Untersuchungen der rhythmischen Struktur von Sprache unter Alkoholeinfluss

Christian Heinrich

München 2014

Untersuchungen der rhythmischen Struktur von Sprache unter Alkoholeinfluss

Christian Heinrich

Inaugural-Dissertation

zur Erlangung des Doktorgrades der Philosophie

an der Ludwig–Maximilians–Universität

München

vorgelegt von

Christian Heinrich

aus München

München 2014

Referent: PD Dr. habil. Dr.-Ing. Florian Schiel

Korreferent: Prof. Dr. Jonathan M. Harrington

Tag der mündlichen Prüfung: 17. Februar 2014

Abstract

This thesis is concerned with the rhythmical structure of speech under the influence of alcohol. All analyses presented are based on the Alcohol Language Corpus, which is a collection of speech uttered by 77 female and 85 male sober and intoxicated speakers. Experimental research was carried out to find robust, automatically extractable features of the speech signal that indicate speaker intoxication. These features included rhythm measures, which reflect the durational variability of vocalic and consonantal elements and are normally used to classify languages into different rhythm classes. The durational variability was found to be greater in the speech of intoxicated individuals than in the speech of sober individuals, which suggests, that speech of intoxicated speakers is more irregular than speech of sober speakers. Another set of features describes the dynamics of the short-time energy function of speech. Therefore different measures are derived from a sequence of energy minima and maxima. The results also reveal a greater irregularity in the speech of intoxicated individuals. A separate investigation about speaking rate included two different measures. One is based on the phonetic segmentation and is an estimate of the number of syllables per second. The other is the mean duration of the time intervals between successive maxima of the short-time energy function of speech. Both measures denote a decreased speaking rate in the speech of intoxicated speakers compared to speech uttered in sober condition. The results of a perception experiment show that a decrease in speaking rate also is an indicator for intoxication in the perception of speech. The last experiment investigates rhythmical features based on the fundamental frequency and energy contours of speech signals. Contours are compared directly with different distance measures (root mean square error, statistical correlation and the Euclidean distance in the spectral space of the contours). They are also compared by parameterization of the contours using Discrete Cosine Transform and the first and second moments of the lower DCT spectrum. A Principal Components Analysis on the contour data was also carried out to find fundamental contour forms regarding the speech of intoxicated and sober individuals. Concerning the distance measures, contours of speech signals uttered by intoxicated speakers differ significantly from contours of speech signals uttered in sober condition. Parameterization of the contours showed that fundamental frequency contours of speech signals uttered by intoxicated speakers consist of faster movements and energy contours of speech signals uttered by intoxicated speakers of slower movements than the respective contours of speech signals uttered in sober condition. Principal Components Analysis did not find any interpretable fundamental contour forms that could help distinguishing contours of speech signals of intoxicated speakers from those of speech uttered in sober condition. All analyses prove that the effects of alcoholic intoxication on different features of speech cannot be generalized but are to a great extent speaker-dependent.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Kapitel 1

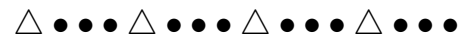
Einführung

1.1 Rhythmus

In der Musik bezeichnet der Begriff Rhythmus die zeitliche Struktur der Töne oder Klänge, die durch eine Abfolge von Dauern und Pausen festgelegt wird. Diese Abfolge schafft eine Art Akzentmuster, das über dem Grundpuls ein rhythmisches Konstrukt errichtet, welches sich in unterschiedlichster Weise in der musikalischen Welt wiederfindet. Rhythmus lässt sich aber auch in der Natur beobachten. Der biologische Rhythmus von Organismen umfasst deren in einer gewissen Regelmäßigkeit wiederkehrende Beschaffenheiten oder Zustände. So sind beispielsweise Schlaf und Wachsein wesentliche Bestandteile des Lebens von Menschen und Tieren und größtenteils durch den Tag-Nacht-Wechsel bedingt. Wiederkehrende Bedingungen in der Natur, die einen Zyklus beschreiben, wie Ebbe und Flut, unterliegen einem universalen Rhythmus. Nach De Groot [1968] wird der Begriff oftmals für jegliche Art von Wiederholungen oder Periodizität in der physischen Welt benutzt.

Niebuhr [2009] verwendet den Begriff *perceptual prominence* und erklärt ihn an einem optischen Beispiel. Durch Formgebung, Farbe und Größe wird innerhalb einer Anordnung von 16 Symbolen (bei Niebuhr stern- und kreisförmige Objekte) eine gewisse Struktur geschaffen. Diese Kette sei beispielhaft durch eine Reihe

großer füllungsloser Dreiecke und kleiner schwarzer Kreisflächen anhand der folgenden Grafik (nach Niebuhr [2009]) dargestellt.



Eine einfache Beschreibung dieser Kette aus Symbolen würde beinhalten, dass 4 der 16 Symbole im Vergleich größer, weniger rund und heller sind als die übrigen 12. Etwas Wesentliches ist in dieser Beschreibung jedoch nicht enthalten. Im Kontext der kleinen schwarzen Kreisflächen sind die großen Dreiecke besonders auffallend, sie erregen mehr Aufmerksamkeit als die Kreisflächen. Dieses Phänomen bezeichnet Niebuhr [2009] als *perceptual prominence* (in der Wahrnehmungspsychologie auch als *perzeptuelle Salienz* bezeichnet). Zu sehen ist demnach keine einfache Kette aus 16 Symbolen, sondern eine Kette aus 4 salienten Dreiecken und je 3 weniger salienten Kreisflächen. Damit schafft dieses Muster eine gewisse Struktur innerhalb der Kette, nämlich eine Sequenz aus 4 Gruppen, die jeweils mit einem salienten Dreieck beginnen. Ein Effekt dieser perzeptiven Gruppierung ist das erleichterte Zählen der Symbole. Bei einer Kette aus 16 gleichen Symbolen lässt sich auf den ersten Blick schwerer eine Gesamtanzahl von 16 erkennen als bei einer 'strukturierten' Kette wie in der Darstellung (4×4 Symbole). Überträgt man dieses Konzept der perzeptuellen Salienz, das sich laut Handel [1986] unter verschiedenen Voraussetzungen ähnlich zeigen kann, auf die Akustik, so können auch akustische Objekte salient sein, insofern sie im Kontext anderer akustischer Objekte durch ihre physikalischen Eigenschaften auffallend abweichen. Allen [1973] sieht als einfachste Form einer hörbaren Sequenz ein rhythmisches Schlagen (z.B. Metronom), das sich durch ein Alternieren von Vorhandensein und Abwesenheit des Signals äußert. Ein vorhandenes Signal ist zweifellos salient in einer solchen Sequenz, aber auch die Stillephasen können als rhythmisches Grundelement angesehen werden. Die Prominenz beeinflussende Parameter im Bereich der Akustik können Frequenz, Dauer, Intensität und Klangqualität sein. Beispielsweise werden Töne mit

ansteigender oder fallender Grundfrequenz (F0), längerer Dauer oder höherer Intensität gegenüber anderen als salient perzipiert (Niebuhr [2009]).
Vergleichbar zeigt sich dies auch in Sprache. Laut Niebuhr [2009] führen die Unterschiede in der Prominenz der Silben zu einer perzeptiven Gruppierung. Falls diese Gruppierung zu einer Reihe von sich wiederholenden Mustern mit jeweils mehr oder weniger prominenten Elementen führt, kann laut Niebuhr [2009] das dadurch entstehende perzeptive Charakteristikum auf auditorischer Ebene auch als *Rhythmus* bezeichnet werden. Dabei gibt es aber keinen direkten Zusammenhang zwischen der perzeptuellen Salienz und einem akustischen Maß, da eine Zuweisung von Prominenzpositionen genau wie in der Musik nur im Kopf des Zuhörers stattfindet. Deswegen ist nach Niebuhr [2009] Rhythmus in der Sprache ein perzeptives Phänomen. Ob er aber anhand von Mustern und Prominenz beschrieben werden kann, ist noch nicht bewiesen, weil weiterhin unklar bleibt, ob und inwieweit solche Muster bei der Perzeption von Rhythmus in der Sprache eine Rolle spielen. Deshalb sollte sich die Untersuchung von Rhythmus in der Sprache nicht nur auf einer Ebene bewegen, wie z.B. Dauerberechnungen (siehe Kapitel 3), sondern die Interaktion verschiedener akustischer Parameter wie der Grundfrequenz, der Energie und auch der Dauer in Betracht gezogen werden. Auch kann Rhythmus kein eindimensionales Phänomen sein, das sich nur durch Muster aus mehr oder weniger prominenten Elementen darstellen lässt. Zusätzlich sind Muster auf Ebene der Töne und Melodien in Betracht zu ziehen (Niebuhr [2009]). Allen [1973] sieht die Bezeichnung 'Rhythmus' als eher problematisch und spricht daher von Mustern mit Spezifikationen auf dynamischer, melodischer und quantitativer Ebene.
Eine weitere wichtige Bedeutung hat Rhythmus in der Poesie, wo er sich durch die Verwendung und Abfolge verschiedener Akzentmuster in der Metrik widerspiegelt. Die Beleuchtung von Sprechrhythmus bei normaler gesprochener Sprache durch eine Analyse der rhythmischen Struktur von Poesie wurde unter anderem von Lehiste [1990] vorgeschlagen.

Die Untersuchungen, die in dieser Arbeit vorgestellt werden, widmen sich einzelnen, auf Basis des akustischen Signals extrahierbaren Parametern, die jeweils einen Teilaspekt von Sprechrhythmus erklären. Eine Kombination verschiedener Parameter wurde dabei nicht vorgenommen. Das zugrundeliegende Sprachmaterial umfasst Sprache, die unter Alkoholeinfluss (im Folgenden auch als *a-Sprache* bezeichnet) und in nüchternem Zustand (im Folgenden auch als *na-Sprache* bezeichnet) geäußert wurde. Deshalb werden zunächst die Auswirkungen von Alkohol auf den menschlichen Körper beschrieben.

1.2 Alkohol

Im Kontext dieser Arbeit bezieht sich der Begriff Alkohol auf *Ethanol* bzw. die chemische Verbindung CH_3CH_2OH, die auch Bestandteil von alkoholischen Getränken ist. Dieses kleine, schwach geladene Molekül besitzt die Eigenschaft, durch Diffusion durch die Membranen im Körper zu gelangen und damit verbreitet zu werden. Durch orale Aufnahme wird bereits eine minimale Menge an Alkohol im Mund durch die Schleimhäute absorbiert, der Großteil wird jedoch im Magen-Darm-Trakt aufgenommen. Die schnellste Aufnahme findet hier im Dünndarm statt, über die Magenschleimhaut und den Dickdarm erfolgt die Absorption langsamer. Insgesamt werden 80% der absorbierten Menge Alkohol im Darm aufgenommen und 20% im Magen. Die Verbreitung im Körper findet wie bei Wasser durch die Gefäßversorgung und Blutbahn statt, die Aufnahme in das jeweilige Gewebe ist dabei abhängig von dessen Wassergehalt. Gut versorgte und durchblutete Gewebe und Organe wie Leber, Lunge, Nieren oder Gehirn nehmen Alkohol durch Diffusion, mit dem Bestreben das Konzentrationsgefälle auszugleichen (Alkohol wird so lange in die Zelle(n) diffundieren, bis die Konzentrationen innerhalb und außerhalb der Zelle(n) gleich sind), schneller auf (innerhalb weniger Minuten) als beispielsweise die Skelettmuskulatur. Knochen und Fett nehmen nur sehr wenig

Alkohol auf. Der Alkoholabbau findet größtenteils über den Stoffwechselprozess statt (90%-98%). Nur ein kleiner Teil Alkohol wird direkt ausgeschieden (2%-10%). Der Metabolismus von Alkohol ist ein schrittweiser enzymatischer Oxidationsprozess, der in Kohlendioxid und Wasser resultiert. Dieser Prozess findet zu einem großen Teil in der Leber statt und nur zu einem kleinen Teil im Magen (Chin and Pisoni [1997]).

