04	.05
Satz09_nOt						
	f4_max	f4_maxtime_rel	f4_range	f4_mean	f4_sd	f4_bandw_q25
	.12	.03	.10	.04	.12	-.06
Satz09_nOt_z						
	f4_max	f4_maxtime_rel	f4_range	f4_mean	f4_sd	f4_bandw_q25
	.21	.09	.14	.19	.15	.09
Satz09_dIen						
	f4_max	f4_maxtime_rel	f4_range	f4_mean	f4_sd	f4_bandw_q25
	.07	.10	.09	.26	.10	-.15
Satz09_dIen_z						
	f4_max	f4_maxtime_rel	f4_range	f4_mean	f4_sd	f4_bandw_q25
	.13	.18	.22	.15	.18	-.09

Tabelle B-10: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rU-fen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE

f4_bandw_q50	f4_bandw_q75	f5_min	f5_mintime_rel	f5_max	f5_maxtime_rel
-0.01	.01	-0.05	.19	.01	-0.06

Satz04_tE_z

f4_bandw_q50	f4_bandw_q75	f5_min	f5_mintime_rel	f5_max	f5_maxtime_rel
-.10	-.05	-.27	.14	-.24	-.03

Satz06_nA

f4_bandw_q50	f4_bandw_q75	f5_min	f5_mintime_rel	f5_max	f5_maxtime_rel
-0.01	-0.03	.19	-.21	.20	.09

Satz06_nA_z

f4_bandw_q50	f4_bandw_q75	f5_min	f5_mintime_rel	f5_max	f5_maxtime_rel
.14	.11	.21	-.14	.23	-.03

Satz09_rUf

f4_bandw_q50	f4_bandw_q75	f5_min	f5_mintime_rel	f5_max	f5_maxtime_rel
-.12	-.09	-.11	.09	-.12	-.29

Satz09_rUf_z

f4_bandw_q50	f4_bandw_q75	f5_min	f5_mintime_rel	f5_max	f5_maxtime_rel
-0.01	.06	-0.05	.02	-.17	-.22

Satz09_dE

f4_bandw_q50	f4_bandw_q75	f5_min	f5_mintime_rel	f5_max	f5_maxtime_rel
.08	.12	.04	.24	-.08	-.11

Satz09_dE_z

f4_bandw_q50	f4_bandw_q75	f5_min	f5_mintime_rel	f5_max	f5_maxtime_rel
.08	.14	-0.05	.18	-.09	-.02

Satz09_nOt

f4_bandw_q50	f4_bandw_q75	f5_min	f5_mintime_rel	f5_max	f5_maxtime_rel
-0.01	-0.07	.05	-.08	.04	.01

Satz09_nOt_z

f4_bandw_q50	f4_bandw_q75	f5_min	f5_mintime_rel	f5_max	f5_maxtime_rel
.16	.06	-0.06	-.30	-.06	-.09

Satz09_dIen

f4_bandw_q50	f4_bandw_q75	f5_min	f5_mintime_rel	f5_max	f5_maxtime_rel
.09	.15	-0.06	-.10	.04	.15

Satz09_dIen_z

f4_bandw_q50	f4_bandw_q75	f5_min	f5_mintime_rel	f5_max	f5_maxtime_rel
.07	.17	-0.09	.04	.21	.29

Tabelle B-11: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rUfen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE						
f5_range	f5_mean	f5_sd	f5_bandw_q25	f5_bandw_q50	f5_bandw_q75	
.12	-.02	-.03	-.01	.02	.02	
Satz04_tE_z						
f5_range	f5_mean	f5_sd	f5_bandw_q25	f5_bandw_q50	f5_bandw_q75	
.04	-.26	-.26	.19	.15	-.01	
Satz06_nA						
f5_range	f5_mean	f5_sd	f5_bandw_q25	f5_bandw_q50	f5_bandw_q75	
.04	.19	-.02	.07	.09	.10	
Satz06_nA_z						
f5_range	f5_mean	f5_sd	f5_bandw_q25	f5_bandw_q50	f5_bandw_q75	
-.01	.24	-.09	.19	.18	.13	
Satz09_rUf						
f5_range	f5_mean	f5_sd	f5_bandw_q25	f5_bandw_q50	f5_bandw_q75	
.00	-.12	-.09	-.02	-.03	-.05	
Satz09_rUf_z						
f5_range	f5_mean	f5_sd	f5_bandw_q25	f5_bandw_q50	f5_bandw_q75	
-.13	-.09	-.15	.02	.02	.02	
Satz09_dE						
f5_range	f5_mean	f5_sd	f5_bandw_q25	f5_bandw_q50	f5_bandw_q75	
-.14	-.03	-.21	-.11	-.06	-.03	
Satz09_dE_z						
f5_range	f5_mean	f5_sd	f5_bandw_q25	f5_bandw_q50	f5_bandw_q75	
-.05	-.09	.01	-.10	.01	.04	
Satz09_nOt						
f5_range	f5_mean	f5_sd	f5_bandw_q25	f5_bandw_q50	f5_bandw_q75	
-.02	.08	.13	.11	.07	.05	
Satz09_nOt_z						
f5_range	f5_mean	f5_sd	f5_bandw_q25	f5_bandw_q50	f5_bandw_q75	
.00	-.02	.08	.04	.00	-.04	
Satz09_dIen						
f5_range	f5_mean	f5_sd	f5_bandw_q25	f5_bandw_q50	f5_bandw_q75	
.11	.00	.25	.02	.08	.11	
Satz09_dIen_z						
f5_range	f5_mean	f5_sd	f5_bandw_q25	f5_bandw_q50	f5_bandw_q75	
.28	.06	.29	-.06	.00	.01	

Tabelle B-12: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rU-fen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE							
	df1f2	df1f3	df1f4	df1f5	df2f3	df2f4	
		-.33	-.06	-.08	-.03	.09	.06
Satz04_tE_z							
	df1f2	df1f3	df1f4	df1f5	df2f3	df2f4	
		-.18	.01	-.09	.21	.08	-.04
Satz06_nA							
	df1f2	df1f3	df1f4	df1f5	df2f3	df2f4	
		-.27	-.25	-.09	-.20	-.12	.01
Satz06_nA_z							
	df1f2	df1f3	df1f4	df1f5	df2f3	df2f4	
		-.10	-.23	.01	-.23	-.16	.04
Satz09_rUf							
	df1f2	df1f3	df1f4	df1f5	df2f3	df2f4	
		-.09	-.02	.00	.16	.12	.13
Satz09_rUf_z							
	df1f2	df1f3	df1f4	df1f5	df2f3	df2f4	
		-.23	-.10	-.03	.14	.21	.25
Satz09_dE							
	df1f2	df1f3	df1f4	df1f5	df2f3	df2f4	
		.13	.04	.18	.05	-.05	.09
Satz09_dE_z							
	df1f2	df1f3	df1f4	df1f5	df2f3	df2f4	
		.21	.01	.15	.05	-.14	-.01
Satz09_nOt							
	df1f2	df1f3	df1f4	df1f5	df2f3	df2f4	
		-.25	-.29	-.17	-.08	.08	.10
Satz09_nOt_z							
	df1f2	df1f3	df1f4	df1f5	df2f3	df2f4	
		-.25	-.23	-.26	.09	.15	.06
Satz09_dIen							
	df1f2	df1f3	df1f4	df1f5	df2f3	df2f4	
		-.10	-.01	.02	.10	.10	.11
Satz09_dIen_z							
	df1f2	df1f3	df1f4	df1f5	df2f3	df2f4	
		.10	.07	.16	-.01	-.02	.06

Tabelle B-13: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rU-fen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE							
	df2f5	df3f4	df3f5	df4f5	MFCC1	MFCC2	
		.11	-.03	.06	.06	.02	-.08
Satz04_tE_z							
	df2f5	df3f4	df3f5	df4f5	MFCC1	MFCC2	
		.23	-.08	.28	.15	-.03	.14
Satz06_nA							
	df2f5	df3f4	df3f5	df4f5	MFCC1	MFCC2	
		-.10	.11	-.03	-.04	.01	-.12
Satz06_nA_z							
	df2f5	df3f4	df3f5	df4f5	MFCC1	MFCC2	
		-.15	.16	-.07	-.14	.02	.06
Satz09_rUf							
	df2f5	df3f4	df3f5	df4f5	MFCC1	MFCC2	
		.04	.03	.07	.00	-.06	-.22
Satz09_rUf_z							
	df2f5	df3f4	df3f5	df4f5	MFCC1	MFCC2	
		.21	.09	.10	.11	-.23	-.06
Satz09_dE							
	df2f5	df3f4	df3f5	df4f5	MFCC1	MFCC2	
		.05	.12	.09	-.03	-.02	-.13
Satz09_dE_z							
	df2f5	df3f4	df3f5	df4f5	MFCC1	MFCC2	
		-.07	.11	.12	-.04	-.01	.12
Satz09_nOt							
	df2f5	df3f4	df3f5	df4f5	MFCC1	MFCC2	
		.15	.03	.16	.06	.07	-.10
Satz09_nOt_z							
	df2f5	df3f4	df3f5	df4f5	MFCC1	MFCC2	
		.28	-.11	.22	.24	-.09	-.12
Satz09_dIen							
	df2f5	df3f4	df3f5	df4f5	MFCC1	MFCC2	
		.12	.04	-.02	.03	.11	-.17
Satz09_dIen_z							
	df2f5	df3f4	df3f5	df4f5	MFCC1	MFCC2	
		-.04	.09	-.04	-.03	-.05	.01

Table B-14: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rU-fen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE	MFCC3	MFCC4	MFCC5	MFCC6	MFCC7	MFCC8
			.23	.05	-.22	-.06
Satz04_tE_z						
Satz06_nA						
Satz06_nA_z						
Satz09_rUf						
Satz09_rUf_z						
Satz09_dE						
Satz09_dE_z						
Satz09_nOt						
Satz09_nOt_z						
Satz09_dIen						
Satz09_dIen_z						

Tabelle B-15: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächst**E**“, [a:] aus „NAchmittag“, [u:] aus „rU-fen“, [e:] aus „d**En**“, [o:] aus „NOtdienst“ und [i:] aus „Notd**I**enst“) und KSS Werten

Satz04_tE						
	MFCC9	MFCC10	MFCC11	MFCC12	MFCCdelta1	MFCCdelta2
	.16	.04	.16	.03	-.11	.17
Satz04_tE_z						
	MFCC9	MFCC10	MFCC11	MFCC12	MFCCdelta1	MFCCdelta2
	.15	.15	.13	-.01	-.07	.07
Satz06_nA						
	MFCC9	MFCC10	MFCC11	MFCC12	MFCCdelta1	MFCCdelta2
	.10	.09	.18	.16	.09	-.13
Satz06_nA_z						
	MFCC9	MFCC10	MFCC11	MFCC12	MFCCdelta1	MFCCdelta2
	-.04	.19	.14	.13	-.01	.03
Satz09_rUf						
	MFCC9	MFCC10	MFCC11	MFCC12	MFCCdelta1	MFCCdelta2
	.18	.03	.22	-.05	-.09	-.15
Satz09_rUf_z						
	MFCC9	MFCC10	MFCC11	MFCC12	MFCCdelta1	MFCCdelta2
	.27	.21	.28	-.18	-.05	.00
Satz09_dE						
	MFCC9	MFCC10	MFCC11	MFCC12	MFCCdelta1	MFCCdelta2
	.06	-.01	.12	.23	.09	.09
Satz09_dE_z						
	MFCC9	MFCC10	MFCC11	MFCC12	MFCCdelta1	MFCCdelta2
	.10	-.12	.13	.20	.08	.13
Satz09_nOt						
	MFCC9	MFCC10	MFCC11	MFCC12	MFCCdelta1	MFCCdelta2
	.08	.14	.06	-.03	-.08	-.05
Satz09_nOt_z						
	MFCC9	MFCC10	MFCC11	MFCC12	MFCCdelta1	MFCCdelta2
	.18	.32	-.02	-.06	.05	.00
Satz09_dIen						
	MFCC9	MFCC10	MFCC11	MFCC12	MFCCdelta1	MFCCdelta2
	.30	.22	.15	.05	.17	.07
Satz09_dIen_z						
	MFCC9	MFCC10	MFCC11	MFCC12	MFCCdelta1	MFCCdelta2
	.35	.32	.28	.02	.35	.18

Tabelle B-16: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rUfen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE						
	MFCCdelta3	MFCCdelta4	MFCCdelta5	MFCCdelta6	MFCCdelta7	MFCCdelta8
	.09	-.07	-.06	.06	-.09	-.06
Satz04_tE_z						
	MFCCdelta3	MFCCdelta4	MFCCdelta5	MFCCdelta6	MFCCdelta7	MFCCdelta8
	.17	.03	-.15	.08	-.26	-.06
Satz06_nA						
	MFCCdelta3	MFCCdelta4	MFCCdelta5	MFCCdelta6	MFCCdelta7	MFCCdelta8
	-.11	-.15	-.06	.14	-.06	-.03
Satz06_nA_z						
	MFCCdelta3	MFCCdelta4	MFCCdelta5	MFCCdelta6	MFCCdelta7	MFCCdelta8
	-.06	-.10	-.15	.12	.00	-.04
Satz09_rUf						
	MFCCdelta3	MFCCdelta4	MFCCdelta5	MFCCdelta6	MFCCdelta7	MFCCdelta8
	-.07	-.33	-.03	.03	.01	.05
Satz09_rUf_z						
	MFCCdelta3	MFCCdelta4	MFCCdelta5	MFCCdelta6	MFCCdelta7	MFCCdelta8
	-.18	-.19	-.07	-.17	.09	-.15
Satz09_dE						
	MFCCdelta3	MFCCdelta4	MFCCdelta5	MFCCdelta6	MFCCdelta7	MFCCdelta8
	-.05	.13	-.10	.02	.24	.06
Satz09_dE_z						
	MFCCdelta3	MFCCdelta4	MFCCdelta5	MFCCdelta6	MFCCdelta7	MFCCdelta8
	-.15	.30	-.02	-.09	.19	.05
Satz09_nOt						
	MFCCdelta3	MFCCdelta4	MFCCdelta5	MFCCdelta6	MFCCdelta7	MFCCdelta8
	-.02	.10	-.07	.00	.34	.04
Satz09_nOt_z						
	MFCCdelta3	MFCCdelta4	MFCCdelta5	MFCCdelta6	MFCCdelta7	MFCCdelta8
	-.11	.12	-.10	.02	.17	-.05
Satz09_dIen						
	MFCCdelta3	MFCCdelta4	MFCCdelta5	MFCCdelta6	MFCCdelta7	MFCCdelta8
	-.19	.04	.16	-.07	-.03	.03
Satz09_dIen_z						
	MFCCdelta3	MFCCdelta4	MFCCdelta5	MFCCdelta6	MFCCdelta7	MFCCdelta8
	-.26	-.08	.03	-.02	.05	.02