Wie bereits angeführt, hat Alkohol pharmakologische Auswirkungen auf den Körper, z.B. im Magen-Darm-Trakt, der Leber, den Nieren und dem Herz-Kreislauf-System. Im Bezug auf Sprache sind aber vor allem seine sedativen Effekte auf das zentrale Nervensystem von Bedeutung. Dabei lassen sich dosierungsabhängige Effekte beobachten, ähnlich wie bei Narkosemitteln, Schlafmitteln oder Beruhigungsmitteln. Unter anderem zählt zu den Effekten ein sinkendes Erregungsniveau. In geringer Dosierung kann Alkohol anregende Effekte hervorrufen, danach wirkt er im Allgemeinen narkotisierend. Symptome bei geringer Dosierung können sich im Gemütszustand der alkoholisierten Person, wie leichter Euphorie und Entspannung, oder wie in einigen wenigen Fällen, einer Steigerung der psychomotorischen Leistung, widerspiegeln. Erhöhte bzw. hohe Dosierungen jedoch können unter anderem zu Wut und Traurigkeit, Störungen in der Bewegungskoordination und eingeschränkten Reflexen, Benommenheit, Bewußtlosigkeit und nicht zuletzt Koma und Tod führen. Sprache scheint dabei ab einer Blutalkoholkonzentration (BAK bzw. Englisch BAC) von 1‰ undeutlicher zu werden (Chin and Pisoni [1997]).

Die Blutalkoholkonzentration ist ein Maß für die Menge von Alkohol im Blut. In Deutschland wird die Blutalkoholkonzentration üblicherweise in $\frac{g}{kg}$ bzw. ‰ (Promille) angegeben. Deshalb wurde dieser Standard auch innerhalb dieser Arbeit verwendet.

1.3 Alkohol und Sprache

Alkohol wirkt sich auf die Sprechweise von Menschen aus, genauso wie Stress, Müdigkeit oder Krankheit. Diese Hypothese findet größtenteils Zustimmung und wird durch die Behauptung vieler Menschen untermauert, dass sie die Alkoholisierung einer Person anhand ihrer sprachlichen Äußerungen erkennen können, auch wenn ihnen diese Person gänzlich unbekannt ist. Entspräche dies tatsächlich der Wahrheit, sollten automatische Erkennungsverfahren ebenfalls in der Lage sein, ein Sprachsignal dahingehend zu analysieren und eine etwaige Alkoholisierung des Sprechers anhand aus dem Sprachsignal extrahierbarer Parameter aufzudecken.
Eine der ersten Veröffentlichungen zum Thema Alkohol und Sprache lieferten Trojan and Kryspin-Exner [1968]. Jedoch sind darin keine Berichte zu Messungen am Sprachsignal enthalten, genauso wie in Sobell et al. [1982], wo lediglich eine Auszählung von Versprechern vorgenommen wurde. In der forensischen Phonetik gibt es diverse Untersuchungen über Alkohol und Sprache. Das Tankerunglück der Exxon Valdez, bei welchem der Verdacht bestand, dass der Kapitän zum Zeitpunkt des Unglücks alkoholisiert war, war der Anlass für eine, frühere Untersuchungen zusammenfassende Studie von Johnson et al. [1990], die folgende phonetische Parameter als indizierend für Alkoholisierung, Stress, Müdigkeit, usw. erachtet: Grundfrequenz und ihre Variabilität, spektrale Neigung und Sprechgeschwindigkeit. Braun [1991] analysierte gelesene Sprache von 33 männlichen Sprechern unter Alkoholeinfluss zwischen 19 und 24 Jahren. Es wird über eine Zunahme der Nasalierung, undeutliche Artikulation, eine erhöhte Versprecherrate und eine Verlängerung von phonetischen Segmenten (entspricht einer Verringerung der Sprechgeschwindigkeit) berichtet. Künzel et al. [1992] erweiterte die Analyse auf Basis desselben Materials auf semi-spontane Sprache (anhand von Bildbeschreibungen) und eine größere Anzahl untersuchter Merkmale. Gemessen wurden fehlerhafte Aussprache (Auslassungen, Einfügungen, Substitutionen und Wiederholungen von Phonemen, Silben und Wörtern), Fehler im Intonationsver-

lauf, Nasalierung und Denasalierung, Lautlängungen, unvollständige Artikulationen, Grundfrequenzmerkmale, Jitter und Pausen. Signifikante Veränderungen der meisten Merkmale zeigten sich hier erst ab einer relativ hohen Alkoholisierung (über 1.6‰). Künzel and Braun [2003] nutzten die Daten ein weiteres Mal, um anhand der Bildbeschreibungen prosodische Parameter erneut zu untersuchen. Es wurde angenommen, dass sich die Grundfrequenz mit steigender Alkoholisierung nicht linear verhält, das heißt bei niedriger Alkoholisierung sinkt und bei erhöhten Werten wieder ansteigt. Auch hier wurde eine Verlangsamung der Sprechgeschwindigkeit beobachtet.

Außerhalb der Forensik untersuchten Sobell et al. [1982] die Parameter Grundfrequenz, Amplitude und Lesegeschwindigkeit anhand der gelesenen Sprache von 16 männlichen Versuchspersonen unter Alkoholeinfluss zwischen 18 und 22 Jahren, jedoch lieferte die Analyse keine signifikanten Ergebnisse. Klingholz et al. [1988] analysierten gelesenes Sprachmaterial von 16 männlichen Sprechern (25-35 Jahre) unter Alkoholeinfluss hinsichtlich der Parameter Grundfrequenz, Signalrauschabstand, Formantlage und Langzeitspektrum. Die mittlere Grundfrequenz stieg nach der Einnahme von Alkohol signifikant an, die Amplitude, abgeleitet aus dem Spektrum, sank. Anhand von gelesenen Spondeen (Wörter mit zwei aufeinanderfolgenden, betonten Silben) untersuchten Behne and Rivera [1990] folgende Merkmale der Sprache 6 männlicher alkoholisierter Sprecher: Grundfrequenz, Amplitude, Formantlagen F1-F3, sowie Rhythmusmerkmale zur Vokaldauer und Intervokaldauer, jeweils auch im Verhältnis zur Wortdauer. Hier wurde von steigender Grundfrequenz der Vokale berichtet, niedrigeren Formantlagen und steigender Intensität. Cooney et al. [1998] schließen aus den Ergebnissen ihrer Analyse von Grundfrequenz, Formantfrequenzen und Sprechdauer in gelesener Sprache, dass diese sich nicht für die Detektion von Alkoholisierung eignet, da keine signifikanten Unterschiede gefunden wurden. Von steigender Grundfrequenz und sinkender Sprechgeschwindigkeit wurde auch in den Studien von Hollien et al. [1999] und Hollien et al. [2001], die Sprache von 19 männlichen und 16 weiblichen Probanden

in nüchternem Zustand und unter Alkoholeinfluss umfassen, berichtet.
Neben den Studien, die sich vorwiegend mit akustischen Eigenschaften von unter Alkoholeinfluss geäußerter Sprache beschäftigen, wurden auch Perzeptionsexperimente durchgeführt, deren Ziel es war herauszufinden, ob sich Sprache unter Alkoholeinfluss von in nüchternem Zustand geäußerter Sprache hörbar unterscheiden lässt. In der Studie von Martin and Yuchtman [1986] (8 männliche Sprecher, BAK 1‰ - 1.9‰, 21 Hörer) wurde im einfachen Identifikationstest eine Erkennungsrate von nur 61.5% erreicht. Beim paarweisen Diskriminationstest lag die Diskriminationsrate mit 73.8% deutlich höher. Die 12 Hörer aus der Studie von Klingholz et al. [1988] (16 männliche Sprecher, 11 davon in alkoholisiertem Zustand, BAK 0.5‰ - 1.5‰) erreichten Werte von 54.2% (einfache Identifikation) und 61.1% (paarweise Identifikation). Über ein signifikantes Ansteigen der Identifikationsrate auf 61% ab einer Blutalkoholkonzentration von 0.8‰ berichten Künzel et al. [1992] in ihrer Studie (Identifikationstest) mit 33 männlichen Sprechern und 30 Hörern. Die 79 Hörer der Studie von Ultes et al. [2011] (77 männliche und 77 weibliche Sprecher, BAK 0.28‰ - 1.75‰) erreichten eine Erkennungsrate von 55.8% (Identifikationstest) im Rahmen der Interspeech Speaker State Challenge (Schuller et al. [2011]).
Zusammenfassend sind in den verschiedenen Studien zu den akustischen Eigenschaften von Sprache unter Alkoholeinfluss zwei der untersuchten phonetischen Parameter am aussagekräftigsten. Die Sprechgeschwindigkeit (da es sich größtenteils um gelesenes Material handelt, auch Lesegeschwindigkeit), die im Falle einer Alkoholisierung als generell sinkend bezeichnet wird (Behne et al. [1991], Hollien et al. [1999, 2001], Künzel and Braun [2003], Sobell et al. [1982]), und die Grundfrequenz. Die Ergebnisse zur Grundfrequenz sind dabei nicht einheitlich. Es wird von einem nicht signifikanten Absinken der durchschnittlichen Grundfrequenz bei Sprache unter Alkoholeinfluss (Aldermann et al. [1995], Watanabe et al. [1994]), einem signifikanten Anstieg (Behne and Rivera [1990], Cummings et al. [1995], Hollien et al. [1999, 2001], Klingholz et al. [1988]), aber auch von der Alkohol-

menge abhängigem Absinken bzw. Ansteigen (Künzel and Braun [2003], Künzel et al. [1992]) berichtet. Der Bereich, in dem sich die Grundfrequenz bewegt, ist laut nahezu allen Studien in alkoholisiertem Zustand der Sprecher größer. Beide Parameter verändern sich aber auch unter Stress (z.B. Hansen and Patil [2007]), bei emotionalen Zuständen wie Freude, Wut oder Traurigkeit (z.B. Mathon and de Abreu [2007], Yildirim et al. [2004]) und beim Lombard-Effekt (z.B. Folk and Schiel [2011]). Die Untersuchung weiterer Parameter gesprochener Sprache unter Einfluss von Alkohol ist nicht nur aus phonetischer Sicht, sondern beispielsweise auch für die Forensik von Interesse.

In diesem Zusammenhang ist es wichtig, dass für die untersuchten Parameter Referenzdaten zum Grad der Alkoholisierung vorliegen, damit auf wissenschaftlicher Basis statistisch valide Aussagen möglich sind. Bis auf Klingholz et al. [1988], Watanabe et al. [1994] und Levit et al. [2001], die in den genannten Studien eine Messung der Blutalkoholkonzentration vornahmen, wurde bisher nur die Atemalkoholkonzentration (AAK bzw. Englisch BrAC) gemessen. Da sich aber laut Schiel et al. [2012] eine Wahrscheinlichkeit von 29% ergibt, dass die gemessene Atemalkoholkonzentration und die Blutalkoholkonzentration derselben Person um 0.0001 voneinander abweichen (Pearson Korrelation zwischen Atemalkoholkonzentration und Blutalkoholkonzentration bei 152 Sprechern $r = 0.89$), wurde für die Erstellung des Sprachkorpus, der der vorliegenden Arbeit zu Grunde liegt, eine Messung der Blutalkoholkonzentration vorgenommen.

Möglicherweise widersprechen sich die Ergebnisse der aufgeführten Studien teilweise deshalb, weil eine zu geringe Sprecherzahl oder zu wenig Sprachmaterial untersucht wurde. Oftmals wurden auch Aussagen auf Basis von eher unausgewogenem bzw. einseitigem Sprachmaterial getroffen. Weiterhin zu kritisieren wäre, dass in keiner Studie bis auf Hollien et al. [1999] und Hollien et al. [2001] weibliche Versuchspersonen vorhanden waren. Die Analysen, die im Folgenden vorgestellt werden, basieren auf divergentem Sprachmaterial (deutsch) von 162 Sprecherinnen und Sprechern und ermöglichen damit statistisch valide Aussagen.

1.4 Inhalt dieser Arbeit

Diese Arbeit berichtet über Grundlagenforschung hinsichtlich der Fragestellung, wie paralinguistische Informationen, in diesem Fall der Zustand der Alkoholisierung, über das Sprachsignal transportiert werden. Insbesondere wird auf die rhythmische Struktur von Sprache unter Alkoholeinfluss anhand verschiedener Analyseansätze eingegangen. Applikationsspezifische Untersuchungen sind nicht Gegenstand dieser Arbeit. Ein Klassifikator zur Detektion von Alkoholisierung anhand der Bewertung eines Merkmals oder der Kombination mehrerer Merkmale, die sich auf Basis des Sprachsignals ermitteln lassen, wurde nicht entworfen. Auch war nicht das Ziel dieser Arbeit, die Blutalkoholkonzentration mit Hilfe eines Merkmals oder der Kombination mehrerer Merkmale des Sprachsignals vorherzusagen. Vielmehr wurden die Sprachsignale nüchterner und alkoholisierter Sprecher dahingehend untersucht, ob Veränderungen einzelner automatisch aus dem Sprachsignal extrahierbarer Parameter in statistisch nachweisbarem Zusammenhang mit der Alkoholisierung der Sprecher stehen. Die Ergebnisse sind vor allem für die forensische Phonetik, aber auch für die Sicherheitstechnik (z.B. im Straßenverkehr) von Interesse. In diesen Bereichen ist es von Bedeutung zu wissen, wie sich ein Parameter bei Veränderung der Sprache durch äußere Einflüsse genau verhält. Des Weiteren wurde untersucht, ob und inwieweit das Geschlecht des Sprechers und der Sprechstil einen signifikanten Einfluss auf die Zusammenhänge zwischen der Veränderung eines Parameters und der Alkoholisierung haben. Außerdem wurden Grundfrequenz- sowie Energiekonturen von Sprachsignalen bezüglich möglicher Gemeinsamkeiten in ihrer globalen Formgebung bei Sprache unter Alkoholeinfluss und Sprache ohne Alkoholeinfluss begutachtet. Die Experimente wurden auf Basis von ausschließlich deutschem Sprachmaterial durchgeführt. Damit sind die Ergebnisse nicht als sprachunabhängig zu bewerten.
In den bisherigen Studien zu Sprache unter Alkoholeinfluss wurden bis auf die Sprech- bzw. Lesegeschwindigkeit nur in Behne and Rivera [1990] Parameter zur

rhythmischen Struktur der Sprache untersucht. Für die Vokaldauer, die Intervokaldauer und die Verhältnisse von Vokaldauer zu Wortdauer bzw. Intervokaldauer zu Wortdauer ergaben sich dabei keine signifikanten Veränderungen. Es wurden jedoch nur Spondeen von 6 männlichen Sprechern analysiert. Zu erwarten wären aber vor allem bei Spontansprache durch die Einschränkungen, die Alkohol hervorrufen kann, Defizite in Planung und Steuerung des Sprachprozesses, welche zu Veränderungen der rhythmischen Struktur, z.B. durch Pausen, Längungen oder Reduzierungen, führen. Die im Folgenden vorgestellten Untersuchungen widmen sich diesem Sachverhalt insofern, als sie verschiedene, aus dem Sprachsignal automatisch extrahierbare Parameter, die jeweils zu einem gewissen Grad die rhythmische Struktur von Sprache widerspiegeln, begutachten. Dazu gehören Parameter, die anhand der Dauern und Dauerverhältnisse von Segmenten den zeitlichen Aspekt von Sprechrhythmus zu einem Teil beleuchten (Kapitel 3), die Analyse der Energiefunktion des Sprachsignals anhand von Parametern zur Dynamik (Kapitel 4), Parameter zur Sprechgeschwindigkeit (Kapitel 5) und die Analyse kompletter Konturen mittels verschiedener Distanzwerte, der Parametrisierung der Konturen und funktionaler Hauptkomponentenanalyse (Kapitel 6). Auf relevante Studien wird dabei jeweils innerhalb der einzelnen Kapitel näher eingegangen. In Kapitel 7 werden die Ergebnisse nochmals zusammenfassend diskutiert.

Zunächst wird in Kapitel 2 die zugrundeliegende Datenbasis vorgestellt.

Kapitel 2

Datenbasis

Im Folgenden wird die Datenbasis beschrieben, die der vorliegenden Arbeit als Grundlage diente. Die Vorstellung des verwendeten Sprachkorpus beschränkt sich hier auf das Korpus-Design (Kapitel 2.1.1), die Aufnahmeprozedur (Kapitel 2.1.2) und die Datenaufbereitung und -verarbeitung (Kapitel 2.2). Eine detaillierte Beschreibung des Korpus findet sich in Schiel et al. [2012]. Ferner wird in Kapitel 2.3 kurz auf die Methoden der Datenauswertung, die im Rahmen dieser Arbeit zur Anwendung kamen, und in Kapitel 2.4 auf die Durchführung von sogenannten Kontrollgruppenversuchen eingegangen.

2.1 Alcohol Language Corpus

Als empirische Datenbasis für alle Auswertungen, die in dieser Arbeit beschrieben werden, diente der Alcohol Language Corpus (ALC) des Bayerischen Archivs für Sprachsignale (BAS) der Universität München. ALC ist der erste öffentlich verfügbare Sprachkorpus mit in alkoholisiertem und nüchternem Zustand geäußerter Sprache von 162 deutschen Sprecherinnen und Sprechern. Von den 162 Sprechern sind 77 weibliche und 85 männliche Sprecher/innen im Alter von 21 bis 64 Jahren (Durchschnitt 31 Jahre). Durchschnittlich wurden von jedem Sprecher und jeder Sprecherin in alkoholisiertem Zustand 4.68 Minuten Sprachmaterial sowie

9.26 Minuten in nüchternem Zustand aufgezeichnet und für jeden Sprecher der Kontrollgruppe durchschnittlich weitere 4.9 Minuten bei der Kontrollgruppenaufnahme (siehe Kapitel 2.4). Im Gegensatz zu den meisten Studien über Sprache unter Alkoholeinfluss, bei welchen nur eine Messung der Atemalkoholkonzentration vorgenommen wurde, wurde bei der Erstellung des ALC zusätzlich zur Atemalkoholkonzentration auch die Blutalkoholkonzentration gemessen. Die Blutalkoholkonzentrationen bewegen sich in einem Bereich von 0.23‰ bis 1.75‰.
In Deutschland liegt die Grenze der zulässigen Blutalkoholkonzentration beim Führen eines Kraftfahrzeugs bei 0.5‰. Deshalb wurde für die meisten im Rahmen dieser Arbeit durchgeführten Untersuchungen das Sprachmaterial der 150 Sprecher des ALC verwendet, deren Blutalkoholkonzentration im alkoholisierten Zustand über $0.4\overline{9}$‰ liegt.

2.1.1 Corpus Design

Der Alcohol Language Corpus beinhaltet drei verschiedene Sprechstile: gelesene Sprache, Spontansprache und sogenannte Command & Control Sprache (im Folgenden auch als Kommandosprache bezeichnet). In Tabelle 2.1 sind die verschiedenen Typen von Aufnahmeelementen sowie der zugehörige Sprechstil zusammen mit ihrer Anzahl in der jeweiligen Aufnahmesitzung aufgelistet. Die Kategorie Command & Control ist anteilig in gelesener Sprache und Spontansprache vertreten.

Enthaltene Zahlenfolgen sind Telefonnummern, Kreditkartennummern und Kfz-Kennzeichen. Die Zungenbrecher wurden mit aufgenommen, um die Hypothese, dass in alkoholisiertem Zustand mehr Artikulationsfehler auftreten, falsifizieren zu können. Dabei wurden eher unbekannte Zungenbrecher und sprachtherapeutische Sprechübungen verwendet, um auswendig aufgesagtes beziehungsweise allzu bekanntes Material zu vermeiden. Die verwendeten Adressen waren bis auf eine Ausnahme, bei der versucht wurde, interessante Lautkombinationen zu integrieren („Madapaka-Betegindis-Straße“), einer Geodatenbank entnommen.

Sprechstil	Aufnahmeelementtyp	alkoholisiert	nüchtern
gelesene Sprache	Zahlenfolge	5	10
	Zungenbrecher	5	10
	Adresse	5	10
	gelesenes Kommando	4	9
	Buchstabierung	1	1
Spontansprache	Bildbeschreibung	3	6
	thematische Frage	2	4
	spontanes Kommando	5	10
Gesamt		30	60

Tabelle 2.1: *ALC Sprachmaterial*

Sämtliche gelesenen Kommandos stammen aus einer sprachgesteuerten Applikation für Entertainment- und Navigationssysteme für Kraftfahrzeuge. Buchstabiert werden musste der Name einer deutschen Stadt. Im Allgemeinen wurde beim gelesenen Material darauf geachtet, dass Laute integriert wurden, die in der Literatur als anfällig gelten im Falle einer vorliegenden Alkoholisierung (siehe Schiel et al. [2012]).

Für die Bildbeschreibungen wurden Bilder aus einem psychologischen Test ausgewählt. Im alkoholischen Fall sind zwei der fünf spontansprachlichen Aufnahmen der Bildbeschreibungen und thematischen Fragen Dialoge, im nüchternen Fall fünf von zehn. Die restlichen Aufnahmen der Bildbeschreibungen und thematischen Fragen sind Monologe. Für eine nähere Beschreibung der Inhalte siehe Schiel et al. [2012].