Tabelle B-17: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rUfen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE							
	MFCC	MFCC	MFCC	MFCC	MFCC	MFCC	
	delta9	delta10	delta11	delta12	deltazwei1	deltazwei2	
	.11	-.02		-.17	.08	.08	-.19
Satz04_tE_z							
	MFCC	MFCC	MFCC	MFCC	MFCC	MFCC	
	delta9	delta10	delta11	delta12	deltazwei1	deltazwei2	
	.03	-.17		-.15	.19	-.12	-.14
Satz06_nA							
	MFCC	MFCC	MFCC	MFCC	MFCC	MFCC	
	delta9	delta10	delta11	delta12	deltazwei1	deltazwei2	
	.01	-.14		.03	.06	.02	-.07
Satz06_nA_z							
	MFCC	MFCC	MFCC	MFCC	MFCC	MFCC	
	delta9	delta10	delta11	delta12	deltazwei1	deltazwei2	
	.01	-.12		.00	.02	.04	.02
Satz09_rUf							
	MFCC	MFCC	MFCC	MFCC	MFCC	MFCC	
	delta9	delta10	delta11	delta12	deltazwei1	deltazwei2	
	.01	-.15		-.03	-.13	-.08	-.02
Satz09_rUf_z							
	MFCC	MFCC	MFCC	MFCC	MFCC	MFCC	
	delta9	delta10	delta11	delta12	deltazwei1	deltazwei2	
	-.14	.14		-.11	-.04	-.09	-.01
Satz09_dE							
	MFCC	MFCC	MFCC	MFCC	MFCC	MFCC	
	delta9	delta10	delta11	delta12	deltazwei1	deltazwei2	
	.14	-.21		.03	.16	-.08	.08
Satz09_dE_z							
	MFCC	MFCC	MFCC	MFCC	MFCC	MFCC	
	delta9	delta10	delta11	delta12	deltazwei1	deltazwei2	
	.09	-.22		.08	.04	-.04	.05
Satz09_nOt							
	MFCC	MFCC	MFCC	MFCC	MFCC	MFCC	
	delta9	delta10	delta11	delta12	deltazwei1	deltazwei2	
	-.08	.00		-.07	.08	.11	.07

Satz09_nOt_z

MFCC	MFCC	MFCC	MFCC	MFCC	MFCC	MFCC
delta9	delta10	delta11	delta12	deltazwei1	deltazwei2	
	-.10	-.06	-.21	.06	.07	.15

Satz09_dIen

MFCC	MFCC	MFCC	MFCC	MFCC	MFCC	MFCC
delta9	delta10	delta11	delta12	deltazwei1	deltazwei2	
	.09	.05	-.07	.08	-.01	.15

Satz09_dIen_z

MFCC	MFCC	MFCC	MFCC	MFCC	MFCC	MFCC
delta9	delta10	delta11	delta12	deltazwei1	deltazwei2	
	-.01	.03	.12	.19	-.05	.03

Tabelle B-18: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rUfen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE

MFCC	MFCC	MFCC	MFCC	MFCC	MFCC
deltazwei3	deltazwei4	deltazwei5	deltazwei6	deltazwei7	deltazwei8
-0.27	-0.14	0.08	0.05	-0.01	-0.10

Satz04_tE_z

MFCC	MFCC	MFCC	MFCC	MFCC	MFCC
deltazwei3	deltazwei4	deltazwei5	deltazwei6	deltazwei7	deltazwei8
-0.26	-0.05	0.04	-0.03	-0.14	-0.06

Satz06_nA

MFCC	MFCC	MFCC	MFCC	MFCC	MFCC
deltazwei3	deltazwei4	deltazwei5	deltazwei6	deltazwei7	deltazwei8
-0.21	0.05	-0.07	0.19	-0.24	0.07

Satz06_nA_z

MFCC	MFCC	MFCC	MFCC	MFCC	MFCC
deltazwei3	deltazwei4	deltazwei5	deltazwei6	deltazwei7	deltazwei8
-0.13	-0.02	-0.15	0.31	-0.15	0.13

Satz09_rUf

MFCC	MFCC	MFCC	MFCC	MFCC	MFCC
deltazwei3	deltazwei4	deltazwei5	deltazwei6	deltazwei7	deltazwei8
0.21	0.02	-0.05	-0.19	-0.21	-0.02

Satz09_rUf_z

MFCC	MFCC	MFCC	MFCC	MFCC	MFCC
deltazwei3	deltazwei4	deltazwei5	deltazwei6	deltazwei7	deltazwei8
0.28	0.00	0.06	-0.08	-0.26	-0.03

Satz09_dE

MFCC	MFCC	MFCC	MFCC	MFCC	MFCC
deltazwei3	deltazwei4	deltazwei5	deltazwei6	deltazwei7	deltazwei8
-0.15	0.14	-0.14	0.00	0.07	0.15

Satz09_dE_z

MFCC	MFCC	MFCC	MFCC	MFCC	MFCC
deltazwei3	deltazwei4	deltazwei5	deltazwei6	deltazwei7	deltazwei8
-0.13	0.14	0.03	-0.02	0.02	0.14

Satz09_nOt

MFCC	MFCC	MFCC	MFCC	MFCC	MFCC
deltazwei3	deltazwei4	deltazwei5	deltazwei6	deltazwei7	deltazwei8
-0.02	-0.15	0.21	-0.02	0.01	-0.26

Satz09_nOt_z

MFCC	MFCC	MFCC	MFCC	MFCC	MFCC
deltazwei3	deltazwei4	deltazwei5	deltazwei6	deltazwei7	deltazwei8
.16	-.05	.19	-.07	-.05	-.14

Satz09_dIen

MFCC	MFCC	MFCC	MFCC	MFCC	MFCC
deltazwei3	deltazwei4	deltazwei5	deltazwei6	deltazwei7	deltazwei8
.05	-.11	-.16	-.02	-.04	-.02

Satz09_dIen_z

MFCC	MFCC	MFCC	MFCC	MFCC	MFCC
deltazwei3	deltazwei4	deltazwei5	deltazwei6	deltazwei7	deltazwei8
.09	-.14	-.07	-.13	.02	.11

Tabelle B-19: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rUfen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE

MFCC	MFCC	MFCC	MFCC		
deltazwei9	deltazwei10	deltazwei11	deltazwei12	int_min	int_timemin
	-.25	.12	-.04	-.03	.03
				.03	.12

Satz04_tE_z

MFCC	MFCC	MFCC	MFCC		
deltazwei9	deltazwei10	deltazwei11	deltazwei12	int_min	int_timemin
	-.15	.04	-.05	.04	-.12
				-.12	.07

Satz06_nA

MFCC	MFCC	MFCC	MFCC		
deltazwei9	deltazwei10	deltazwei11	deltazwei12	int_min	int_timemin
	-.02	-.18	.04	.12	.00
				.12	.05

Satz06_nA_z

MFCC	MFCC	MFCC	MFCC		
deltazwei9	deltazwei10	deltazwei11	deltazwei12	int_min	int_timemin
	.07	-.26	.29	.25	-.25
				-.25	-.06

Satz09_rUf

MFCC	MFCC	MFCC	MFCC		
deltazwei9	deltazwei10	deltazwei11	deltazwei12	int_min	int_timemin
	.02	.11	-.19	-.05	-.05
				-.05	-.10

Satz09_rUf_z

MFCC	MFCC	MFCC	MFCC		
deltazwei9	deltazwei10	deltazwei11	deltazwei12	int_min	int_timemin
	-.05	.03	-.07	.06	-.30
				-.30	.01

Satz09_dE

MFCC	MFCC	MFCC	MFCC		
deltazwei9	deltazwei10	deltazwei11	deltazwei12	int_min	int_timemin
	-.19	-.17	-.12	-.17	-.03
				-.17	-.18

Satz09_dE_z

MFCC	MFCC	MFCC	MFCC		
deltazwei9	deltazwei10	deltazwei11	deltazwei12	int_min	int_timemin
	-.18	-.12	-.09	-.04	-.30
				-.30	.04

Satz09_nOt

MFCC	MFCC	MFCC	MFCC		
deltazwei9	deltazwei10	deltazwei11	deltazwei12	int_min	int_timemin
	-.27	.20	.05	.01	.12
				.12	-.08

Tabelle B-20: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rU-fen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE						
	int_max	int_timemax	int_mean	int_sd	int_25quant	int_75quant
	.01	.14	.02	-.16	.03	.02
Satz04_tE_z						
	int_max	int_timemax	int_mean	int_sd	int_25quant	int_75quant
	-.16	.15	-.16	-.11	-.14	-.16
Satz06_nA						
	int_max	int_timemax	int_mean	int_sd	int_25quant	int_75quant
	.00	-.03	.00	-.02	.00	.00
Satz06_nA_z						
	int_max	int_timemax	int_mean	int_sd	int_25quant	int_75quant
	-.24	-.14	-.25	.24	-.25	-.24
Satz09_rUf						
	int_max	int_timemax	int_mean	int_sd	int_25quant	int_75quant
	-.08	.13	-.08	-.10	-.05	-.07
Satz09_rUf_z						
	int_max	int_timemax	int_mean	int_sd	int_25quant	int_75quant
	-.32	.04	-.34	-.05	-.30	-.32
Satz09_dE						
	int_max	int_timemax	int_mean	int_sd	int_25quant	int_75quant
	-.05	.00	-.04	-.13	-.03	-.04
Satz09_dE_z						
	int_max	int_timemax	int_mean	int_sd	int_25quant	int_75quant
	-.30	-.03	-.30	-.10	-.30	-.30
Satz09_nOt						
	int_max	int_timemax	int_mean	int_sd	int_25quant	int_75quant
	.03	-.08	.05	-.29	.12	.03
Satz09_nOt_z						
	int_max	int_timemax	int_mean	int_sd	int_25quant	int_75quant
	-.24	-.04	-.19	-.21	-.06	-.23
Satz09_dIen						
	int_max	int_timemax	int_mean	int_sd	int_25quant	int_75quant
	.04	-.15	.04	-.21	.05	.04
Satz09_dIen_z						
	int_max	int_timemax	int_mean	int_sd	int_25quant	int_75quant
	-.29	.10	-.31	.05	-.32	-.30

Tabelle B-21: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rU-fen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE

int_mintime_rel	int_maxtime_rel	int_range	openQuot_ OQG	glottOpen_ GOG	skewness_ SKG
.06	-.07	-.05	.01	.01	.07

Satz04_tE_z

int_mintime_rel	int_maxtime_rel	int_range	openQuot_ OQG	glottOpen_ GOG	skewness_ SKG
.05	-.02	-.03	.00	-.01	.05

Satz06_nA

int_mintime_rel	int_maxtime_rel	int_range	openQuot_ OQG	glottOpen_ GOG	skewness_ SKG
.01	-.06	-.01	-.11	-.09	-.11

Satz06_nA_z

int_mintime_rel	int_maxtime_rel	int_range	openQuot_ OQG	glottOpen_ GOG	skewness_ SKG
-.02	-.12	.05	-.08	-.04	-.05

Satz09_rUf

int_mintime_rel	int_maxtime_rel	int_range	openQuot_ OQG	glottOpen_ GOG	skewness_ SKG
-.21	.11	-.05	-.17	-.24	-.26

Satz09_rUf_z

int_mintime_rel	int_maxtime_rel	int_range	openQuot_ OQG	glottOpen_ GOG	skewness_ SKG
-.09	-.03	.02	-.16	-.17	-.15

Satz09_dE

int_mintime_rel	int_maxtime_rel	int_range	openQuot_ OQG	glottOpen_ GOG	skewness_ SKG
-.16	.10	-.09	.07	-.05	-.03

Satz09_dE_z

int_mintime_rel	int_maxtime_rel	int_range	openQuot_ OQG	glottOpen_ GOG	skewness_ SKG
-.04	-.08	.06	.08	-.07	-.07

Satz09_nOt

int_mintime_rel	int_maxtime_rel	int_range	openQuot_ OQG	glottOpen_ OG	skewness_ SKG
.02	-.06	-.24	-.11	-.04	-.06

Satz09_nOt_z

int_mintime_rel	int_maxtime_rel	int_range	openQuot_ OQG	glottOpen_ GOG	skewness_ SKG
.07	-.01	-.17	-.12	-.11	-.07

Satz09_dIen

int_mintime_rel	int_maxtime_rel	int_range	openQuot_ OQG	glottOpen_ GOG	skewness_ SKG
-.02	-.16	-.14	-.01	-.14	-.07

Satz09_dIen_z

int_mintime_rel	int_maxtime_rel	int_range	openQuot_ OQG	glottOpen_ GOG	skewness_ SKG
-.08	-.01	.12	.00	.05	.02

Tabelle B-22: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rUfen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE

RateClos_	imcomplclos_					
RCG	IC	energmax0_2k	energmax2k_5k	hammarberg	hammarberg_s0	
.08	-.02	.03	.05	-.05	-.03	

Satz04_tE_z

rateClos_	imcomplclos_					
RCG	IC	energmax0_2k	energmax2k_5k	hammarberg	hammarberg_s0	
.03	-.03	-.09	-.28	.28	-.16	

Satz06_nA

rateClos_	imcomplclos_					
RCG	IC	energmax0_2k	energmax2k_5k	hammarberg	hammarberg_s0	
-.10	-.05	.01	.06	-.10	.07	

Satz06_nA_z

rateClos_	imcomplclos_					
RCG	IC	energmax0_2k	energmax2k_5k	hammarberg	hammarberg_s0	
-.07	-.12	-.16	-.24	.23	-.16	

Satz09_rUf

rateClos_	imcomplclos_					
RCG	IC	energmax0_2k	energmax2k_5k	hammarberg	hammarberg_s0	
-.28	-.13	-.10	.02	-.25	-.08	

Satz09_rUf_z

rateClos_	imcomplclos_					
RCG	IC	energmax0_2k	energmax2k_5k	hammarberg	hammarberg_s0	
-.24	-.05	-.32	-.20	-.13	-.41	

Satz09_dE

rateClos_	imcomplclos_					
RCG	IC	energmax0_2k	energmax2k_5k	hammarberg	hammarberg_s0	
-.07	-.05	-.08	-.01	-.08	-.03	

Satz09_dE_z

rateClos_	imcomplclos_					
RCG	IC	energmax0_2k	energmax2k_5k	hammarberg	hammarberg_s0	
-.07	-.08	-.34	-.33	.19	-.25	