2.1.2 Aufnahmeprozedur

In einem Zeitraum von ca. 3 Jahren (2007-2009) wurde Sprachmaterial von 162 Personen gesammelt. Dabei handelt es sich um Sprache alkoholisierter Sprecher sowie Sprache derselben Sprecher in nüchternem Zustand. Die Aufnahmen unter Alkoholeinfluss fanden in Zusammenarbeit mit dem Institut für Rechtsmedizin der Universität München statt, welches in regelmäßigen Abständen sogenannte

Trinktests durchführt, vorwiegend für Rechtsreferendare, Staatsanwälte, Polizeikräfte und medizinisches Personal. Ziel dieser Tests ist es, den Probanden eine Vorstellung dafür zu vermitteln, welcher Effekt Alkohol auf den menschlichen Körper hat. Darum werden die Probanden ersucht, in einem Zeitraum von ungefähr zwei Stunden eine bestimmte Menge Alkohol in Form von Bier oder Wein zu konsumieren. Die Menge richtet sich dabei nach Körpergröße, Gewicht, Alter und Geschlecht und wird auf Basis einer vom Probanden vorab angegebenen gewünschten Promillezahl errechnet. Dabei kamen die Formeln von Erik M. P. Widmark (Widmark [1932]) und P. E. Watson (Watson et al. [1980]) zum Einsatz, die in Kombination eine Berechnung der aufzunehmenden Alkoholmenge ermöglichen.

$$c = \frac{V}{m \cdot r} \tag{2.1}$$

Widmark Formel 2.1: c ist die Alkoholkonzentration im Blut pro kg Blut in Gramm, V die aufgenommene Masse des Alkohols in Gramm, m die Masse der Person in kg und r der Verteilungsfaktor im Körper.
Um den Verteilungsfaktor r zu ermitteln, wurde mit Hilfe der erweiterten Formeln von Watson und der Modifikation von Axel Eicker für Frauen zunächst der Gesamtkörperwasseranteil für Männer und Frauen berechnet:

$$GKW_{männlich} = 2,447 - 0,09516 \cdot \frac{t}{\text{Jahre}} + 0,1074 \cdot \frac{h}{\text{cm}} + 0,3362 \cdot \frac{m}{\text{kg}} \tag{2.2}$$

$$GKW_{weiblich} = 0.203 - 0,07 \cdot \frac{t}{\text{Jahre}} + 0.1069 \cdot \frac{h}{\text{cm}} + 0,2466 \cdot \frac{m}{\text{kg}} \tag{2.3}$$

Dabei ist t das Alter in Jahren und h die Körpergröße in cm. Wird der GKW mit der Dichte von Blut $\rho_{Blut} = 1,055\frac{g}{cm^3}$ und dem Anteil von Wasser im Blut $f = 0.8$ kombiniert, ergibt sich:

$$r = \frac{\rho_{Blut} \cdot GKW}{f \cdot m} \tag{2.4}$$

Durch Einsetzten der Gleichung 2.4 in Gleichung 2.1 kann die Menge des zu konsumierenden Alkohols in Gramm wie folgt berechnet werden (Schiel et al. [2012]):

$$V = \frac{c \cdot \rho_{Blut} \cdot GKW}{f} \tag{2.5}$$

Damit sich der im Mundraum befindliche Restalkohol verflüchtigt, muss nach dem Alkoholkonsum eine Mindestwartezeit von 20 Minuten eingehalten werden, erst dann ist eine Messung der Atemalkoholkonzentration aussagekräftig. Entsprechend kann auch eine Messung der Blutalkoholkonzentration erst nach diesem Zeitraum verlässliche Werte liefern. Beide Messungen wurden im Rahmen des Trinktests möglichst zeitnah nacheinander vorgenommen. Direkt im Anschluss wurde mit Teilnehmern des Trinktests die Sprachaufnahme unter Alkoholeinfluss durchgeführt, um eine messbare Veränderung im Grad der Alkoholisierung der Sprecher möglichst auszuschließen. Diese Sprachaufnahme dauerte maximal 15 Minuten. Nach einer mindestens zweiwöchigen Wartezeit fand daraufhin die Aufnahme in nüchternem Zustand statt. Auf Grund der doppelten Anzahl von Aufnahmeelementen im Vergleich zur Aufnahme unter Alkoholeinfluss dauerte diese Aufnahme ungefähr 30 Minuten. Von insgesamt 162 Sprechern wurden weiterhin 20 zufällig als Kontrollgruppensprecher ausgewählt (10 weiblich und 10 männlich) und wiederum mindestens 2 Wochen später erneut aufgenommen. Dabei wurden dieselben Aufnahmeelemente wie in der Aufnahme unter Alkoholeinfluss verwendet. Anhand der Kontrollgruppendaten kann untersucht werden, ob andere Faktoren außer der Alkoholisierung Einfluss auf das Sprachsignal haben bzw. ob gefundene Effekte wirklich auf die Alkoholisierung zurückzuführen sind.
Als Aufnahmeumgebung für alle Versuche dienten zwei Personenkraftwagen, die technisch ausgerüstet wurden und somit als mobiles Studio fungierten. Damit wurde auch sichergestellt, dass die akustischen Aufnahmebedingungen für alle Aufnahmen dieselben waren. Die Trinkversuche und somit auch die Aufnahmen unter Alkoholeinfluss fanden in vier verschiedenen bayerischen Städten statt: München, Landshut, Augsburg und Traunstein.

Die Sprachdaten wurden unter laborähnlichen Bedingungen erhoben, in vollem Bewusstsein der Sprecher, sich in einer Aufnahmesituation zu befinden. Dieser Sachverhalt trägt dazu bei, dass die Sprachdaten vor allem für die forensische Phonetik und die Sicherheitstechnik von Interesse sind, da auch in diesem Bereich vor allem Sprache von Sprechern mit dem Wissen, aufgenommen bzw. hinsichtlich ihrer Sprache bewertet zu werden, analysiert wird. Sprachaufnahmen, die nicht unter Beobachtung stattfinden bzw. bei welchen der Sprecher nicht weiß, dass er aufgenommen wird, wären zwar im Bezug auf das Verhalten des Sprechers und seine lautsprachlichen Äußerungen bei einer vorliegenden Alkoholisierung realistischer bzw. mehr am alltäglichen Leben orientiert, doch findet in den relevanten Arbeitsbereichen eine Bewertung der Sprache im Allgemeinen nicht unter denselben Bedingungen statt. In der Praxis weiß der Sprecher grundsätzlich, dass eine objektive oder subjektive Bewertung seiner Sprache bevorsteht, sei es bei der Spracherkennung einer sprachgesteuerten Applikation oder beim Tätigen eines Telefonanrufs. Somit ist die „Aufnahmesituation“ auch dann eher laborähnlich. Der ALC bietet mit den verschiedenen Sprechstilen ein großes Repertoire an Sprachdaten, die auf Grund der ebenfalls laborähnlichen Aufnahmebedingungen die nötigen Voraussetzungen erfüllen, um Untersuchungen zur Bewertung von praxisnaher Sprache unter Alkoholeinfluss durchführen zu können.

2.2 Datenaufbereitung und -verarbeitung

Alle Audiodateien wurden in Stereo mit einer Abtastrate von 44100 Hz und 16 Bit Samplingtiefe von zwei Mikrofonen aufgezeichnet, einem Headset Beyerdynamic Opus 54.16/3 und einem Grenzflächenmikrofon AKG Q400. Der jeweils linke Kanal der stereo Audiodateien enthält das Signal des Headset Mikrofons, der jeweils rechte Kanal das Signal des Grenzflächenmikrofons. Zur weiteren Verarbeitung wurden die stereo Audiosignale in zwei unabhängig voneinander verwertbare Monosignal-Dateien aufgeteilt.

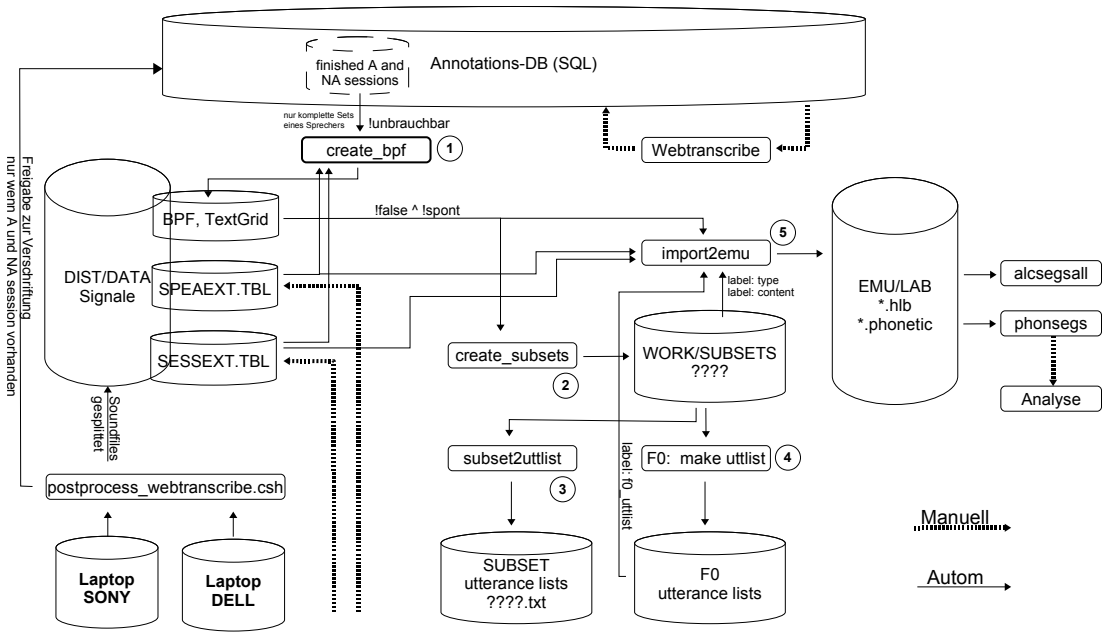


Abbildung 2.1: *Datenflussdiagramm zum ALC.*

Für die Analyse kann somit auf ein sehr direktes Nahbesprechungsmikrofonsignal (Headset), aber auch auf ein Raumakustik-erfassendes Mikrofonsignal (Grenzflächenmikrofon) zurückgegriffen werden. Jedoch basieren die in dieser Arbeit genannten Untersuchungen ausschließlich auf Signalen des Headset Mikrofons. Diese stellen eine kaum durch Nebengeräusche belastete und damit gut zu verwertende Grundlage für die durchgeführten akustischen Analysen dar.
Der gesamte Datenfluss den ALC betreffend ist aus dem Diagramm in Abbildung 2.1 ersichtlich. Im Folgenden sind nur die wichtigsten Schritte der Datenverarbeitung näher beschrieben.
Unter Zuhilfenahme des Online Annotationswerkzeugs **Webtranscribe** (Draxler [2005]) wurden auf Basis der Headset Audiosignale größtenteils orthographische Transkriptionen für jede Signaldatei durch drei phonetisch gebildete Verschrifter manuell angefertigt. Die dafür verwendeten Verschriftungskonventionen werden hier nicht näher beschrieben, können aber in Schiel et al. [2012] nachgeschlagen werden. Aus diesen phonetischen Transkriptionen erstellte **MAUS** (Münchner

AUtomatische Segmentierungen, Schiel [1999]) eine phonetische Segmentierung nach SAMPA (Speech Assessment Methods Phonetic Alphabet)[1], die als Grundlage für alle im Rahmen dieser Arbeit durchgeführten Analysen, die einer Segmentierung bedürfen, diente. Alle vorhandenen Informationen, sprecher- und aufnahmespezifische, sowie die einzelnen Segmentierungen, wurden in **EMU** (EMU Speech Database System, Cassidy and Harrington [2001]) Hierarchie Dateien abgelegt. Diese Daten können in der verwendeten Statistiksoftware (siehe Kapitel 2.3) eingelesen und weiterverarbeitet werden.

Die für eine Vernetzung der ALC Daten notwendige grundlegende Datenmatrix, die in der verwendeten Statistiksoftware verarbeitet werden kann, enthält Informationen zu jeder Äußerung bzw. jedem Aufnahmeelement sowie die Metadaten des jeweiligen Sprechers (Ausschnitt der grundlegenden Datenmatrix [*alcsegsall*] in Anhang D; Detailbeschreibung der Metadaten in Schiel et al. [2012]).

Das zweite grundlegende Konstrukt ist eine Segmentliste, welche die phonetische Segmentierung mit Beginn- und Endzeitpunkt des jeweiligen Lautes sowie den Dateinamen der Äußerung enthält. Auf Basis dieser Segmentliste wurden in einer weiteren Segmentliste alle zusammenhängenden vokalischen Elemente zu V Elementen und alle zusammenhängenden konsonantischen Elemente zu C Elementen zusammengefasst (siehe auch Kapitel 3.1). Beginn- und Endzeitpunkt der Elemente sowie der Dateiname der Äußerung sind ebenfalls vorhanden (Ausschnitte von beiden Segmentlisten [*phonsegs* und entsprechender Ausschnitt von *cvsegs*] in Anhang D). Diese drei Konstrukte zusammen mit den Audiodateien dienten als Grundlage für die hier vorgestellten und zahlreiche weitere Untersuchungen, die bisher unter Verwendung des ALC durchgeführt wurden.

EMU bietet durch das *ASSP* Modul verschiedene Werkzeuge zur Analyse von Audiosignalen an. Für die im Rahmen dieser Arbeit durchgeführten Untersuchungen sind jedoch nur die Werkzeuge *f0ana* und *rmsana* von Bedeutung. *f0ana* berech-

[1] phonetisches Alphabet, das auf der Zeichenkodierung American Standard Code for Information Interchange (ASCII) basiert.

net die F0-Werte (siehe Kapitel 6) zu einer oder mehreren Äußerungen, *rmsana* die RMS (root mean square) Werte (siehe Kapitel 4 und 6). Beide Werkzeuge und die dabei verwendeten Optionen werden in den jeweiligen Kapiteln näher beschrieben. Eine detailliertere Beschreibung des ALC findet sich in Schiel et al. [2012].

2.3 Datenauswertung

Die statistische Datenauswertung fand ausschließlich mit Hilfe der Statistiksoftware **R**[2] statt. Bei der Auswertung kamen verschiedene statistische Verfahren zum Einsatz, die *repeated measures multivariate analysis of variance* (RM-MANOVA = ANOVA mit Faktoren, die mehr als zwei Stufen aufweisen sowie 'repeated measures' Design), *Tukey HSD Post-hoc-Tests*, *Mixed Effects Models* (MEM) sowie die Berechnung von *Pearson Produkt-Moment-Korrelationen.*

Um die Bedingung der statistischen Unabhängigkeit einzuhalten, muss der Faktor Sprecher bei RM-MANOVA und MEM als 'random factor' modelliert werden. Die RM-MANOVA setzt voraus, dass die abhängige Variable ein Messwert ist, die Daten normalverteilt und einigermaßen homogen und die unabhängigen Variablen Faktoren sind. Zu jeder Versuchsperson und Faktorkombination muss genau ein Wert existieren, daher ist gegebenenfalls vorher eine Mittelung der Daten notwendig. Die RM-MANOVA ist ein etabliertes Verfahren und liefert leicht zu interpretierende p-Werte.

Bei signifikanten Interaktionen mit einem interessierenden Faktor (Beim ALC z.B. Alkoholisierung) wurden Tukey HSD Post-hoc-Tests durchgeführt, um festzustellen, ob eventuell nur eine Teilgruppe für den signifikanten Effekt verantwortlich ist. Besteht beispielsweise ein signifikanter Unterschied bezüglich der Alkoholisierung, aber auch eine signifikante Interaktion zwischen der Alkoholisierung und dem Sprechstil, dann kann es sein, dass nur ein Sprechstil signifikante Unter-

[2] http://www.r-project.org/, Version 2.13.0

schiede aufweist. Beim Tukey HSD Post-hoc-Test wird jede Kombination separat getestet.

MEMs haben gegenüber der ANOVA den Vorteil, dass die Daten nicht vollständig, homogen besetzt oder normalverteilt sein und vorher nicht gemittelt werden müssen. Faktoren können ordinal angeordnet sein und es sind mehrere 'random factors' möglich. Bei der MEM Analyse werden standardmäßig keine p-Werte, sondern nur F-Werte ausgegeben. Im Rahmen dieser Arbeit wurden alle p-Werte aus den F-Werten auf eher konservative Weise geschätzt. Dabei wurde nach Reubold et al. [2010] ein festgelegter, sehr niedriger Wert für die Freiheitsgrade ($Df = 60$) verwendet, weil innerhalb der F-Statistik ungefähr an dieser Stelle ein Plateau für die p-Werte erreicht wird, d.h. dass sich der F-Wert, ab welchem eine Signifikanz erreicht wird, für Freiheitsgrade über 60 nur noch geringfügig ändert. Beispielsweise sind die F-Werte bei $\alpha = 0.01$ $F[1, 60] = 8.49$ und $F[1, 600] = 7.94$, das entspricht einer Änderung von 0.55 im F-Wert gegenüber einer Änderung der Freiheitsgrade von 60 zu 600 (Reubold et al. [2010]).

Um den Grad des linearen Zusammenhangs verschiedener Merkmale zu bestimmen, wurden weiterhin in einigen Fällen Pearson Produkt-Moment-Korrelationen berechnet. Die Korrelationsberechnungen wurden immer auf Basis aller 162 Sprecher/innen des ALC durchgeführt (Sprecher, deren BAK $< 0.5‰$ ist, sind bei den Korrelationsberechnungen ebenfalls zu berücksichtigen).

2.4 Kontrollgruppenversuche

ALC bietet neben den Sprachdaten von 162 Personen in alkoholisiertem und nüchternem Zustand zusätzliches Sprachmaterial von 20 (10 weiblich, 10 männlich) der 162 Sprecher in ebenfalls nüchternem Zustand. Dieses Sprachmaterial wurde in einer dritten Sprachaufnahme erhoben, um auszuschließen, dass ein durch die Experimente gefundener Effekt zwischen Sprache unter Alkoholeinfluss und Sprache ohne Alkoholeinfluss auf einen verborgenen Faktor zurückzuführen ist.

Mit Hilfe dieser Kontrollgruppendaten können dieselben Experimente, die etwaige Unterschiede zwischen Sprache unter Alkoholeinfluss und in nüchternem Zustand geäußerter Sprache aufdecken, mit den Sprachdaten, die in nüchternem Zustand der Sprecher erhoben wurden und den Kontrollgruppendaten erneut durchgeführt werden (jeweils auf Basis der 20 Sprecher). Treten dort dieselben Effekte auf, muss davon ausgegangen werden, dass sie nicht auf die Alkoholisierung zurückzuführen sind. Bleiben die Effekte zwischen Kontrollgruppenaufnahmensprache (im Folgenden auch als *cna-Sprache* bezeichnet) und in nüchternem Zustand geäußerter Sprache aus und treten gleichzeitig zwischen Sprache unter Alkoholeinfluss und in nüchternem Zustand geäußerter Sprache auch auf Basis der 20 Sprecher auf, sind sie sehr wahrscheinlich durch die Alkoholisierung bedingt. Diese Kontrollgruppenversuche wurden, soweit möglich bzw. sinnvoll, für alle in dieser Arbeit vorgestellten Experimente durchgeführt und die jeweiligen Ergebnisse in den einzelnen Kapiteln aufgeführt. Sie basieren immer auf den Sprachdaten der 20 Sprecher, von welchen zusätzlich zu den in alkoholisiertem und nüchternem Zustand durchgeführten Sprachaufnahmen eine weitere in nüchternem Zustand durchgeführte Kontrollgruppenaufnahme existiert.

Kapitel 3

Rhythmusparameter

Für die erste hier vorgestellte Untersuchung der rhythmischen Struktur des ALC Sprachmaterials diente die phonetische Segmentierung als Grundlage. Die verfügbaren zeitlichen Informationen der einzelnen Phoneme ermöglichen verschiedene auch relationale Dauerberechnungen zwischen vokalischen und konsonantischen Elementen. Zahlreiche Studien beschäftigen sich mit Dauern und Dauerverhältnissen von vokalischen und konsonantischen Elementen. Vor allem wurde damit versucht, eine Einteilung von verschiedenen Sprachen in Rhythmusklassen zu erreichen. Bevor aber mit Hilfe von dauerbasierten Parametern Klassifizierungen vorgenommen wurden, prägten Pike [1946] und Abercrombie [1967] den Begriff der *rhythm class hypothesis*, die besagt, dass alle Sprachen in zwei verschiedene Rhythmusklassen eingeteilt werden können, silbenzählende und akzentzählende Sprachen. In silbenzählenden Sprachen sind aus ihrer Sicht aufeinanderfolgende Silben von annähernd gleicher Länge und damit isochron, in akzentzählenden Sprachen die Intervalle zwischen Betonungen. Das Konzept der Isochronie wurde in der Vergangenheit kontrovers diskutiert und vielfach abgelehnt, rein akustisch konnte Isochronie nicht nachgewiesen werden (siehe z.B. Bolinger [1965], Dauer [1983]). Neben anderen sieht auch Lehiste [1977] Isochronie eher als perzeptives Phänomen. Später wurde die Klassifizierung verschiedener Sprachen nach ihrem Rhythmus mit Hilfe von *Rhythmusparametern*, die akustisch phonetische Korrelate des Sprachrhythmus zur Klassenunterscheidung liefern sollten, oftmals

wiederholt und teilweise weiterentwickelt (z.B. Barry et al. [2003], Dellwo [2006], Dellwo et al. [2004], Loukina et al. [2009], Ramus et al. [1999], Wagner and Dellwo [2004], White and Mattys [2007], Wiget et al. [2010]). In einem kartesischen Koordinatensystem gegeneinander aufgetragene, für verschiedene Sprachen ermittelte Parameter, scheinen in vielen Fällen eine Gruppierung der Sprachen nach ihrer zugesprochenen Rhythmusklasse zu ermöglichen (siehe z.B. Dellwo [2006], Grabe and Low [2002], Ramus et al. [1999], Wiget et al. [2010]), doch gibt es auch Untersuchungen, die einer Trennbarkeit durch die propagierten Parameter und der Konzeptualisierung von Rhythmus, auf der sie beruhen, entgegensprechen (z.B. Arvaniti [2009]). Eine dritte Rhythmusklasse wurde von Bloch [1950], Han [1962] und Ladefoged [1975] benannt, die der morenzählenden Sprachen, bei der aufeinanderfolgende Moren (eine More entspricht einer offenen Silbe mit kurzem Vokal oder Kurzvokal und höchstens einem nachfolgenden Konsonanten) in ihrer Dauer ungefähr gleich lang sein sollen. Die verschiedenen Rhythmusparameter wurden wie bereits angeführt dahingehend entwickelt, um akustisch phonetische Korrelate des Sprachrhythmus zur Rhythmusklassenunterscheidung zu liefern, jedoch ist durch sie eine Klassifikation von Sprachen nicht unbedingt von Erfolg gekrönt. Vielmehr scheinen viele Sprachen hinsichtlich der Rhythmusparameter sowohl einer, als auch der anderen Rhythmusklasse jeweils zu einem gewissen Grad zugehörig (Arvaniti [2009]).

Unabhängig von ihren bisherigen Anwendungsbereichen basieren die Rhythmusparameter auf den Dauern und Dauerverhältnissen von vokalischen und konsonantischen Elementen eines Sprachsignals. Genauso wie es möglich bzw. unmöglich ist, mit ihnen Sprachen in verschiedene Rhythmusklassen einzuteilen, kann mit ihrer Hilfe auch unter Alkoholeinfluss und in nüchternem Zustand geäußerte Sprache hinsichtlich vokalischer und konsonantischer Dauern untersucht werden. Als Hypothese hinsichtlich Sprache unter Alkoholeinfluss gilt:

Hypothese 1: Veränderungen in der rhythmischen Struktur der Sprache in Form von Unregelmäßigkeiten, die durch Alkoholisierung hervorgerufen werden, spiegeln sich auch in Veränderungen der Dauern und Dauerverhältnisse von phonetischen Einheiten wider.

Zunächst wird in Kapitel 3.1 die Methode zur Extraktion einzelner Rhythmusparameter anhand der phonetischen Segmentierung und der damit verfügbaren zeitlichen Informationen der einzelnen Phoneme beschrieben. Dann folgen in Kapitel 3.2 die Auswertungsergebnisse zu den Parametern für alle Sprecher, die eine Blutalkoholkonzentration $\geq 0.5‰$ aufweisen und in Kapitel 3.3 die Auswertungsergebnisse der Kontrollgruppenversuche. Die anschließende Diskussion (Kapitel 3.4) zu den Rhythmusparametern beendet das Kapitel.