Satz09_nOt

rateClos_	imcomplclos_					
RCG	IC	energmax0_2k	energmax2k_5k	hammarberg	hammarberg_s0	
-.12	.12	.05	.07	-.04	.03	

Satz09_nOt_z

rateClos_	imcomplclos_					
RCG	IC	energmax0_2k	energmax2k_5k	hammarberg	hammarberg_s0	
-.15	.08	-.15	-.02	-.12		-.22

Satz09_dIen

rateClos_	imcomplclos_					
RCG	IC	energmax0_2k	energmax2k_5k	hammarberg	hammarberg_s0	
-.04	-.16	.03	.10	-.16		.11

Satz09_dIen_z

rateClos_	imcomplclos_					
RCG	IC	energmax0_2k	energmax2k_5k	hammarberg	hammarberg_s0	
.04	.10	-.20	-.22	-.09		-.24

Tabelle B-23: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rUfen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE

hammarberg_s1	hammarberg_s2	hammarberg_s3	hammarberg_s4	spec_cm	spec_kurt
-0.12	-0.24	0.10	-0.06	0.05	-0.05

Satz04_tE_z

hammarberg_s1	hammarberg_s2	hammarberg_s3	hammarberg_s4	spec_cm	spec_kurt
0.02	-0.07	0.17	0.09	0.03	0.28

Satz06_nA

hammarberg_s1	hammarberg_s2	hammarberg_s3	hammarberg_s4	spec_cm	spec_kurt
0.13	0.00	-0.08	0.01	-0.05	-0.13

Satz06_nA_z

hammarberg_s1	hammarberg_s2	hammarberg_s3	hammarberg_s4	spec_cm	spec_kurt
-0.02	-0.04	-0.08	0.17	-0.02	0.03

Satz09_rUf

hammarberg_s1	hammarberg_s2	hammarberg_s3	hammarberg_s4	spec_cm	spec_kurt
0.13	-0.11	0.00	0.01	-0.01	-0.21

Satz09_rUf_z

hammarberg_s1	hammarberg_s2	hammarberg_s3	hammarberg_s4	spec_cm	spec_kurt
-0.03	-0.33	-0.21	-0.22	0.06	-0.01

Satz09_dE

hammarberg_s1	hammarberg_s2	hammarberg_s3	hammarberg_s4	spec_cm	spec_kurt
0.20	-0.12	-0.02	-0.06	0.03	-0.16

Satz09_dE_z

hammarberg_s1	hammarberg_s2	hammarberg_s3	hammarberg_s4	spec_cm	spec_kurt
0.18	-0.09	0.02	0.06	-0.23	0.24

Satz09_nOt

hammarberg_s1	hammarberg_s2	hammarberg_s3	hammarberg_s4	spec_cm	spec_kurt
-0.02	-0.07	-0.09	-0.02	0.01	0.05

Satz09_nOt_z

hammarberg_s1	hammarberg_s2	hammarberg_s3	hammarberg_s4	spec_cm	spec_kurt
-0.02	-0.10	-0.20	-0.12	-0.02	0.08

Satz09_dIen

hammarberg_s1	hammarberg_s2	hammarberg_s3	hammarberg_s4	spec_cm	spec_kurt
0.05	0.08	0.05	0.04	-0.01	-0.16

Satz09_dIen_z

hammarberg_s1	hammarberg_s2	hammarberg_s3	hammarberg_s4	spec_cm	spec_kurt
0.00	-0.16	-0.11	-0.11	-0.11	0.02

Tabelle B-24: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rUfen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE						
spec_skew	spec_sd	spec_cog	baen0_1k	baen1k_8k	baen_hf_lf1k	
	-0.05	.05	-0.03	.12	.15	-0.07
Satz04_tE_z						
spec_skew	spec_sd	spec_cog	baen0_1k	baen1k_8k	baen_hf_lf1k	
	.22	-0.07	-.22	-0.02	-0.03	-0.17
Satz06_nA						
spec_skew	spec_sd	spec_cog	baen0_1k	baen1k_8k	baen_hf_lf1k	
	-0.06	.06	-0.05	.05	.06	-0.11
Satz06_nA_z						
spec_skew	spec_sd	spec_cog	baen0_1k	baen1k_8k	baen_hf_lf1k	
	.03	-0.03	-.33	-0.05	-0.15	-.25
Satz09_rUf						
spec_skew	spec_sd	spec_cog	baen0_1k	baen1k_8k	baen_hf_lf1k	
	-.27	.08	.04	-0.03	-0.04	.09
Satz09_rUf_z						
spec_skew	spec_sd	spec_cog	baen0_1k	baen1k_8k	baen_hf_lf1k	
	-0.02	.10	-.20	-0.14	-0.09	.10
Satz09_dE						
spec_skew	spec_sd	spec_cog	baen0_1k	baen1k_8k	baen_hf_lf1k	
	-0.15	.04	.08	-0.04	-0.04	.01
Satz09_dE_z						
spec_skew	spec_sd	spec_cog	baen0_1k	baen1k_8k	baen_hf_lf1k	
	.21	-.24	-0.13	-0.10	-0.05	-0.09
Satz09_nOt						
spec_skew	spec_sd	spec_cog	baen0_1k	baen1k_8k	baen_hf_lf1k	
	.07	.03	-0.11	-0.03	-0.07	-.22
Satz09_nOt_z						
spec_skew	spec_sd	spec_cog	baen0_1k	baen1k_8k	baen_hf_lf1k	
	.10	.01	-0.13	-0.10	-0.09	-0.03
Satz09_dIen						
spec_skew	spec_sd	spec_cog	baen0_1k	baen1k_8k	baen_hf_lf1k	
	-.21	.00	-0.01	-0.06	-0.06	.04
Satz09_dIen_z						
spec_skew	spec_sd	spec_cog	baen0_1k	baen1k_8k	baen_hf_lf1k	
	-0.05	-0.18	-.25	-0.14	-0.10	-0.05

Tabelle B-25: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rUfen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE

baen_	baen_	baen_	baen_	baen_	bade
125_200	200_300	500_600	1k_1600	5k_8k	200_1k
.03	.14	.10	.03	.05	.11

Satz04_tE_z

baen_	baen_	baen_	baen_	baen_	bade200_1k
125_200	200_300	500_600	1k_1600	5k_8k	bade200_1k
.08	-.11	.04	-.09	-.01	-.02

Satz06_nA

baen_	baen_	baen_	baen_	baen_	bade200_1k
125_200	200_300	500_600	1k_1600	5k_8k	bade200_1k
-.06	.08	.11	.03	.20	.06

Satz06_nA_z

baen_	baen_	baen_	baen_	baen_	bade200_1k
125_200	200_300	500_600	1k_1600	5k_8k	bade200_1k
-.04	.03	-.01	-.17	.08	-.04

Satz09_rUf

baen_	baen_	baen_	baen_	baen_	bade200_1k
125_200	200_300	500_600	1k_1600	5k_8k	bade200_1k
-.17	.05	.02	-.04	-.05	-.02

Satz09_rUf_z

baen_	baen_	baen_	baen_	baen_	bade200_1k
125_200	200_300	500_600	1k_1600	5k_8k	bade200_1k
-.13	-.01	-.11	-.09	-.04	-.14

Satz09_dE

baen_	baen_	baen_	baen_	baen_	bade200_1k
125_200	200_300	500_600	1k_1600	5k_8k	bade200_1k
-.13	-.16	.06	-.05	-.02	-.03

Satz09_dE_z

baen_	baen_	baen_	baen_	baen_	bade200_1k
125_200	200_300	500_600	1k_1600	5k_8k	bade200_1k
-.13	-.26	.02	-.05	-.08	-.09

Satz09_nOt

baen_	baen_	baen_	baen_	baen_	bade200_1k
125_200	200_300	500_600	1k_1600	5k_8k	bade200_1k
-.05	-.05	-.04	-.07	-.06	-.03

Satz09_nOt_z

baen_ 125_200	baen_ 200_300	baen_ 500_600	baen_ k_1600	baen_ 5k_8k	bade200_1k	
-0.13	-0.02		-0.03	-0.09	-0.10	-0.09

Satz09_dIen

baen_ 125_200	baen_ 200_300	baen_ 500_600	baen_ 1k_1600	baen_ 5k_8k	bade200_1k	
.02	-0.10		.01	-0.07	.03	-0.08

Satz09_dIen_z

baen_ 125_200	baen_ 200_300	baen_ 500_600	baen_ 1k_1600	baen_ 5k_8k	bade200_1k	
-0.11	-0.11	-0.25	-0.10	-0.09	-0.09	-0.15

Tabelle B-26: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rUfen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE

baendi	badeddiff	ltas_slope				
0_500_500_4k	0_500_500_1k	0_1k_1k_4k	ltas_min	ltas_minfreq	ltas_max	
.03	.03	-.01	.02	-.01	.03	

Satz04_tE_z

baendi	badeddiff	ltas_slope				
0_500_500_4k	0_500_500_1k	0_1k_1k_4k	ltas_min	ltas_minfreq	ltas_max	
.00	.00	-.25	-.04	.13	-.09	

Satz06_nA

baendi	badeddiff	ltas_slope				
0_500_500_4k	0_500_500_1k	0_1k_1k_4k	ltas_min	ltas_minfreq	ltas_max	
.02	.02	-.12	-.01	.11	.01	

Satz06_nA_z

baendi	badeddiff	ltas_slope				
0_500_500_4k	0_500_500_1k	0_1k_1k_4k	ltas_min	ltas_minfreq	ltas_max	
-.15	-.15	-.28	-.04	-.03	-.16	

Satz09_rUf

baendi	badeddiff	ltas_slope				
0_500_500_4k	0_500_500_1k	0_1k_1k_4k	ltas_min	ltas_minfreq	ltas_max	
.08	.08	.06	-.08	-.08	-.10	

Satz09_rUf_z

baendi	badeddiff	ltas_slope				
0_500_500_4k	0_500_500_1k	0_1k_1k_4k	ltas_min	ltas_minfreq	ltas_max	
-.06	-.06	.02	-.31	-.02	-.32	

Satz09_dE

baendi	badeddiff	ltas_slope				
0_500_500_4k	0_500_500_1k	0_1k_1k_4k	ltas_min	ltas_minfreq	ltas_max	
.07	.07	.10	-.08	-.07	-.08	

Satz09_dE_z

baendi	badeddiff	ltas_slope				
0_500_500_4k	0_500_500_1k	0_1k_1k_4k	ltas_min	ltas_minfreq	ltas_max	
.00	.00	-.10	-.23	-.10	-.34	

Satz09_nOt

baendi	badeddiff	ltas_slope				
0_500_500_4k	0_500_500_1k	0_1k_1k_4k	ltas_min	ltas_minfreq	ltas_max	
-.06	-.06	-.17	.09	.05	.05	

Satz09_nOt_z

baendi	badeddiff	ltas_slope				
0_500_500_4k	0_500_500_1k	0_1k_1k_4k	ltas_min	ltas_minfreq	ltas_max	
-.06	-.06	-.07	-.04	.09	-.15	

Satz09_dIen

baendi	badeddiff	ltas_slope				
0_500_500_4k	0_500_500_1k	0_1k_1k_4k	ltas_min	ltas_minfreq	ltas_max	
.13	.13	.13	-.04	.06	.03	

Satz09_dIen_z

baendi	badeddiff	ltas_slope				
0_500_500_4k	0_500_500_1k	0_1k_1k_4k	ltas_min	ltas_minfreq	ltas_max	
.03	.06	.11	-.19	-.09	-.20	

Tabelle B-27: Korrelationen von akustischen Roh- und sprecherspezifisch z-transformierten Kennzahlen extrahierter Vokale ([ə] aus „nächstE“, [a:] aus „NAchmittag“, [u:] aus „rU-fen“, [e:] aus „dEn“, [o:] aus „NOtdienst“ und [i:] aus „NotdIenst“) und KSS Werten

Satz04_tE						
	ltas_maxfreq	ltas_sd	ltas_locpeak	ltas_trend_max	pe1k_baenqout	
	-0.14	0.03	0.01	0.08		-0.08
Satz04_tE_z						
	ltas_maxfreq	ltas_sd	ltas_locpeak	ltas_trend_max	pe1k_baenqout	
	-0.15	-0.12	-0.14	-0.08		0.18
Satz06_nA						
	ltas_maxfreq	ltas_sd	ltas_locpeak	ltas_trend_max	pe1k_baenqout	
	-0.09	0.01	0.05	-0.03		0.05
Satz06_nA_z						
	ltas_maxfreq	ltas_sd	ltas_locpeak	ltas_trend_max	pe1k_baenqout	
	-0.33	-0.19	0.00	-0.24		0.07
Satz09_rUf						
	ltas_maxfreq	ltas_sd	ltas_locpeak	ltas_trend_max	pe1k_baenqout	
	0.00	-0.10	0.32	-0.09		-0.07
Satz09_rUf_z						
	ltas_maxfreq	ltas_sd	ltas_locpeak	ltas_trend_max	pe1k_baenqout	
	-0.18	-0.34	0.08	-0.28		0.04
Satz09_dE						
	ltas_maxfreq	ltas_sd	ltas_locpeak	ltas_trend_max	pe1k_baenqout	
	0.20	-0.08	-0.02	0.07		-0.17
Satz09_dE_z						
	ltas_maxfreq	ltas_sd	ltas_locpeak	ltas_trend_max	pe1k_baenqout	
	0.21	-0.34	-0.09	-0.10		0.04
Satz09_nOt						
	ltas_maxfreq	ltas_sd	ltas_locpeak	ltas_trend_max	pe1k_baenqout	
	-0.13	0.06	-0.01	-0.02		0.10
Satz09_nOt_z						
	ltas_maxfreq	ltas_sd	ltas_locpeak	ltas_trend_max	pe1k_baenqout	
	-0.08	-0.16	0.03	-0.21		-0.04
Satz09_dIen						
	ltas_maxfreq	ltas_sd	ltas_locpeak	ltas_trend_max	pe1k_baenqout	
	-0.27	0.03	0.03	0.07		-0.21
Satz09_dIen_z						
	ltas_maxfreq	ltas_sd	ltas_locpeak	ltas_trend_max	pe1k_baenqout	
	-0.46	-0.21	-0.07	-0.19		-0.11

ANHANG C:
MITTELWERTVERGLEICHS-BEZOGENE ERGEBNISSE
(EXEMPLARISCH “FAHRERASSISTENZSATZ 9“)

Tabelle C-1: Sprachmaterial "Fahrerassistenzsatz 9": Mittelwerte (MW), Standardabweichungen (SD) für die Bedingungen "wach" ($KSS_{kat}=0$; $KSS \leq 5$) und "schläfrig" ($KSS_{kat}=1$; $KSS > 5$) (Rohwerte, Frauen)

Kennwert	KSS_kat	MW	SD	T	df	p
f0_timemin	0	1.32	0.30	1.77	46	.08
	1	1.14	0.36			
f0_75quant	0	252.60	28.74	2.07	46	.04
	1	233.46	30.10			
f0_mintime_rel	0	0.80	0.11	3.47	46	.00
	1	0.64	0.20			
f0_mean	0	219.81	20.30	2.28	46	.03
	1	203.84	26.09			
shimmer	0	0.11	0.02	-2.34	46	.02
	1	0.14	0.04			
f1_maxtime_rel	0	0.65	0.23	1.99	46	.05
	1	0.50	0.26			
f2_sd	0	583.15	68.66	2.45	46	.02
	1	526.01	84.08			
f3_sd	0	534.93	110.07	1.69	46	.10
	1	480.03	79.38			
f4_max	0	5813.34	117.74	-2.27	46	.03
	1	5889.02	61.15			
f4_range	0	2413.84	299.79	-1.84	46	.07
	1	2585.79	293.00			
f5_bandw_q50	0	699.28	133.03	-2.29	46	.03
	1	781.97	162.01			
f5_bandw_q75	0	1012.27	168.89	-1.84	46	.07
	1	1134.50	166.27			
df1f4	0	-3910.98	108.69	-1.66	46	.10
	1	-3849.24	135.38			
df2f3	0	-1316.20	127.22	-1.66	46	.10

ANHANG C: MITTELWERTVERGLEICHS-BEZOGENE ERGEBNISSE

	1	-1255.76	74.92			
df2f4	0	-2641.14	141.88	-2.07	46	.04
	1	-2553.01	111.55			
MFCC2	0	83.08	28.03	2.36	46	.02
	1	63.44	21.13			
Δ MFCC3	0	-0.03	0.10	-1.66	46	.10
	1	0.03	0.09			
MFCC4	0	-20.71	16.62	-1.71	46	.09
	1	-11.94	14.86			
$\Delta\Delta$ MFCC8	0	-0.01	0.07	-1.69	46	.10
	1	0.03	0.06			
MFCC11	0	-1.20	8.78	-1.66	46	.10
	1	3.29	7.78			
int_timemax	0	0.37	0.23	-2.33	46	.02
	1	0.58	0.37			
int_maxtime_rel	0	0.23	0.16	-1.66	46	.10
	1	0.33	0.23			
baen_hf_lf1000	0	0.02	0.02	-1.91	46	.06
	1	0.04	0.04			
baendi0_500_500_4000	0	-7.97	3.30	-2.07	46	.04
	1	-5.87	2.89			
badediff0_500_500_4000	0	-16.42	3.30	-2.07	46	.04
	1	-14.33	2.89			

Tabelle C-2: Sprachmaterial “Fahrerassistenzsatz 9“: Mittelwerte (MW), Standardabweichungen (SD) für die Bedingungen “wach“ (KSS_kat=0; $KSS \leq 5$) und “schläfrig“ (KSS_kat=1; $KSS > 5$) (Rohwerte, Männer)

Kennwert	KSS	MW	SD	T	df	p
f0_75quant	0	158.71	32.46	1.82	22	.08
	1	135.54	12.25	1.98	22	.06
f0_mean	0	143.20	29.66	2.39	22	.03
	1	119.95	12.69			
df1f2	0	-1302.80	76.60	1.76	22	.09
	1	-1415.23	156.79			
df1f3	0	-2621.38	114.74	1.82	22	.08
	1	-2727.92	176.95			
MFCC5	0	39.83	13.37	-2.18	22	.04
	1	27.85	17.66			
Δ MFCC5	0	-0.06	0.11	1.75	22	.09
	1	0.06	0.14			
MFCC8	0	-9.94	10.30	2.30	22	.03
	1	-20.40	19.07			
Δ MFCC10	0	0.03	0.06	2.30	22	.03
	1	-0.04	0.09			
$\Delta\Delta$ MFCC10	0	-0.02	0.07	-1.85	22	.08
	1	0.04	0.08			
Δ MFCC11	0	0.03	0.08	3.49	22	.00
	1	-0.10	0.09			
Δ MFCC12	0	0.00	0.06	2.89	22	.01
	1	-0.08	0.07			
int_min	0	52.72	2.71	2.15	22	.04
	1	50.29	1.93			
hammarberg_s4	0	17.97	5.10	-1.87	22	.08
	1	22.16	4.68			
spec_cm	0	8062267099. 22	5229765437. 65	1.99	22	.06

ANHANG C: MITTELWERTVERGLEICHS-BEZOGENE ERGEBNISSE

	1	3570717477. 94	4393029607. 43			
spec_kurt	0	78.22	52.59	-3.26	22	.00
	1	187.65	114.39			
spec_skew	0	7.88	2.50	-2.66	22	.01
	1	11.32	3.73			
spec_sd	0	994.03	326.33	2.29	22	.03
	1	654.83	337.47			

Tabelle C-3: Sprachmaterial “Fahrerassistenzsatz 9“: Mittelwerte (MW), Standardabweichungen (SD) für die Bedingungen “wach“ ($KSS_{kat}=0$; $KSS \leq 5$) und “schläfrig“ ($KSS_{kat}=1$; $KSS > 5$) (Rohwerte, alle Probanden)

Kennwert	KSS	MW	SD	T	df	p
f0_timemin	0	1.28	0.34	1.73	70	.09
	1	1.12	0.36			
f0_mintime_rel	0	0.78	0.15	2.43	70	.02
	1	0.68	0.18			
Shimmer	0	0.12	0.02	-2.22	70	.03
	1	0.14	0.04			
f2_mean	0	1930.93	123.14	-1.69	70	.10
	1	1983.50	112.64			
f2_sd	0	585.52	78.36	1.74	70	.09
	1	547.79	95.57			
f3_range	0	3054.53	471.43	1.77	70	.08
	1	2835.65	490.98			
f3_sd	0	573.71	110.32	1.81	70	.08
	1	523.62	97.71			
df1f2	0	-1280.83	93.57	2.02	70	.05
	1	-1335.90	129.11			
MFCC2	0	77.32	28.94	1.67	70	.10
	1	65.55	22.18			
Δ MFCC3	0	-0.01	0.11	-2.19	70	.03
	1	0.05	0.11			
MFCC4	0	-18.31	17.72	-2.34	70	.02
	1	-7.44	18.39			
Δ MFCC4	0	-0.04	0.11	-1.85	70	.07
	1	0.01	0.12			
Δ MFCC5	0	-0.04	0.09	-1.82	70	.07
	1	0.00	0.10			
$\Delta\Delta$ MFCC8	0	-0.01	0.08	-2.19	70	.03

	1	0.03	0.06			
MFCC9	0	2.99	10.14	-1.83	70	.07
	1	7.91	11.01			
Δ MFCC10	0	0.03	0.07	2.25	70	.03
	1	-0.01	0.07			
$\Delta\Delta$ MFCC10	0	-0.01	0.07	-1.85	70	.07
	1	0.03	0.07			
Δ MFCC11	0	0.02	0.07	2.32	70	.02
	1	-0.03	0.08			
int_timemax	0	0.40	0.27	-2.12	70	.04
	1	0.57	0.36			
int_maxtime_rel	0	0.25	0.17	-2.05	70	.05
	1	0.34	0.21			

Tabelle C-4: Sprachmaterial “Fahrerassistenzsatz 9“: Mittelwerte (MW), Standardabweichungen (SD) für die Bedingungen “wach“ ($KSS_{kat}=0$; $KSS \leq 5$) und “schläfrig“ ($KSS_{kat}=1$; $KSS > 5$) (z-Werte, Frauen)

Kennwert	KSS	MW	SD	T	df	p
f0_75quant	0	0.17	1.01	1.88	46	.07
	1	-0.42	0.91			
f0_mintime_rel	0	0.23	1.01	2.58	46	.01
	1	-0.55	0.81			
Jitter	0	0.19	1.03	2.10	46	.04
	1	-0.46	0.84			
f1_maxtime_rel	0	0.15	0.95	1.68	46	.10
	1	-0.37	1.09			
f1_mean	0	-0.16	0.94	-1.76	46	.09
	1	0.39	1.10			
f2_mean	0	-0.18	1.02	-1.99	46	.05
	1	0.44	0.87			
f2_sd	0	0.17	1.03	1.92	46	.06
	1	-0.42	0.86			
f2_bandw_q25	0	-0.20	1.00	-2.27	46	.03
	1	0.50	0.88			
f3_min	0	-0.16	1.04	-1.74	46	.09
	1	0.39	0.84			
f3_mintime_rel	0	0.18	1.06	1.93	46	.06
	1	-0.43	0.74			
f3_max	0	0.21	0.88	2.34	46	.02
	1	-0.51	1.16			
f3_maxtime_rel	0	0.25	0.92	2.89	46	.01
	1	-0.61	0.98			
f3_range	0	0.20	1.00	2.19	46	.03
	1	-0.48	0.91			
f3_bandw_q75	0	0.17	1.03	1.83	46	.07

ANHANG C: MITTELWERTVERGLEICHS-BEZOGENE ERGEBNISSE

	1	-0.41	0.86			
f4_mean	0	0.19	0.98	2.15	46	.04
	1	-0.47	0.96			
f5_mean	0	0.19	0.95	2.09	46	.04
	1	-0.46	1.03			
f5_bandw_q50	0	-0.23	0.98	-2.55	46	.01
	1	0.55	0.89			
f5_bandw_q75	0	-0.29	1.00	-3.37	46	.00
	1	0.69	0.66			
df1f4	0	-0.27	0.96	-3.13	46	.00
	1	0.65	0.83			
df1f5	0	-0.21	0.95	-2.37	46	.02
	1	0.51	1.00			
df2f4	0	-0.28	1.00	-3.25	46	.00
	1	0.67	0.69			
df2f5	0	-0.22	0.98	-2.52	46	.02
	1	0.54	0.89			
df3f4	0	-0.16	1.00	-1.73	46	.09
	1	0.39	0.96			
MFCC2	0	0.29	0.92	3.49	46	.00
	1	-0.71	0.86			
MFCC4	0	-0.28	1.01	-3.22	46	.00
	1	0.67	0.64			
MFCC6	0	0.17	0.97	1.81	46	.08
	1	-0.40	1.02			
Δ MFCC6	0	-0.20	0.95	-2.20	46	.03
	1	0.48	1.02			
MFCC11	0	-0.21	0.95	-2.38	46	.02
	1	0.52	0.99			
int_timemax	0	-0.16	0.98	-1.73	46	.09
	1	0.39	1.01			
int_mean	0	0.17	0.97	1.88	46	.07

ANHANG C: MITTELWERTVERGLEICHS-BEZOGENE ERGEBNISSE

	1	-0.42	1.03			
int_25quant	0	0.21	1.03	2.33	46	.02
	1	-0.51	0.80			
int_75quant	0	0.20	0.97	2.18	46	.03
	1	-0.48	0.99			
energmax0_2000	0	0.22	0.92	2.54	46	.02
	1	-0.55	1.05			
baen0_1000	0	0.16	0.99	1.76	46	.09
	1	-0.39	0.99			
baen200_300	0	0.22	0.98	2.51	46	.02
	1	-0.54	0.90			
baendi0_500_500_4000	0	-0.16	0.99	-1.79	46	.08
	1	0.40	0.98			
badediff0_500_500_4000	0	-0.16	0.99	-1.79	46	.08
	1	0.40	0.98			
ltas_max	0	0.22	0.92	2.54	46	.02
	1	-0.55	1.05			
ltas_sd	0	0.21	0.93	2.38	46	.02
	1	-0.52	1.04			
ltas_locpeak	0	-0.23	1.01	-2.57	46	.01
	1	0.55	0.81			

Tabelle C-5: Sprachmaterial “Fahrerassistenzsatz 9“: Mittelwerte (MW), Standardabweichungen (SD) für die Bedingungen “wach“ ($KSS_{kat}=0$; $KSS \leq 5$) und “schläfrig“ ($KSS_{kat}=1$; $KSS > 5$) (Rohwerte, Frauen)

Kennwert	KSS	MW	SD	T	df	p
f0_timemax	0	-0.29	0.93	-2.38	22	.03
	1	0.71	0.94			
f0_maxtime_rel	0	-0.24	0.89	-1.94	22	.07
	1	0.60	1.14			
f1_mean	0	0.25	1.07	2.01	22	.06
	1	-0.61	0.59			
f3_sd	0	0.22	1.03	1.70	22	.10
	1	-0.53	0.84			
f4_maxtime_rel	0	0.22	0.97	1.73	22	.10
	1	-0.54	1.00			
f4_sd	0	0.24	0.88	1.92	22	.07
	1	-0.59	1.16			
df1f2	0	0.25	0.96	2.02	22	.06
	1	-0.62	0.96			
df1f3	0	0.26	1.04	2.10	22	.05
	1	-0.64	0.68			
df1f4	0	0.25	0.99	1.97	22	.06
	1	-0.60	0.88			
MFCC1	0	0.27	1.04	2.16	22	.04
	1	-0.65	0.66			
MFCC4	0	-0.25	0.99	-1.99	22	.06
	1	0.61	0.90			
MFCC5	0	0.24	1.04	1.86	22	.08
	1	-0.57	0.75			
Δ MFCC5	0	-0.30	0.97	-2.46	22	.02
	1	0.72	0.80			
Δ MFCC11	0	0.35	0.87	3.08	22	.01

ANHANG C: MITTELWERTVERGLEICHS-BEZOGENE ERGEBNISSE

	1	-0.86	0.87			
int_min	0	0.33	0.95	2.76	22	.01
	1	-0.79	0.76			
energmax2000_5000	0	0.33	1.04	2.83	22	.01
	1	-0.80	0.21			
hammarberg_s0	0	0.23	1.00	1.80	22	.09
	1	-0.56	0.90			
spec_kurt	0	-0.24	1.01	-1.92	22	.07
	1	0.59	0.84			
spec_skew	0	-0.23	1.04	-1.80	22	.09
	1	0.56	0.77			
spec_sd	0	0.23	1.06	1.82	22	.08
	1	-0.56	0.69			
baen1000_8000	0	0.30	1.06	2.45	22	.02
	1	-0.72	0.35			
baen125_200	0	0.25	0.95	1.95	22	.06
	1	-0.60	1.01			
baen1000_1600	0	0.30	1.06	2.52	22	.02
	1	-0.74	0.31			
baen5000_8000	0	0.23	0.99	1.78	22	.09
	1	-0.55	0.94			
ltas_trend_max	0	0.35	1.00	3.06	22	.01
	1	-0.85	0.36			