3.1 Rhythmusparameter - Methode

Die genannten sowie zusätzliche in diesem Kapitel eingeführte Rhythmusparameter wurden dazu verwendet, den Sprechrhythmus der a-Sprache und na-Sprache des ALC zu beschreiben und etwaige Unterschiede zwischen beiden, zwischen den Sprechstilen und zwischen den Geschlechtern aufzudecken.
Um eine Berechnung der einzelnen Parameter für den gesamten Korpus zu bewerkstelligen, wurden in R einzelne Funktionen entwickelt, die mit Hilfe der vorhandenen automatisch erzeugten phonetischen Segmentierung und den damit verfügbaren zeitlichen Informationen der Phoneme über eine oder mehrere Äußerungen hinweg automatisch Werte für alle Parameter ermitteln. Hierbei ist eine Einteilung der durch die automatische Segmentierung vorgegebenen Phoneme in vokalische und konsonantische Elemente unabdingbar. Diese Zuweisung wurde durch ein **Perl**[1]-Skript erreicht, das vokalische Elemente auf V Elemente abbildet und konsonantische Elemente auf C Elemente (Einteilung nach Tabelle 3.1).

[1] http://www.perl.org/

Um eine Schätzung der Silbendauer zu erhalten, wurde vereinfacht angenommen, dass die Silbengrenzen mit den zeitlichen Mittelpunkten der C Elemente zusammenfallen, die jeweils einen Silbennukleus, und damit vereinfacht angenommen ein V Element, umschließen. Eine Silbe S entspricht dem Abstand jeweils benachbarter Silbengrenzen bzw. den zeitlichen Mittelpunkten der C Elemente. Diese eher triviale Definition des Konstrukts Silbe soll keinen Anspruch einer Erklärung zur Silbe liefern. Im Rahmen dieser Arbeit legte man sich aber darauf sowie auf die in Tabelle 3.1 ersichtliche Einteilung in C und V Elemente und deren o.g. Beteiligung im Kontext der Silbe fest. Die Silben S wurden innerhalb jeder Äußerung und damit einer Reihe von C und V Elementen bestimmt. Sie können dabei auch wort- und phrasenübergreifend sein. Eine manuelle Silbensegmentierung war auf Grund der großen Datenmenge nicht realisierbar. Da sich innerhalb der Äußerungen und damit auch zwischen zwei C Elementen neben den V Elementen Pausen befinden können, wurden alle Parameter, deren Berechnung die Silbendauer zu Grunde liegt, auf zwei Arten ermittelt. Zum Einen sind die Pausen in der Berechnung und damit auch indirekt im Wert des Parameters enthalten (die jeweiligen Parameter sind in den Ergebnistabellen mit dem Index P markiert), d.h. dass sich die Silbendauer bei Vorkommen einer Pause zwischen zwei C Elementen aus jeweils der Hälfte der Dauer der umschließenden C Elemente, der Dauer des Nukleus bzw. V Elements und zusätzlich aus der Dauer der beinhalteten Pause berechnet. Zum Anderen wurden jegliche Silbendauern, die 500 ms (Millisekunden) überschreiten, von der Berechnung ausgeschlossen. Hierbei wurde vereinfacht angenommen, dass die durchschnittliche Silbendauer bei ungefähr 200 ms liegt und demnach alle Elemente, deren Dauer einen Maximalwert übersteigt (im Rahmen dieser Arbeit festgelegt bei 500 ms), nicht mehr als Silben zu werten sind. Weiterhin wurden einige Parameter auf Basis der Silbenkernabstände SN berechnet. Die Silbenkerne entsprechen den zeitlichen Mittelpunkten der V Elemente, ein Silbenkernabstand SN dem Abstand zwischen zwei Silbenkernen bzw. den Mittelpunkten zweier benachbarter V Elemente.

V Elemente	aː,ɑː,eː,iː,oː,uː,ɛː,yː,øː,a,ɛ,i,o,u,e,ɔ,ɪ,ʊ,ə,ʏ,œ,ɐ aʊ,aɪ,ɔʏ,ɛɪ,ɛ̃ː,ɛ̃,ãː,ã,õː
C Elemente	p,b,t,d,k,g,ʔ,f,v,ð,s,z,ʃ,ʒ,ç,x,h,m,n,ŋ,l,r,ʊ,j

Tabelle 3.1: *Einteilung vokalischer und konsonantischer Laute in die Kategorien C und V nach IPA.*

Die genaue Lauteinteilung in C und V Elemente nach den Richtlinien der *International Phonetic Association* (IPA) zeigt Tabelle 3.1.

Im Zuge einer alternativen Lauteinteilung wurden der Approximant /j/, der Lateral /l/ und die Nasale /m,n,ŋ/ (in IPA) auf Grund ihrer hohen Sonorität den Vokalen zugesprochen. Diese alternative Lauteinteilung sowie die zugehörigen Auswertungsergebnisse wurden hier nicht aufgeführt, da die Ergebnisse zu beiden Ansätzen der Lauteinteilung nicht maßgeblich voneinander abweichen. Sie sind jedoch in Anhang A zu finden.

Letztendlich handelt es sich um eine Transformation, bei der eine Aneinanderreihung von Phonemen in eine Kette aus C und V Elementen überführt wird. Wichtig für die eigentliche Berechnung der Parameter sind vokalische und konsonantische Teile von lautsprachlichen Äußerungen. Die vokalischen und konsonantischen Elemente können dabei sowohl einzeln als auch in Gruppen vorkommen. Deshalb findet innerhalb des Perl-Skripts zusammen mit der Zuweisung zusätzlich eine Zusammenfassung jeweils mehrerer nebeneinanderliegender C beziehungsweise V Elemente zu Gruppen von C und V Elementen statt, die demnach auch lautübergreifend sein können. Damit werden vokalische Regionen sowie konsonantische Regionen innerhalb der Äußerung voneinander abgegrenzt. Diese Vereinfachung und Zusammenfassung der ursprünglichen Segmentierung macht eine effektive, schnelle und zuverlässige Berechnung der Parameter für eine so große Datenmenge überhaupt erst realisierbar. Folgende Parameter wurden für jeden Sprecher und jeden Sprechstil, sowohl für die Aufnahmen im alkoholisierten und nicht alkoholisierten Zustand, als auch gegebenenfalls für die vorhandenen Kontrollgruppenaufnahmen, separat berechnet.

In den Formeln werden die Abkürzungen *dur* beziehungsweise *durs* für Dauer und Dauern eingesetzt, *mean* für den Mittelwert und *sd* für die Standardabweichung.

3.1.1 Vokalischer Anteil der Äußerung $\%V$

Zur Berechnung des vokalischen Anteils einer Äußerung $\%V$ (Ramus et al. [1999]) wurde die zeitliche Summe aller V Elemente durch die Gesamtdauer der lautsprachlichen Äußerung dividiert. Diese wiederum errechnet sich aus der Dauersumme aller C und V Elemente. Stillephasen innerhalb der Äußerung werden demnach nicht mit einbezogen.

$$\%V = \frac{durs(V)}{durs(V) + durs(C)} \tag{3.1}$$

3.1.2 Standardabweichung der vokalischen Dauern ΔV

Die Standardabweichung der vokalischen Dauern ΔV (Ramus et al. [1999]) errechnet sich wie folgt aus den vokalischen Elementen V:

$$\Delta V = sd\big(durs(V)\big) \tag{3.2}$$

3.1.3 Standardabweichung der konsonantischen Dauern ΔC

Die Standardabweichung der konsonantischen Dauern ΔC (Ramus et al. [1999]) wird wie folgt aus den konsonantischen Elementen C berechnet:

$$\Delta C = sd\big(durs(C)\big) \tag{3.3}$$

3.1.4 Standardabweichung der Silbendauern ΔS

Die Standardabweichung der silbischen Dauern ΔS errechnet sich aus den Silben S wie folgt (in Analogie zu ΔV nach Ramus et al. [1999]):

$$\Delta S = sd\big(durs(S)\big) \tag{3.4}$$

3.1.5 Standardabweichung der Silbenkernabstände ΔSN

Die Standardabweichung der Silbenkernabstände ΔSN errechnet sich aus den Silbenkernabständen SN wie folgt (in Analogie zu ΔV nach Ramus et al. [1999]):

$$\Delta SN = sd\big(durs(SN)\big) \tag{3.5}$$

3.1.6 Variationskoeffizient von ΔV, $Varco\Delta V$

Der Variationskoeffizient von ΔV, $Varco\Delta V$ (Dellwo [2006]) entspricht der Standardabweichung der vokalischen Dauern dividiert durch die durchschnittliche Vokaldauer und multipliziert mit 100. Damit ist der Wert sprechgeschwindigkeitsnormalisiert.

$$Varco\Delta V = \frac{sd\big(durs(V)\big)}{mean\big(durs(V)\big)} \cdot 100 \tag{3.6}$$

3.1.7 Variationskoeffizient von ΔC, $Varco\Delta C$

Der Variationskoeffizient von ΔC, $Varco\Delta C$ (Dellwo [2006]) entspricht der Standardabweichung der konsonantischen Dauern dividiert durch die durchschnittliche Konsonantendauer und multipliziert mit 100. Damit ist der Wert sprechgeschwindigkeitsnormalisiert.

$$Varco\Delta C = \frac{sd\big(durs(C)\big)}{mean\big(durs(C)\big)} \cdot 100 \tag{3.7}$$

3.1.8 Variationskoeffizient von ΔS, $Varco\Delta S$

Der Variationskoeffizient von ΔS, $Varco\Delta S$ entspricht der Standardabweichung der silbischen Dauern dividiert durch die durchschnittliche Silbendauer und multipliziert mit 100 (in Analogie zu $Varco\Delta C$ nach Dellwo [2006]). Damit ist der Wert sprechgeschwindigkeitsnormalisiert.

$$Varco\Delta S = \frac{sd\big(durs(S)\big)}{mean\big(durs(S)\big)} \cdot 100 \tag{3.8}$$

3.1.9 *Raw Pairwise Variability Index* der Vokalcluster $rPVI_V$

Der *raw Pairwise Variability Index* der Vokalcluster $rPVI_V$ (Grabe and Low [2002]) ist der durchschnittliche Dauerunterschied aufeinanderfolgender Vokalcluster V.

$$rPVI_V = \left(\sum_{i=1}^{N-1} \left| dur_i(V) - dur_{i+1}(V) \right| \right) \cdot (N-1)^{-1} \tag{3.9}$$

3.1.10 *Raw Pairwise Variability Index* der Konsonantencluster $rPVI_C$

Der *raw Pairwise Variability Index* der Konsonantencluster $rPVI_C$ (Grabe and Low [2002]) ist der durchschnittliche Dauerunterschied aufeinanderfolgender Konsonantencluster C.

$$rPVI_C = \left(\sum_{i=1}^{N-1} \left| dur_i(C) - dur_{i+1}(C) \right| \right) \cdot (N-1)^{-1} \tag{3.10}$$

3.1.11 *Raw Pairwise Variability Index* der Silben $rPVI_S$

Der *raw Pairwise Variability Index* der Silben $rPVI_S$ ist der durchschnittliche Dauerunterschied aufeinanderfolgender Silben S (in Analogie zu $rPVI_C$ nach Grabe and Low [2002]).

$$rPVI_S = \left(\sum_{i=1}^{N-1} \left| dur_i(S) - dur_{i+1}(S) \right| \right) \cdot (N-1)^{-1} \tag{3.11}$$

3.1.12 *Raw Pairwise Variability Index* der Silbenkernabstände $rPVI_{SN}$

Der *raw Pairwise Variability Index* (Grabe and Low [2002]) der Silbenkernabstände $rPVI_{SN}$ ist der durchschnittliche Dauerunterschied aufeinanderfolgender

Silbenkernabstände SN (in Analogie zu $rPVI_C$ nach Grabe and Low [2002]).