Tabelle C-6: Sprachmaterial “Fahrerassistenzsatz 9“: Mittelwerte (MW), Standardabweichungen (SD) für die Bedingungen “wach“ (KSS_kat=0; $KSS \leq 5$) und “schläfrig“ (KSS_kat=1; $KSS > 5$) (Rohwerte, Frauen)

Kennwert	KSS	MW	SD	T	df	p
f0_timemax	0	-0.14	1.01	-1.83	70	.07
	1	0.33	0.95			
f0_75quant	0	0.15	1.00	1.97	70	.05
	1	-0.36	0.95			
f0_mintime_rel	0	0.17	1.02	2.28	70	.03
	1	-0.41	0.87			
f0_maxtime_rel	0	-0.13	0.99	-1.67	70	.10
	1	0.31	1.01			
f2_mean	0	-0.16	1.02	-2.13	70	.04
	1	0.38	0.88			
f2_sd	0	0.14	1.05	1.88	70	.06
	1	-0.34	0.82			
f2_bandw_q25	0	-0.15	1.01	-2.00	70	.05
	1	0.36	0.91			
f3_max	0	0.17	0.92	2.30	70	.03
	1	-0.41	1.12			
f3_maxtime_rel	0	0.16	0.93	2.10	70	.04
	1	-0.38	1.11			
f3_range	0	0.14	0.99	1.90	70	.06
	1	-0.34	0.99			
f3_sd	0	0.17	1.01	2.34	70	.02
	1	-0.42	0.90			
f5_mean	0	0.14	0.95	1.94	70	.06
	1	-0.35	1.08			
f5_bandw_q50	0	-0.13	1.01	-1.70	70	.09
	1	0.31	0.96			
f5_bandw_q75	0	-0.22	1.01	-3.02	70	.00

ANHANG C: MITTELWERTVERGLEICHS-BEZOGENE ERGEBNISSE

	1	0.53	0.80			
df1f2	0	0.18	0.96	2.43	70	.02
	1	-0.43	1.00			
df2f4	0	-0.19	1.01	-2.60	70	.01
	1	0.46	0.86			
df2f5	0	-0.17	0.99	-2.37	70	.02
	1	0.43	0.93			
df3f5	0	-0.14	1.00	-1.89	70	.06
	1	0.34	0.95			
MFCC1	0	0.18	1.01	2.38	70	.02
	1	-0.43	0.89			
MFCC2	0	0.15	0.99	2.06	70	.04
	1	-0.37	0.97			
MFCC4	0	-0.27	0.99	-3.83	70	.00
	1	0.65	0.71			
MFCC5	0	0.13	1.04	1.67	70	.10
	1	-0.31	0.87			
MFCC6	0	0.17	0.98	2.28	70	.03
	1	-0.41	0.97			
Δ MFCC8	0	-0.13	0.98	-1.67	70	.10
	1	0.30	1.04			
Δ MFCC10	0	0.15	1.03	1.94	70	.06
	1	-0.35	0.88			
MFCC11	0	-0.14	0.96	-1.87	70	.07
	1	0.34	1.05			
Δ MFCC11	0	0.13	1.01	1.78	70	.08
	1	-0.32	0.95			
int_mean	0	0.17	0.97	2.32	70	.02
	1	-0.42	1.00			
int_25quant	0	0.20	1.00	2.78	70	.01
	1	-0.49	0.86			
int_75quant	0	0.19	0.96	2.57	70	.01

ANHANG C: MITTELWERTVERGLEICHS-BEZOGENE ERGEBNISSE

	1	-0.46	0.99			
energmax0_2000	0	0.20	0.93	2.68	70	.01
	1	-0.48	1.04			
energmax2000_5000	0	0.15	1.04	2.05	70	.04
	1	-0.37	0.82			
baen0_1000	0	0.16	0.99	2.14	70	.04
	1	-0.39	0.96			
baen200_300	0	0.14	0.97	1.88	70	.07
	1	-0.34	1.05			
baen1000_1600	0	0.17	1.02	2.30	70	.02
	1	-0.41	0.88			
bade200_1000	0	0.14	0.99	1.83	70	.07
	1	-0.33	1.00			
ltas_max	0	0.20	0.93	2.68	70	.01
	1	-0.48	1.04			
ltas_sd	0	0.19	0.94	2.64	70	.01
	1	-0.47	1.03			
ltas_locpeak	0	-0.14	1.02	-1.81	70	.07
	1	0.33	0.91			

**ANHANG D:
ERGEBNISSE DER MERKMALSSELEKTION
(1-NN GENETIC ALGORITHM)**

Tabelle D-1: Sprachmaterial “extrahierter Vokal [o:]“: Selektierte Kennwerte

Selektierte Kennwerte	
Rohwerte	z-Werte
f0_25quant	f0_min_z
f0_75quant	f0_maxtime_rel_z
f0_maxtime_rel	f0_range_z
f1_mintime	f0_slope_z
f1_maxtime	f1_mintime_z
f1_range	f1_maxtime_z
f1_mean	f1_sd_z
f2_bandw_q75	f2_min_z
f3_maxtime_rel	f2_maxtime_rel_z
f3_sd	f2_bandw_q50_z
f4_maxtime	f3_mintime_rel_z
f4_bandw_q50	f3_maxtime_rel_z
f5_min	f4_mintime_rel_z
df1f3	f4_bandw_q50_z
df2f4	f5_min_z
df3f4	f5_mintime_rel_z
df4f5	f5_mean_z
Δ MFCC3	f5_bandw_q25_z
$\Delta\Delta$ MFCC4	f5_bandw_q50_z
MFCC6	MFCC4_z
Δ MFCC6	$\Delta\Delta$ MFCC6_z
MFCC7	$\Delta\Delta$ MFCC7_z
Δ MFCC8	MFCC11_z
$\Delta\Delta$ MFCC8	Δ MFCC11_z
$\Delta\Delta$ MFCC11	Δ MFCC12_z
int_maxtime_rel	int_mintime_rel_z
int_range	openQuot_OQG_z

glottOpen_GOG	skewness_SKG_z
energmax0_2000	imcompleclos_IC_z
hammarberg_s0	spec_cm_z
hammarberg_s3	ltas_slope0_1k_1k_4k_z
spec_kurt	ltas_min_z
spec_sd	ltas_maxfreq_z
baendi0_500_500_4000	

Tabelle D-2: Sprachmaterial "Isolierte Vokalphonation [a:]": Selektierte Kennwerte

Selektierte Kennwerte	
Rohwerte	z-Werte
f0_timemin	f0_75quant_z
f0_25quant	f0_mintime_rel_z
f0_mintime_rel	Shimmer_z
f0_range	f1_min_z
hnr	f2_mean_z
f1_max	f2_sd_z
f1_bandw_q50	f2_bandw_q50_z
f2_min	f3_maxtime_z
f2_bandw_q25	f3_range_z
f3_maxtime_rel	f4_bandw_q50_z
f3_range	$\Delta\Delta\text{MFCC6}_z$
f3_mean	$\Delta\Delta\text{MFCC7}_z$
f3_bandw_q25	$\Delta\Delta\text{MFCC12}_z$
f4_maxtime_rel	int_mean_z
f4_bandw_q50	int_sd_z
f5_maxtime	int_maxtime_rel_z
f5_bandw_q75	imcompclos_IC_z
df1f2	energmax2000_5000_z
df2f4	hammarberg_s1_z
ΔMFCC1	spec_cog_z
MFCC4	baen200_300_z
$\Delta\Delta\text{MFCC4}$	baen1000_1600
ΔMFCC5	ltas_min_z
$\Delta\Delta\text{MFCC5}$	ltas_max
ΔMFCC6	ltas_maxfreq_z
$\Delta\Delta\text{MFCC7}$	ltas_locpeak
$\Delta\Delta\text{MFCC8}$	pe1000_baenqout_z
$\Delta\Delta\text{MFCC9}$	

Δ MFCC10	
int_timemax	
int_25quant	
hammarberg_s0	
spec_sd	
baen200_300	
baen1000_1600	
baen5000_8000	
bade200_1000	
baendi0_500_500_4000	
badediff0_500_500_4000	
ltas_minfreq	
ltas_locpeak	

Table D-3: Sprachmaterial "Fahrerassistenzsatz 4 – Wort 1": Selektierte Kennwerte

Selektierte Kennwerte	
Rohwerte	z-Werte
f0_timemax	f0_75quant_z
f0_range	f0_maxtime_rel_z
f1_mintime	f1_mintime_z
f1_maxtime_rel	f1_range_z
f1_bandw_q75	f1_mean_z
f2_min	f1_sd_z
f2_maxtime_rel	f2_max_z
f2_sd	f2_bandw_q25_z
f2_bandw_q25	f3_mintime_rel_z
f3_min	f3_maxtime_rel_z
f3_bandw_q25	f3_range_z
f4_mintime	f3_bandw_q25_z
f4_bandw_q75	f4_bandw_q25_z
f5_maxtime_rel	f4_bandw_q50_z
f5_range	f5_mintime_z
df1f5	df3f4_z
df3f5	df3f5_z
df4f5	$\Delta\Delta\text{MFCC1}_z$
MFCC1	MFCC4_z
ΔMFCC1	ΔMFCC4_z
$\Delta\Delta\text{MFCC5}$	$\Delta\Delta\text{MFCC5}_z$
$\Delta\Delta\text{MFCC6}$	MFCC6_z
$\Delta\Delta\text{MFCC7}$	ΔMFCC7_z
$\Delta\Delta\text{MFCC10}$	$\Delta\Delta\text{MFCC7}_z$
ΔMFCC11	$\Delta\Delta\text{MFCC8}_z$
ΔMFCC12	ΔMFCC10_z
int_timemax	MFCC11_z
hammarberg_s3	ΔMFCC11_z

hammarberg_s4	int_max_z
baen_hf_lf1000	int_maxtime_rel_z
baen125_200	skewness_SKG_z
bade200_1000	hammarberg_s0_z
badediff0_500_500_4000	spec_kurt_z
ltas_maxfreq	ltas_minfreq_z
ltas_sd	ltas_locpeak_Z
	ltas_trend_max_z

Tabelle D-4: Sprachmaterial “Fahrerassistenzsatz 3“ : Selektierte Kennwerte

Selektierte Kennwerte	
Rohwerte	z-Werte
f0_timemin	f0_75quant_z
f0_range	f1_range_z
f1_min	f1_bandw_q25_z
f1_range	f2_maxtime_rel_z
f2_bandw_q25	f3_maxtime_z
f2_bandw_q50	f3_bandw_q25_z
f2_bandw_q75	f3_bandw_q50_z
f3_min	f4_maxtime_z
f3_range	f4_bandw_q75_z
f3_bandw_q75	f5_mintime_z
f4_maxtime_rel	f5_mintime_rel_z
f5_maxtime	df3f4_z
f5_sd	df4f5_z
f5_bandw_q25	MFCC8_z
df1f2	MFCC9_z
df2f3	Δ MFCC10_z
df2f4	$\Delta\Delta$ MFCC12_z
df2f5	int_min_z
df3f4	int_timemin_z
$\Delta\Delta$ MFCC4	int_max_z
Δ MFCC6	int_25quant_z
Δ MFCC7	int_mintime_rel_z
$\Delta\Delta$ MFCC7	glottOpen_GOG_z
$\Delta\Delta$ MFCC8	rateClos_RCG_z
MFCC9	imcomplclos_IC_z
Δ MFCC12	energmax2000_5000_z
int_75quant	hammarberg_s1_z
rateClos_RCG	baen200_300_z

spec_kurt	pe1000_baenqout_z
spec_skew	
spec_sd	
baen125_200	
bade200_1000	
baendi0_500_500_4000	
badediff0_500_500_4000	
ltas_maxfreq	
ltas_locpeak	

Tabelle D-5: Sprachmaterial “Fahrerassistenzsatz 9“: Selektierte Kennwerte

Selektierte Kennwerte	
Rohwerte	z-Werte
f0_sd	f0_sd_z
f0_25quant	f0_mintime_rel_z
f0_75quant	f0_range_z
f0_maxtime_rel	f0_slope_z
f1_maxtime_rel	f0_slope_octjump_z
f1_mean	f0_mean_z
f1_sd	f1_mintime_z
f2_mintime	f1_sd_z
f3_sd	f1_bandw_q75_z
f4_mintime	f2_maxtime_rel_z
f4_maxtime	f3_mintime_z
f4_sd	f3_max_z
f4_bandw_q25	f3_bandw_q50_z
f4_bandw_q50	f5_mintime_rel_z
f4_bandw_q75	f5_mean_z
f5_min	f5_bandw_q75_z
f5_mintime	df3f4_z
df2f5	MFCC2_z
MFCC2	MFCC4_z
MFCC4	Δ MFCC6_z
MFCC6	$\Delta\Delta$ MFCC6_z
deltazwei8	MFCC8_z
MFCC11	$\Delta\Delta$ MFCC8_z
int_min	MFCC10_z
int_maxtime_rel	$\Delta\Delta$ MFCC11_z
int_range	MFCC12_z
glottOpen_GOG	openQuot_OQG_z
hammarberg_s3	glottOpen_GOG_z

hammarberg_s4	skewness_SKG_z
spec_kurt	rateClos_RCG_z
spec_sd	energmax2000_5000_z
ltas_min	spec_cm_z
ltas_sd	baendi0_500_500_4000_z
	ltas_slope0_1000_1000_4000_z
	ltas_maxfreq_z

Tabelle D-6: Sprachmaterial "Intro": Selektierte Kennwerte

Selektierte Kennwerte	
Rohwerte	z-Werte
f0_timemax	f0_timemin_z
f0_sd	f0_mean_z
f0_25quant	f1_mintime_z
f0_range	f1_range_z
f1_mintime	f1_mean_z
f1_maxtime_rel	f1_sd_z
f1_mean	f1_bandw_q25_z
f2_maxtime_rel	f2_max_z
f2_mean	f2_maxtime_rel_z
f2_sd	f2_bandw_q25_z
f2_bandw_q25	f3_mintime_rel_z
f3_min	f3_max_z
f3_bandw_q75	f3_range_z
f4_max	f4_bandw_q50_z
f4_maxtime	f5_mintime_z
f5_maxtime_rel	df3f4_z
f5_range	df3f5_z
df3f4	$\Delta\Delta\text{MFCC2}_z$
df4f5	$\Delta\Delta\text{MFCC3}_z$
MFCC1	ΔMFCC4_z
ΔMFCC1	$\Delta\Delta\text{MFCC5}_z$
MFCC2	$\Delta\Delta\text{MFCC7}_z$
ΔMFCC3	ΔMFCC8_z
$\Delta\Delta\text{MFCC5}$	$\Delta\Delta\text{MFCC8}_z$
$\Delta\Delta\text{MFCC6}$	ΔMFCC10_z
MFCC9	MFCC11_z
ΔMFCC9	ΔMFCC11_z
ΔMFCC11	int_min_z

int_min	int_maxtime_rel_z
int_timemax	skewness_SKG_z
int_mean	baen_hf_lf1000_z
int_75quant	baendi0_500_500_4000_z
hammarberg_s4	ltas_locpeak_z
baen0_1000	ltas_trend_max_z
baen125_200	
bade200_1000	
badeddiff0_500_500_1000	
ltas_maxfreq	
ltas_sd	
ltas_trend_max	

ANHANG E:

PROBANDENBEZOGENE DETAILERGEBNISSE (SPEKTROGRAMM)

Sprachmaterial „extrahierter Vokal [a:]“

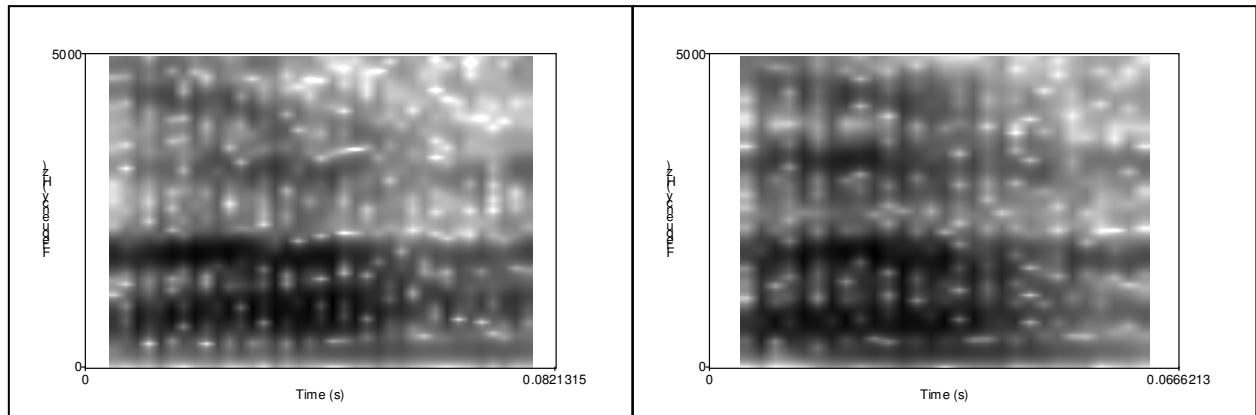


Abbildung E-1: Spektrogramm des extrahierten Vokals [a:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 10)

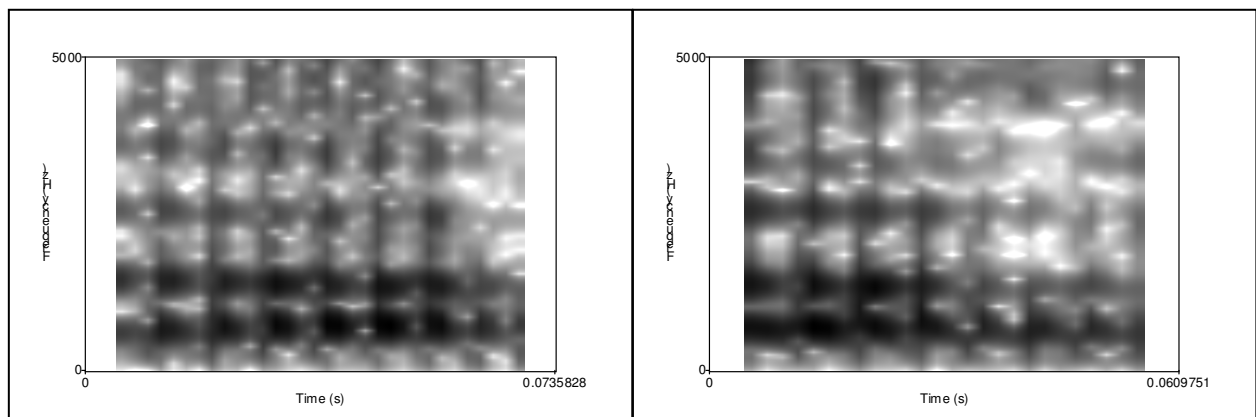


Abbildung E-2: Spektrogramm des extrahierten Vokals [a:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 22)

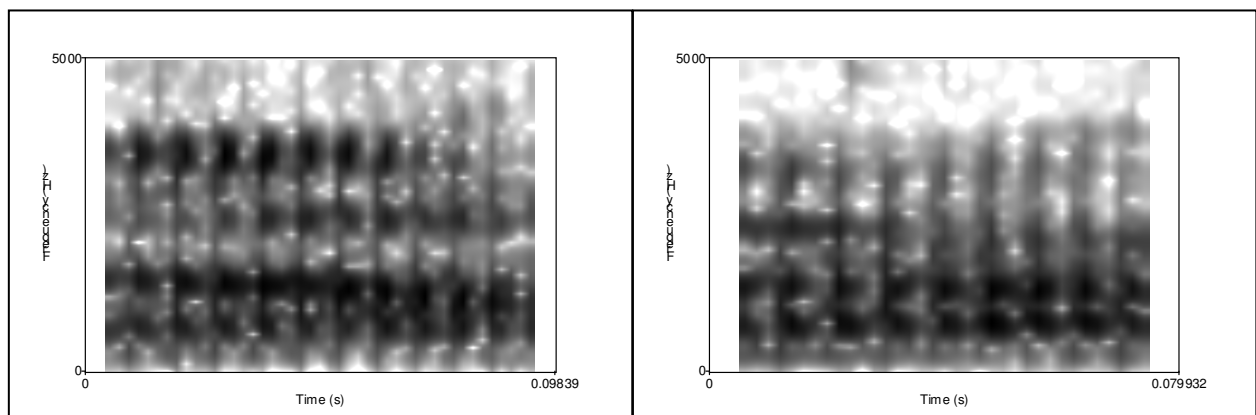


Abbildung E-3: Spektrogramm des extrahierten Vokals [a:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 26)

Sprachmaterial „extrahierter Vokal [o:]“

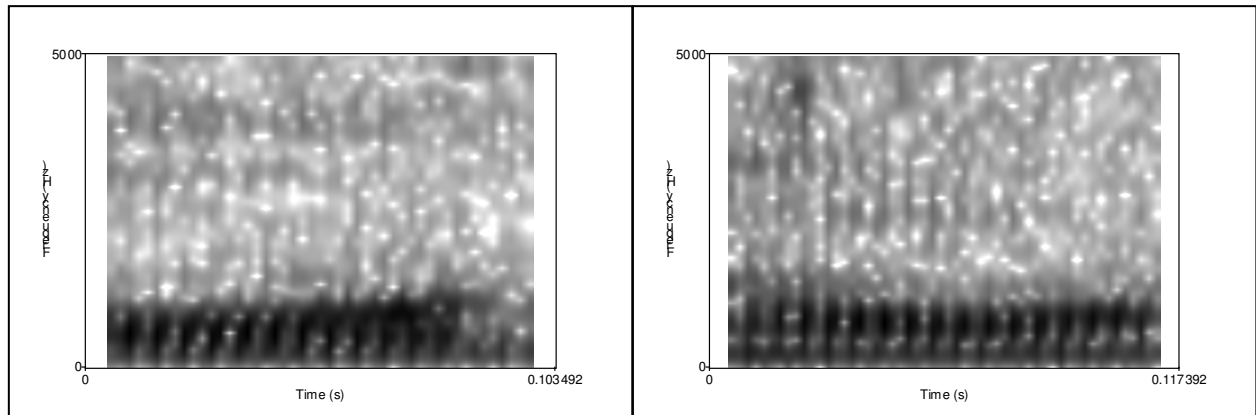


Abbildung E-4: Spektrogramm des extrahierten Vokals [o:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 10)

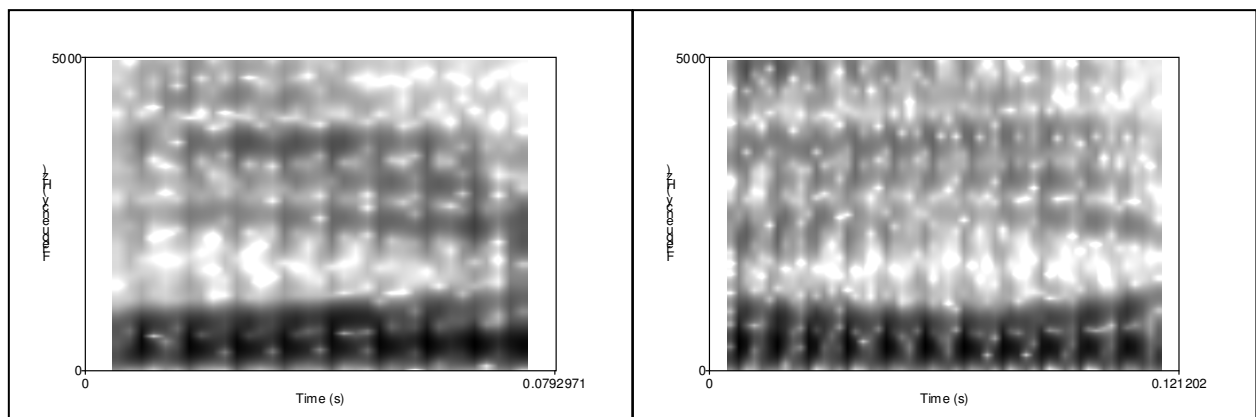


Abbildung E-5: Spektrogramm des extrahierten Vokals [o:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 22)

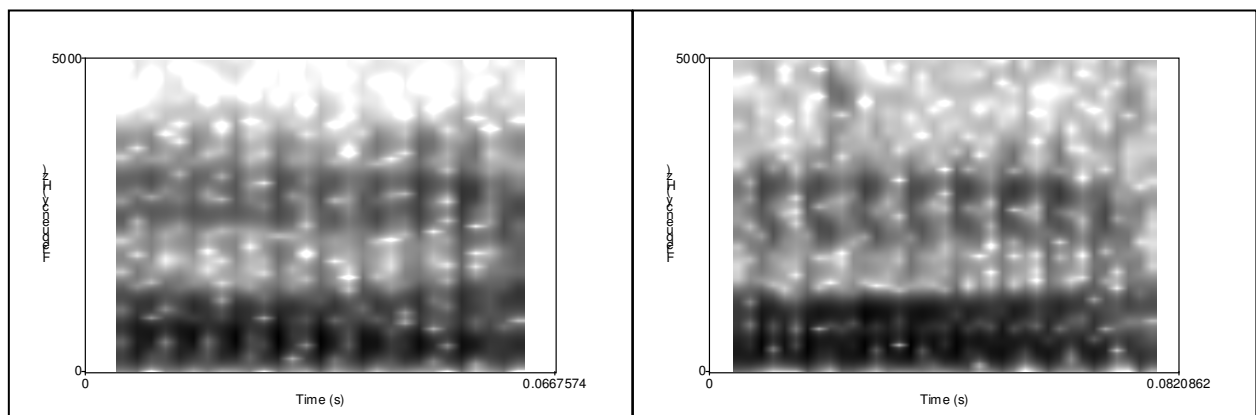


Abbildung E-6: Spektrogramm des extrahierten Vokals [o:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 26)

Sprachmaterial „Vokalphonation [a:]“

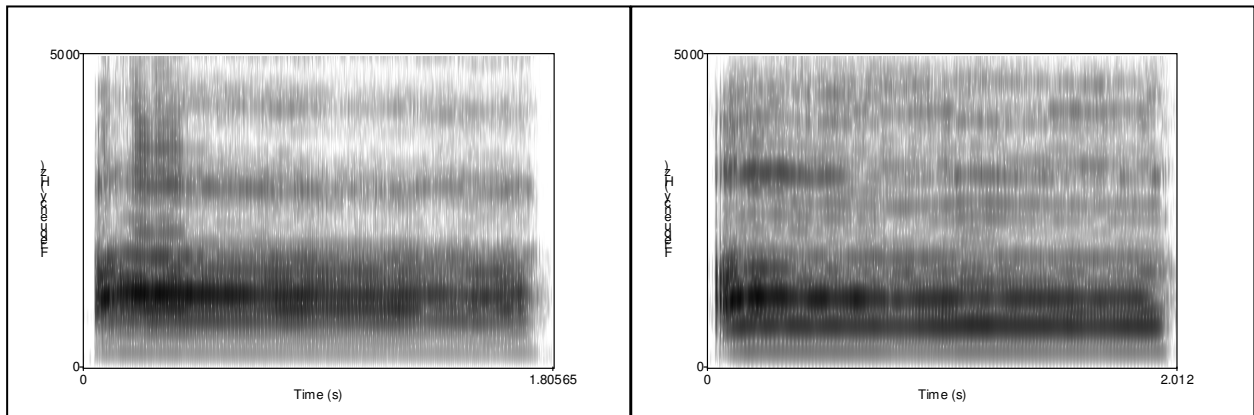


Abbildung E-7: Spektrogramm der 2 Sek. isolierten Vokalphonation [a:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 10)

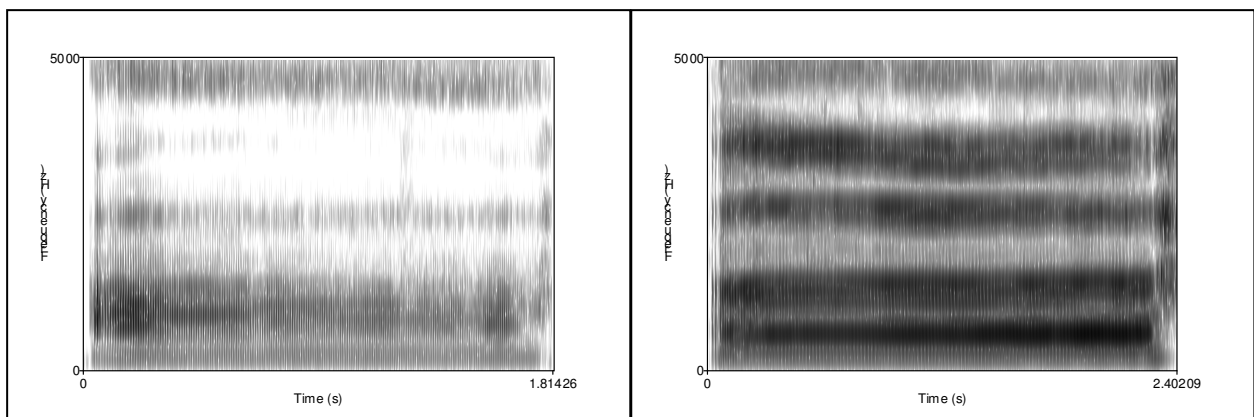


Abbildung E-8: Spektrogramm der 2 Sek. isolierten Vokalphonation [a:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 22)