$$rPVI_{SN} = \left(\sum_{i=1}^{N-1} \left| dur_i(SN) - dur_{i+1}(SN) \right| \right) \cdot (N-1)^{-1} \tag{3.12}$$

3.1.13 *Normalized Pairwise Variability Index* $nPVI$

Für die Vokalcluster V, die Konsonantencluster C, die Silben S und die Silbenkernabstände SN wurde zusätzlich jeweils auch der *normalized Pairwise Variability Index* berechnet ($nPVI_V$, $nPVI_C$, $nPVI_S$, $nPVI_{SN}$). Zusammenfassend wurden V, C, S und SN in folgender Beschreibung und Formel als Int für Intervall bezeichnet. Der *normalized Pairwise Variability Index* der Intervalle $nPVI_{Int}$ (Grabe and Low [2002]) ist der durchschnittliche Dauerunterschied aufeinanderfolgender Intervalle Int dividiert durch die Summe der Dauern derselben Intervalle, damit sprechgeschwindigkeitsnormalisiert.

$$nPVI_{Int} = 100 \cdot \left(\sum_{i=1}^{N-1} \left| \frac{dur_i(Int) - dur_{i+1}(Int)}{\big(dur_i(Int) + dur_{i+1}(Int)\big)/2} \right| \right) \cdot (N-1)^{-1} \tag{3.13}$$

3.1.14 *Yet Another Rhythm Determination* $YARD$

$YARD$ (Wagner and Dellwo [2004]) ist der durchschnittliche Dauerunterschied der standardisierten (z-normalisierten) Silbendauer *dur(S)*.

$$YARD = \sum_{i=1}^{N-1} \left| dur_i(S) - dur_{i+1}(S) \right| \cdot (N-1)^{-1} \tag{3.14}$$

Des Weiteren wurden für ALC die Sprechgeschwindigkeit und einige Parameter zu Pausen, die ebenfalls ein rhythmisches Element darstellen können, wie folgt berechnet.

3.1.15 Sprechgeschwindigkeit SR

Die hier angewandte Schätzung der Sprechgeschwindigkeit SR beruht auf der automatisch erzeugten phonetischen Segmentierung durch MAUS und diese wie-

derum auf dem manuell erstellten orthographischen Transkript des Sprachmaterials und dem Sprachsignal. Sie ist damit weder rein manuellen noch automatischen Ursprungs. Berechnet wurde die Anzahl der Silben pro Sekunde. Dabei wurden einmal die äußerungsinternen Pausen miteinbezogen, da diese im Kontext einer Äußerung sprechgeschwindigkeitsspezifische Informationen tragen und damit Auswirkungen auf die globale Sprechgeschwindigkeit haben können. Zum Anderen wurde die Silbenrate auch exklusive äußerungsinterner Pausen berechnet, so dass nur Teile der Äußerung berücksichtigt wurden, die Sprache enthalten, Stillephasen jedoch ausgeschlossen wurden.
Die Anzahl der Silben entspricht der Anzahl der in der Äußerung vorkommenden vokalischen Elemente V (Silbennuklei). Die Sprechgeschwindigkeit SR ist die Anzahl der vokalischen Elemente V geteilt durch die Gesamtdauer der Äußerung $dur(utt)$.

$$SR = \frac{N_V}{dur(utt)} \tag{3.15}$$

3.1.16 Pausenparameter

Pausen wurden willkürlich in zwei Klassen eingeteilt. Kurze Pausen entsprechen äußerungsinternen Stillephasen bis 1 Sekunde Länge, lange Pausen solchen größer 1 Sekunde. Aus allen kurzen sowie langen Pausen wurden jeweils Mittelwert und Standardabweichung bestimmt $\big(mean(P_{short})$, $sd(P_{short})$, $mean(P_{long})$, $sd(P_{long})\big)$. Des Weiteren wurden auch die Raten kurzer (SP) und langer Pausen pro Sekunde (LP) berechnet ($SP = \frac{N_{SP}}{dur(utt)}$ sowie $LP = \frac{N_{LP}}{dur(utt)}$).
Da bei der Berechnung der Pausenparameter für die Kontrollgruppe der 20 Sprecher lange Pausen nicht in ausreichendem Maße vorhanden sind, wurden bei den Kontrollgruppenversuchen Mittelwert und Standardabweichung aller Pausen berechnet und nur die Anzahl kurzer Pausen bestimmt (siehe Kapitel 3.3).

3.2 Rhythmusparameter - Ergebnisse

Die genannten Parameter wurden für die 150 Sprecher des ALC mit einer Blutalkoholkonzentration über $0.4\overline{9}‰$ (im alkoholisierten Zustand) und jeden Sprechstil (gelesene, spontane und Kommandosprache) der in alkoholisiertem und nüchternem Zustand durchgeführten Sprachaufnahmen berechnet. Äußerungsübergreifend wurde dabei jeweils ein Wert für jeden Sprechstil ermittelt, damit also für jeden Sprecher 6 Werte (3 Sprechstile $\times$ 2 Sprecherzustände). Die Auswertung erfolgte per RM-MANOVA mit dem jeweiligen Parameter als abhängiger Variable, den 'within factors' Alkoholisierung (Stufen: alkoholisiert, nüchtern) und Sprechstil (Stufen: gelesen, spontan, Kommando), dem 'between factor' Geschlecht (Stufen: weiblich, männlich) sowie dem 'random factor' Sprecher. Da sich größtenteils signifikante Interaktionen der Faktoren Alkoholisierung und Sprechstil ergeben, wurde gegebenenfalls ein Tukey HSD Post-hoc-Test durchgeführt, und nur die relevanten Sprechstile ermittelt. Tabelle 3.2 zeigt die einzelnen Parameter mit den prozentualen Anteilen der Sprecher, bei welchen sich der Parameter vom nüchternen zum alkoholisierten Zustand hin erhöht bzw. verringert, die Signifikanzniveaus, sowie die jeweils betroffenen Sprechstile.

Parameter	% $\Uparrow$ na zu a	% $\Downarrow$ na zu a	p-level	Sprechstil(e)
$\%V$	63.33	36.67	$p < 0.001$	r
ΔV	75.33	24.67	$p < 0.001$	r,s
ΔC	65.33	34.67	$p < 0.001$	r,s
ΔS	68	32	$p < 0.001$	r,s
ΔS_P	65.33	34.67	$p < 0.001$	r,s
ΔSN	66	34	$p < 0.001$	r
ΔSN_P	66	34	$p < 0.001$	r,s
$Varco\Delta V$	64.67	35.33	$p < 0.001$	r
$Varco\Delta C$	59.33	40.67	$p < 0.001$	r
$Varco\Delta S$	54	46	$p < 0.001$	r
$Varco\Delta S_P$	58.67	41.33	$p < 0.001$	r
$rPVI_V$	74	26	$p < 0.001$	r,s,c
$rPVI_C$	69.33	30.67	$p < 0.1$	s
$rPVI_S$	70	30	$p < 0.001$	r,s

Parameter	% ⇑ na zu a	% ⇓ na zu a	p-level	Sprechstil(e)
$rPVI_{S_P}$	70	30	$p < 0.001$	r,s
$rPVI_{SN}$	68.67	31.33	$p < 0.001$	s
$rPVI_{SN_P}$	68.67	31.33	$p < 0.001$	r,s
$nPVI_V$	61.33	38.67	$p < 0.001$	s,c
$nPVI_C$	60	40	$p < 0.1$	s
$nPVI_S$	62	38	$p < 0.001$	s
$nPVI_{S_P}$	66.67	33.33	$p < 0.001$	r
$nPVI_{SN}$	52	48	n.s.	-
$nPVI_{SN_P}$	57.33	42.67	$p < 0.01$	r
$YARD$	53.33	46.67	$p < 0.01$	r
$YARD_P$	48.67	51.33	$p < 0.1$	r
SR	25.33	74.67	$p < 0.001$	r,s
SR_P	25.33	74.67	$p < 0.001$	r,s
$mean(P_{short})$	52	48	n.s.	-
$sd(P_{short})$	54.67	45.33	n.s.	-
$mean(P_{long})$	53.33	46.67	n.s.	-
$sd(P_{long})$	46	54	n.s.	-
SP	56	44	$p < 0.001$	r
LP	58	42	n.s.	-

Tabelle 3.2: *Auswertungsergebnisse zu den Rhythmusparametern mit Prozentwerten der Sprecher, bei welchen sich der jeweilige Parameter bei a-Sprache gegenüber na-Sprache erhöht (% ⇑ na zu a) bzw. verringert (% ⇓ na zu a). Sprechstile r (gelesen), s (spontan) und c (Kommando). Ein P im Index des Parameters bedeutet Berechnung inklusive Pausen.*

Die Ergebnisse, ersichtlich aus Tabelle 3.2, zeigen nahezu für alle untersuchten Parameter signifikante Unterschiede zwischen a-Sprache und na-Sprache an, jedoch nicht immer gleichermaßen für alle Sprechstile. Vor allem gelesenes Material scheint gut geeignet, um mit Hilfe der Parameter a-Sprache gegen na-Sprache abzugrenzen. Bis auf $YARD_P$, SR, SR_P und $sd(P_{long})$ sind die Parameter im statistischen Mittel bei a-Sprache höher als bei na-Sprache. Für einen Großteil der Sprecher stellt sich damit zwar für a-Sprache gegenüber na-Sprache eine Erhöhung der Parameterwerte ein, jedoch verringern sich die Werte auch für viele Sprecher. Demnach verhalten sich die Parameter sprecherabhängig. Es zeigten sich keine geschlechtsspezifischen Effekte.

3.3 Rhythmusparameter - Kontrollgruppenversuche - Ergebnisse

Um auszuschließen, dass es sich bei den gefundenen Effekten um Einflüsse verborgener Faktoren handelt, wurden zusätzlich Versuche anhand der Kontrollgruppe (20 Sprecher) durchgeführt. Treten hierbei Unterschiede zwischen a- und na-Sprache, nicht aber zwischen cna- und na-Sprache auf, kann davon ausgegangen werden, dass die gefundenen Effekte tatsächlich nur in Zusammenhang mit Sprache unter Alkoholeinfluss auftreten. Die genannten Parameter wurden für jeden der 20 Kontrollgruppensprecher und jeden Sprechstil (gelesene, spontane und Kommandosprache) der in alkoholisiertem und nüchternem Zustand durchgeführten Sprachaufnahmen sowie der Kontrollgruppenaufnahmen berechnet. Äußerungsübergreifend wurde dabei jeweils ein Wert für jeden Sprechstil ermittelt, damit insgesamt 9 Werte pro Sprecher (3 Sprechstile $\times$ 3 Sprecherzustände). Die Auswertung erfolgte ebenfalls per RM-MANOVA mit dem jeweiligen Parameter als abhängiger Variable, den 'within factors' Alkoholisierung (Stufen: alkoholisiert, nüchtern) und Sprechstil (Stufen: gelesen, spontan, Kommando), dem 'between factor' Geschlecht (Stufen: weiblich, männlich) sowie dem 'random factor' Sprecher. Da sich auch hier signifikante Interaktionen der Faktoren Alkoholisierung und Sprechstil ergeben, wurde gegebenenfalls ein Tukey HSD Post-hoc-Test durchgeführt, und nur die relevanten Sprechstile ermittelt. Tabelle 3.3 enthält die Auswertungsergebnisse zu a- und na-Sprache, Tabelle 3.4 die Ergebnisse zu na-Sprache und cna-Sprache für die einzelnen Parameter, den jeweiligen Signifikanzniveaus, den betroffenen Sprechstilen und den Prozentwerten der Sprecher, bei welchen sich der jeweilige Parameter bei a-Sprache gegenüber na-Sprache erhöht bzw. verringert.