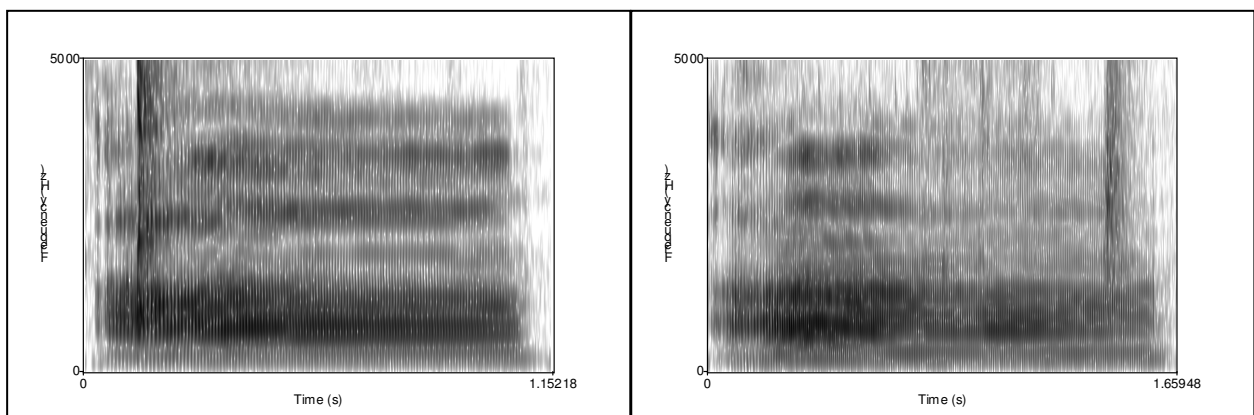


Abbildung E-9: Spektrogramm der 2 Sek. isolierten Vokalphonation [a:] für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 26)

Sprachmaterial Wort „suchen“

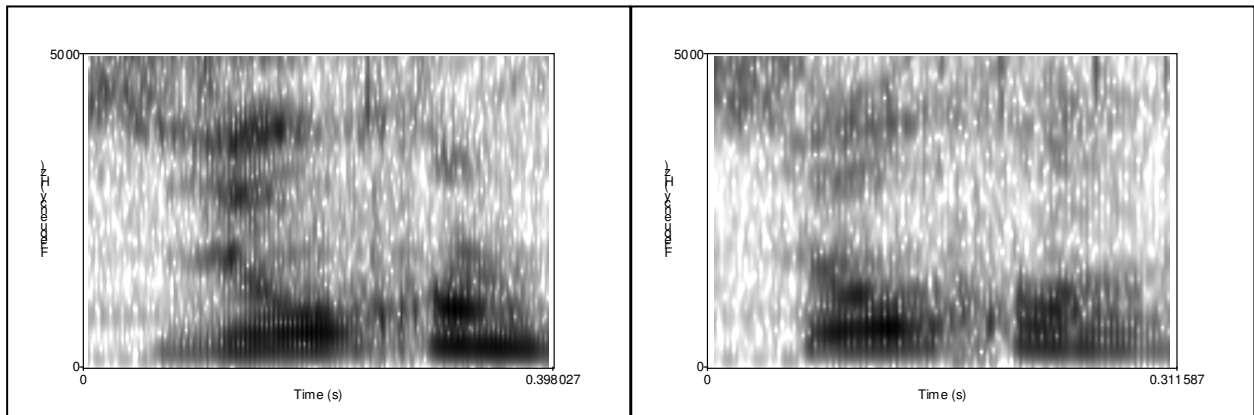


Abbildung E-10: Spektrogramm des extrahierten Wortes „suchen“ für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 10)

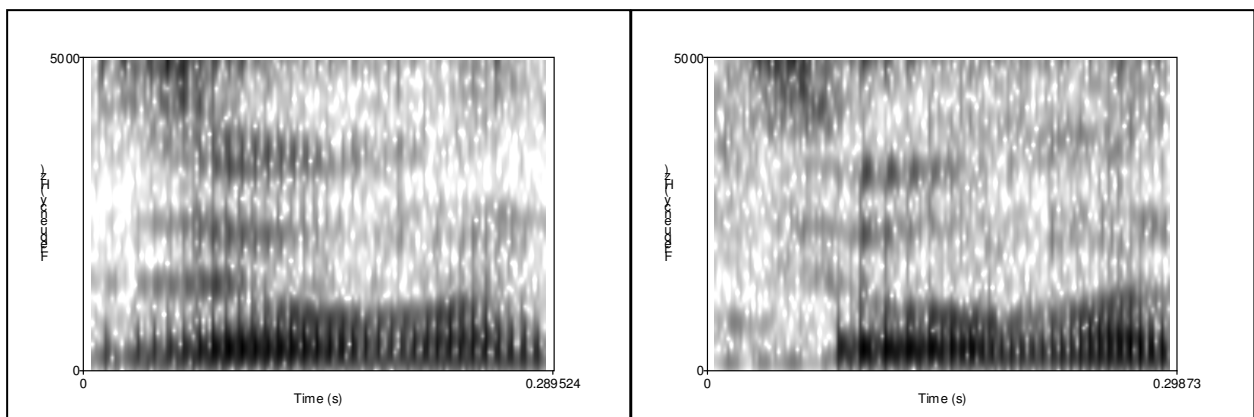


Abbildung E-11: Spektrogramm des extrahierten Wortes „suchen“ für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 22)

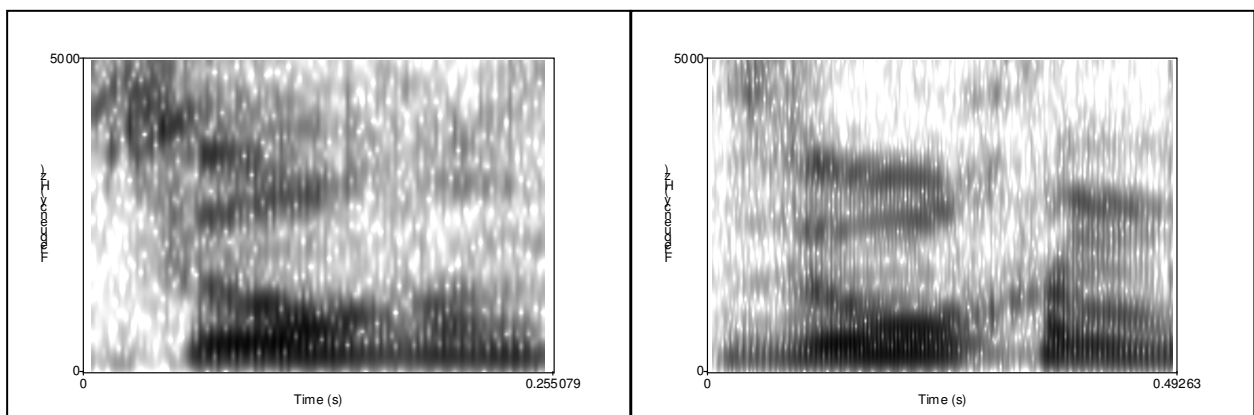


Abbildung E-12: Spektrogramm des extrahierten Wortes „suchen“ für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 26)

Sprachmaterial „Fahrerassistenzsatz 3“

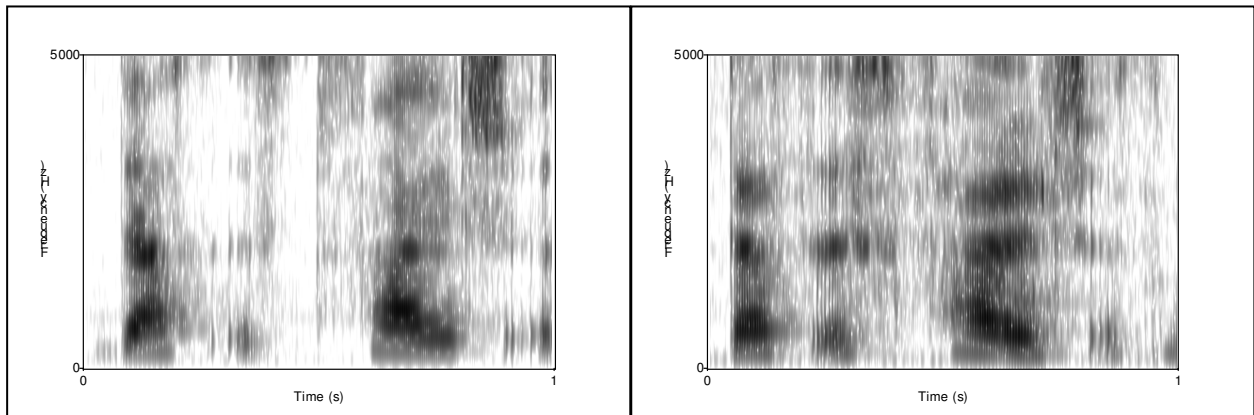


Abbildung E-13: *Spektrogramm des Fahrerassistenzsatzes 3 für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 10)*

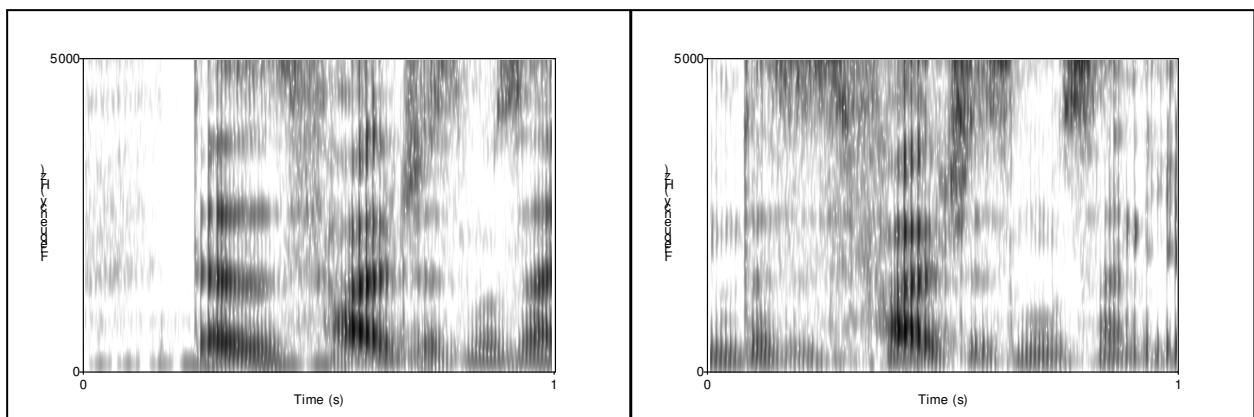


Abbildung E-14: *Spektrogramm des Fahrerassistenzsatzes 3 für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 22)*

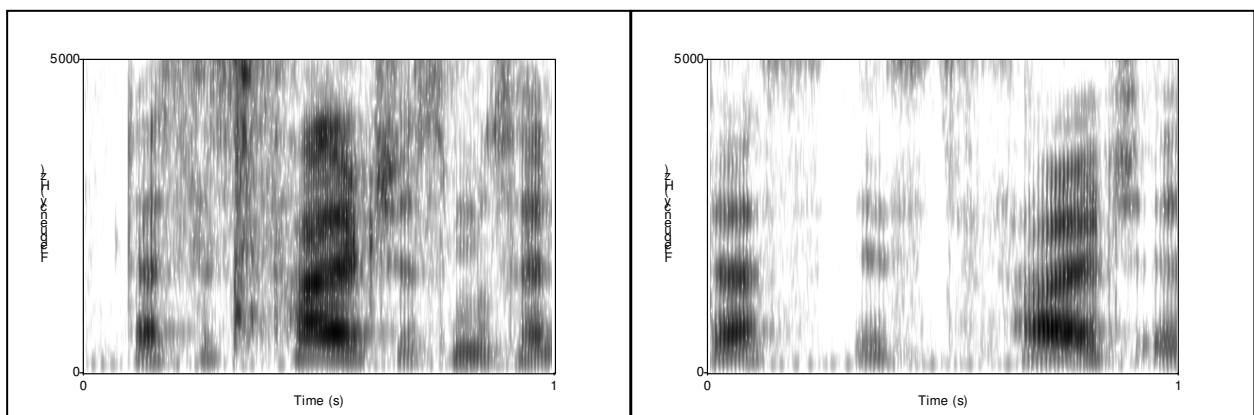


Abbildung E-15: *Spektrogramm des Fahrerassistenzsatzes 3 für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 26)*

Sprachmaterial „Fahrerassistenzsatz 9“

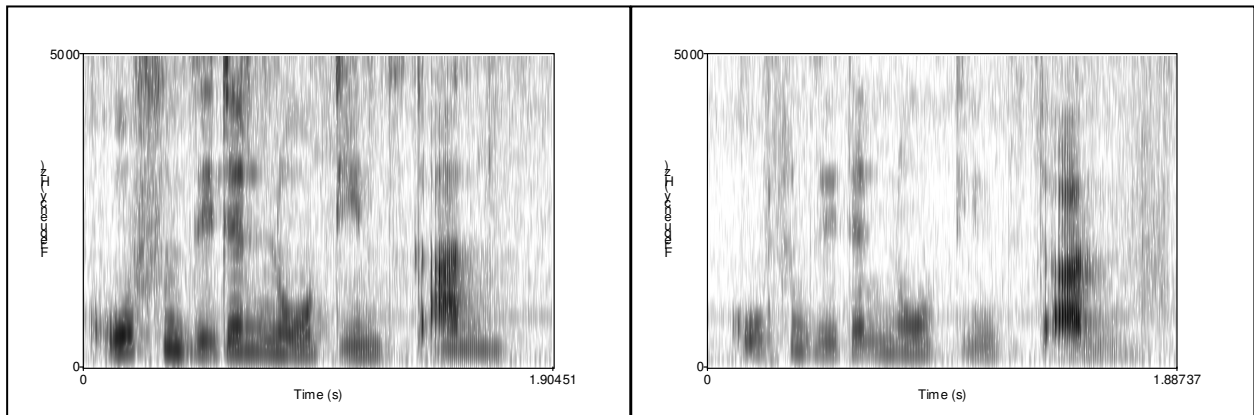


Abbildung E-16: Spektrogramm des Fahrerassistenzsatzes 9 („Rufen Sie den Notdienst an“) für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 10)

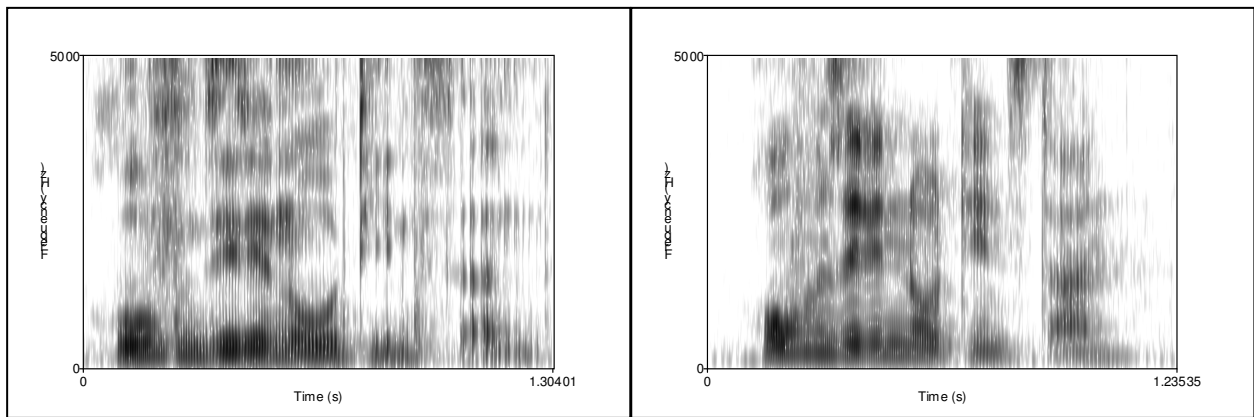


Abbildung E-17: Spektrogramm des Fahrerassistenzsatzes 9 („Rufen Sie den Notdienst an“) für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 22)

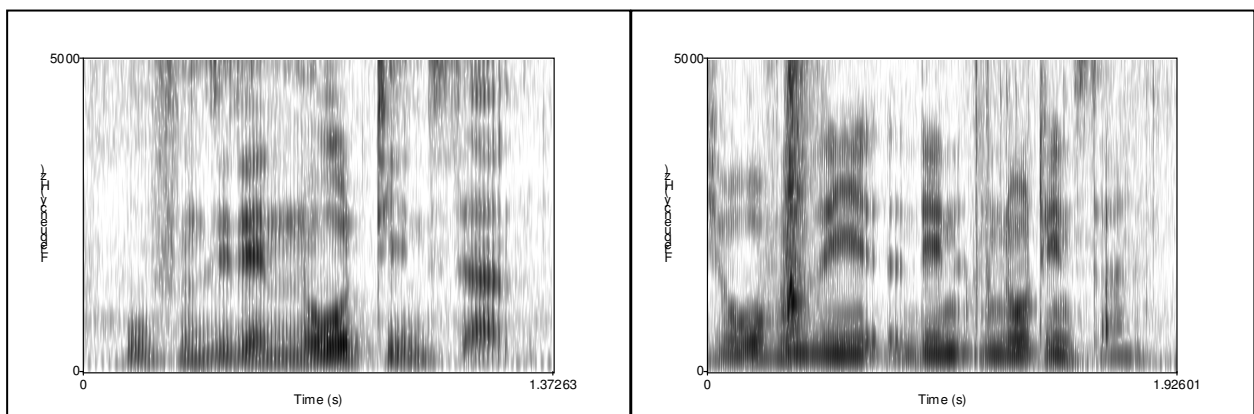


Abbildung E-18: Spektrogramm des Fahrerassistenzsatzes 9 („Rufen Sie den Notdienst an“) für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 26)

Sprachmaterial „Introphrase“

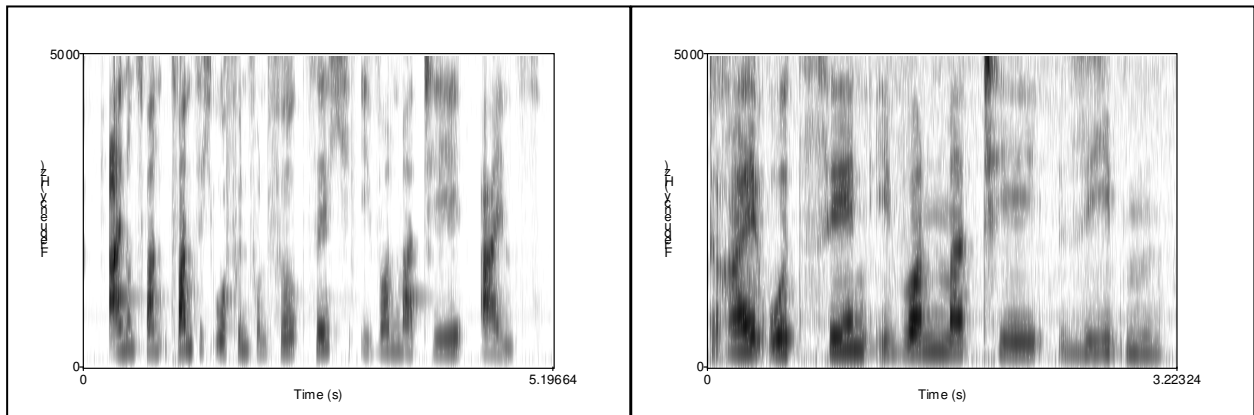


Abbildung E-19: Spektrogramm der Introphrase (z.B. „19.50 Uhr, Vp 10, Sleepiness Scale Wert 2“) für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 10)

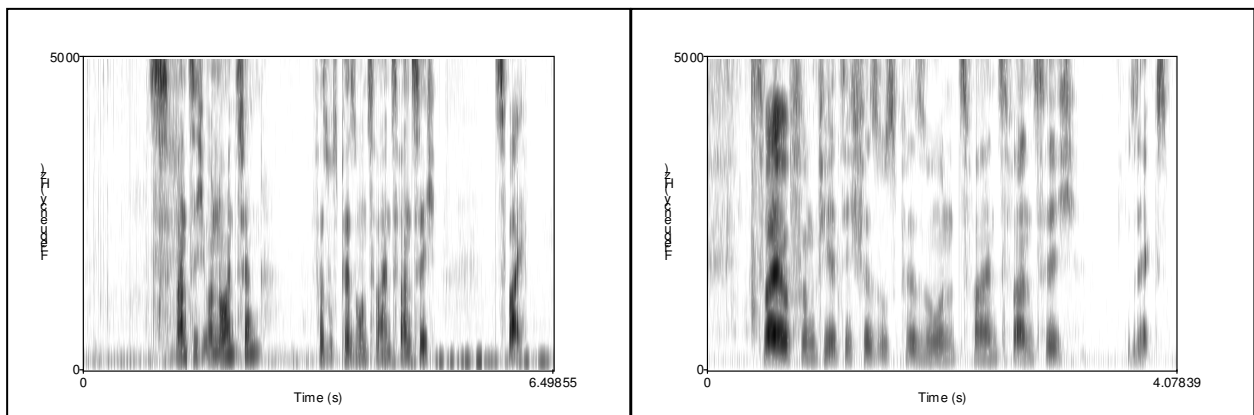


Abbildung E-20: Spektrogramm der Introphrase (z.B. „19.50 Uhr, Vp 10, Sleepiness Scale Wert 2“) für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 22)

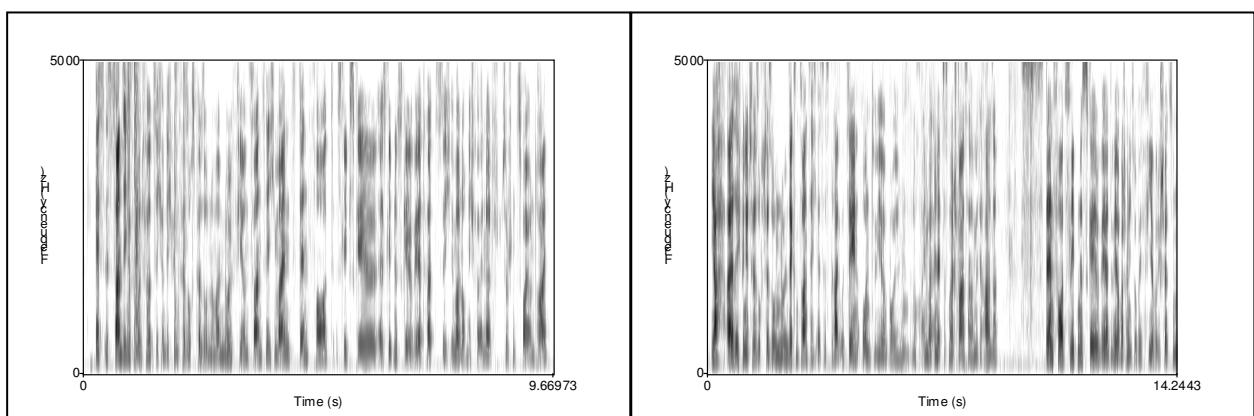


Abbildung E-21: Spektrogramm der Introphrase (z.B. „19.50 Uhr, Vp 10, Sleepiness Scale Wert 2“) für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 26)

Sprachmaterial „freie Bildbeschreibung“

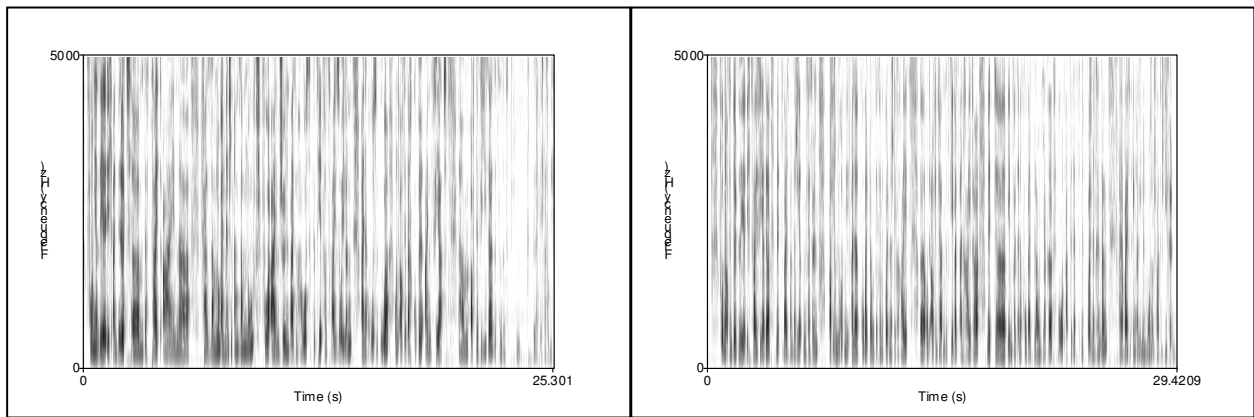


Abbildung E-22: *Spektrogramm der freien Bildbeschreibung für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 10)*

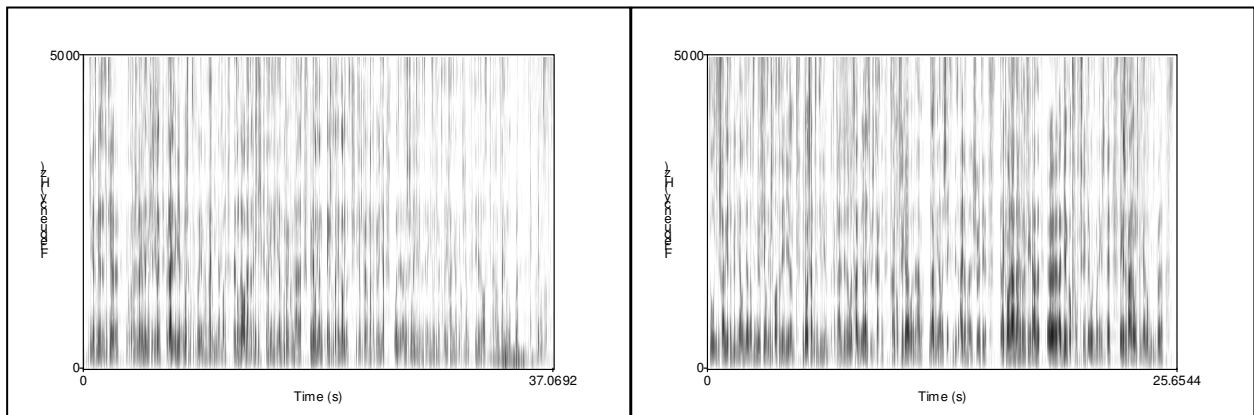


Abbildung E-23: *Spektrogramm der freien Bildbeschreibung für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 22)*

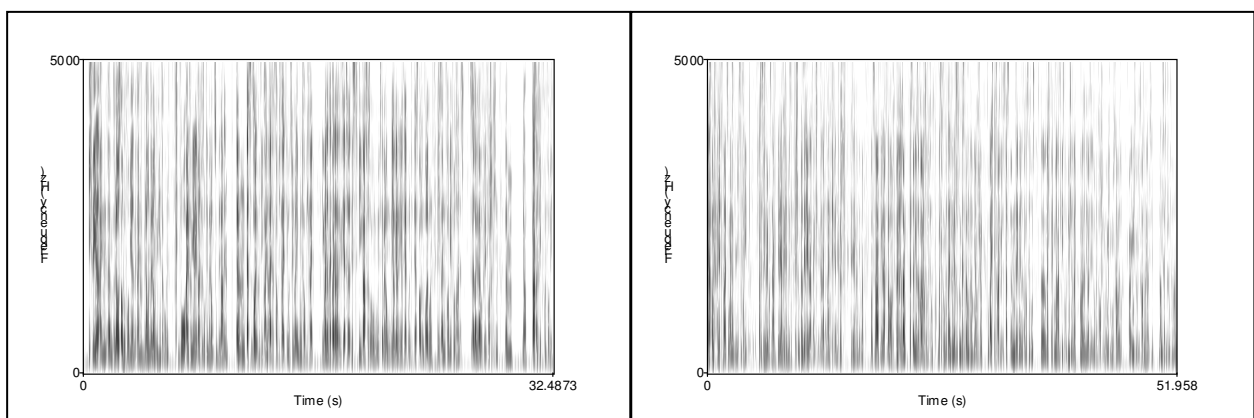


Abbildung E-24: *Spektrogramm der freien Bildbeschreibung für ein waches ($KSS \leq 5$) und ein müdes ($KSS > 5$) Sprachbeispiel (Vp 26)*