Parameter	% $\Uparrow$ na zu a	% $\Downarrow$ na zu a	p-level	Sprechstil(e)
$\%V$	60	40	n.s.	-
ΔV	80	20	$p < 0.001$	r
ΔC	80	20	$p < 0.1$	r

Parameter	% ⇑ na zu a	% ⇓ na zu a	p-level	Sprechstil(e)
ΔS	90	10	$p < 0.05$	r
ΔS_P	70	30	$p < 0.001$	r
ΔSN	80	20	$p < 0.05$	r
ΔSN_P	75	25	$p < 0.001$	r
$Varco\Delta V$	75	25	$p < 0.01$	r
$Varco\Delta C$	65	35	$p < 0.1$	r
$Varco\Delta S$	55	45	$p < 0.05$	r
$Varco\Delta S_P$	60	40	$p < 0.01$	r
$rPVI_V$	90	10	$p < 0.001$	r,s
$rPVI_C$	85	15	$p < 0.01$	r
$rPVI_S$	90	10	$p < 0.01$	r
$rPVI_{S_P}$	85	15	$p < 0.001$	r
$rPVI_{SN}$	80	20	$p < 0.1$	r,s
$rPVI_{SN_P}$	85	15	$p < 0.001$	r
$nPVI_V$	75	25	$p < 0.01$	s
$nPVI_C$	80	20	$p < 0.01$	r
$nPVI_S$	80	20	$p < 0.01$	s
$nPVI_{S_P}$	90	10	$p < 0.001$	r
$nPVI_{SN}$	70	30	n.s.	-
$nPVI_{SN_P}$	65	35	$p < 0.1$	r
$YARD$	55	45	n.s.	-
$YARD_P$	50	50	n.s.	-
SR	10	90	$p < 0.01$	r
SR_P	5	95	$p < 0.001$	r
$mean(P)$	70	30	$p < 0.01$	r
$sd(P)$	55	45	$p < 0.1$	r
SP	45	55	$p < 0.05$	r,s

Tabelle 3.3: *Auswertungsergebnisse zu den Rhythmusparametern Kontrollgruppe alkoholisiert-nüchtern mit Prozentwerten der Sprecher, bei welchen sich der jeweilige Parameter bei a-Sprache gegenüber na-Sprache erhöht (% ⇑ na zu a) bzw. verringert (% ⇓ na zu a). Sprechstile r (gelesen), s (spontan) und c (Kommando). Ein P im Index des Parameters bedeutet Berechnung inklusive Pausen.*

Parameter	% ⇑ na zu cna	% ⇓ na zu cna	p-level	Sprechstil(e)
$\%V$	45	55	n.s.	-
ΔV	60	40	n.s.	-
ΔC	40	60	n.s.	-
ΔS	45	55	n.s.	-
ΔS_P	50	50	n.s.	-

Parameter	% ⇑ na zu cna	% ⇓ na zu cna	p-level	Sprechstil(e)
ΔSN	55	45	n.s.	-
ΔSN_P	55	45	n.s.	-
$Varco\Delta V$	55	45	n.s.	-
$Varco\Delta C$	45	55	n.s.	-
$Varco\Delta S$	50	50	n.s.	-
$Varco\Delta S_P$	45	55	n.s.	-
$rPVI_V$	55	45	n.s.	-
$rPVI_C$	60	40	n.s.	-
$rPVI_S$	60	40	n.s.	-
$rPVI_{S_P}$	65	35	n.s.	-
$rPVI_{SN}$	60	40	n.s.	-
$rPVI_{SN_P}$	65	35	n.s.	-
$nPVI_V$	30	70	n.s.	-
$nPVI_C$	55	45	n.s.	-
$nPVI_S$	65	35	n.s.	-
$nPVI_{S_P}$	85	15	$p < 0.1$	s
$nPVI_{SN}$	60	40	n.s.	-
$nPVI_{SN_P}$	60	40	n.s.	-
$YARD$	50	50	n.s.	-
$YARD_P$	60	40	n.s.	-
SR	30	70	$p < 0.1$	s
SR_P	40	60	$p < 0.1$	c
$mean(P)$	50	50	n.s.	-
$sd(P)$	40	60	n.s.	-
SP	60	40	n.s.	-

Tabelle 3.4: *Auswertungsergebnisse zu den Rhythmusparametern Kontrollgruppe Kontrollgruppenaufnahmen-nüchtern mit Prozentwerten der Sprecher, bei welchen sich der jeweilige Parameter bei cna- gegenüber na-Sprache erhöht (% ⇑ na zu cna) bzw. verringert (% ⇓ na zu cna). Sprechstile r (gelesen), s (spontan) und c (Kommando). Ein P im Index des Parameters bedeutet Berechnung inklusive Pausen.*

Die Ergebnisse der Kontrollgruppe zeigen, dass sich die untersuchten Parameter zur Unterscheidung von a- und na-Sprache eignen, da sie nur zwischen a- und na-Sprache einen signifikanten Unterschied anzeigen (Tabelle 3.3). Zwischen der Aufnahme in nüchternem Zustand und der Kontrollgruppenaufnahme lässt sich anhand der Parameter statistisch kein Unterschied feststellen (Tabelle 3.4). Die Rhythmusparameter lassen also erkennen, dass die Segmentdauern bei a-Sprache

von den Segmentdauern bei na-Sprache abweichen. Dies hängt nicht nur von der Sprechgeschwindigkeit, sondern größtenteils vom Sprecherzustand ab, da auch die Parameter, deren Berechnung eine Normalisierung bezüglich der Sprechgeschwindigkeit beinhaltet, einen signifikanten Unterschied anzeigen (siehe Kapitel 3.4).

3.4 Rhythmusparameter - Diskussion

Die Rhythmusparameter, welche eine Einteilung von verschiedenen Sprachen in Rhythmusklassen ermöglichen sollen, scheinen bis zu einem gewissen Grad ebenfalls zur Unterscheidung von Sprache unter Alkoholeinfluss und in nüchternem Zustand geäußerter Sprache geeignet zu sein. Eine Erhöhung, welche bei einem Großteil der Parameter und Sprecher auftritt, bedeutet in der Regel mehr Variation, das heißt, dass unter Alkoholeinfluss geäußerte Sprache unregelmäßiger ist als Sprache ohne Alkoholeinfluss. Betrachtet man beispielsweise die Vokalcluster V, so deuten die Ergebnisse (z.B. ΔV) auf längere aber auch kürzere Dauern der Cluster bei a-Sprache als bei na-Sprache hin. Dies geschieht auch unabhängig von einer Veränderung der Sprechgeschwindigkeit. Geht man im Allgemeinen davon aus, dass die Sprechgeschwindigkeit bei a-Sprache niedriger ist als bei na-Sprache (siehe Parameter SR bzw. SR_P, Kapitel 1.3 und Kapitel 5), dann sind die Ergebnisse der Parameter, deren Berechnungen keine Normalisierung hinsichtlich der Sprechgeschwindigkeit beinhalten (ΔV, ΔC, ΔS, ΔS_P, ΔSN, ΔSN_P, $rPVI_V$, $rPVI_C$, $rPVI_S$, $rPVI_{S_P}$, $rPVI_{SN}$, $rPVI_{SN_P}$) weniger aussagekräftig als die Ergebnisse der Parameter, die durch Normalisierung unabhängig von der Sprechgeschwindigkeit bewerten ($\%V$, $Varco\Delta V$, $Varco\Delta C$, $Varco\Delta S$, $Varco\Delta S_P$, $nPVI_V$, $nPVI_C$, $nPVI_S$, $nPVI_{S_P}$, $nPVI_{SN}$, $nPVI_{SN_P}$, $YARD$, $YARD_P$; Pausen nehmen hierbei eine gesonderte Stellung ein, da sie weder als sprechgeschwindigkeitsabhängig noch -unabhängig bezeichnet werden können). Denn ohne Normalisierung ändern sich mit der Sprechgeschwindigkeit gleichzeitig die absoluten Dauern von V und C Elementen und damit die Wer-

te der betroffenen Parameter. Dies allein könnte bereits die Unterschiede zwischen a-Sprache und na-Sprache der Parameter, die ohne eine Normalisierung hinsichtlich der Sprechgeschwindigkeit berechnet wurden, hervorrufen. Findet jedoch eine Normalisierung statt und decken die entsprechenden Parameter trotz der Normalisierung bei einer Veränderung der Sprechgeschwindigkeit (und damit einer Abweichung in den absoluten Dauern von V und C Elementen) Unterschiede zwischen a- und na-Sprache auf, sind diese Unterschiede sehr wahrscheinlich auf die Alkoholisierung zurückzuführen und keinesfalls durch die Veränderung der Sprechgeschwindigkeit bedingt. Bei den Parametern, deren Berechnung eine Normalisierung beinhaltet, stellt sich zwar bei weniger Sprechern eine Erhöhung (na zu a) ein als bei den Parametern, die ohne Normalisierung berechnet werden, dennoch bleibt der Trend erkennbar.

Ein Anstieg für $\%V$ (gelesene Sprache) bei gleichzeitig sinkender Sprechgeschwindigkeit bedeutet, dass keine gleichmäßige oder lineare Dehnung des Sprachsignals stattfindet. Vielmehr scheinen die Dauern vokalischer Bereiche in größerem Maße zu steigen als die Dauern konsonantischer Bereiche. Die Zahlen bestätigen dies. Einer durchschnittlichen Dauerzunahme von 7 ms (na nach a) bei den Vokalclustern steht sogar eine durchschnittliche Dauerreduzierung um 1 ms (na nach a) bei den Konsonantenclustern gegenüber. Ein vermehrtes Auftreten von kurzen Pausen bei a-Sprache könnte auf Probleme in der Steuerung des Artikulationsprozesses hinweisen. Die Pausen könnten hierbei als zeitliche Puffer dienen, um die Defizite im Steuerungsprozess auszugleichen.

Inwieweit die untersuchten Parameter den Rhythmus von gesprochener Sprache widerspiegeln, soll nicht Untersuchungsgegenstand dieser Arbeit sein. Indiskutabel ist jedoch, dass sich die Dauern der Vokal- und Konsonantencluster im Falle einer vorliegenden Alkoholisierung verändern und diese Dauerveränderungen mit Hilfe der Parameter erfasst werden können. Damit wird auch Hypothese 1 statistisch bestätigt. Veränderungen in der rhythmischen Struktur der Sprache in Form von Unregelmäßigkeiten, die durch Alkoholisierung hervorgerufen werden,

spiegeln sich auch in Veränderungen der Dauern und Dauerverhältnisse von phonetischen Einheiten wider.

Analog zu den Rhythmusklassen bei Sprachen wurden auch im Falle des ALC die Werte für $\%V$ und ΔC, $\%V$ und ΔV, ΔV und ΔC (Ramus et al. [1999]), $\%V$ und $Varco\Delta C$ (Dellwo [2006]), $\%V$ und $Varco\Delta V$ (Wiget et al. [2010]) sowie $rPVI_C$ und $nPVI_V$ (Grabe and Low [2002]) in ein Koordinatensystem eingetragen (Abbildung 3.1), um zu untersuchen, ob sich die Daten hinsichtlich a-Sprache und na-Sprache äquivalent zu den Daten verschiedener Sprachen gruppieren lassen. Jedoch zeigte sich kein solches Bild. Die schwarzen Kreise (alkoholisiert) vermischen sich mit den grauen Dreiecken (nüchtern) in allen Fällen zu Punktwolken, die keinerlei Gruppierung erkennen lassen. Das heißt, dass mit Hilfe der verschiedenen Parameter keine eindeutig klassentrennende Gruppenbildung in die Daten von a- und na-Sprache möglich ist.

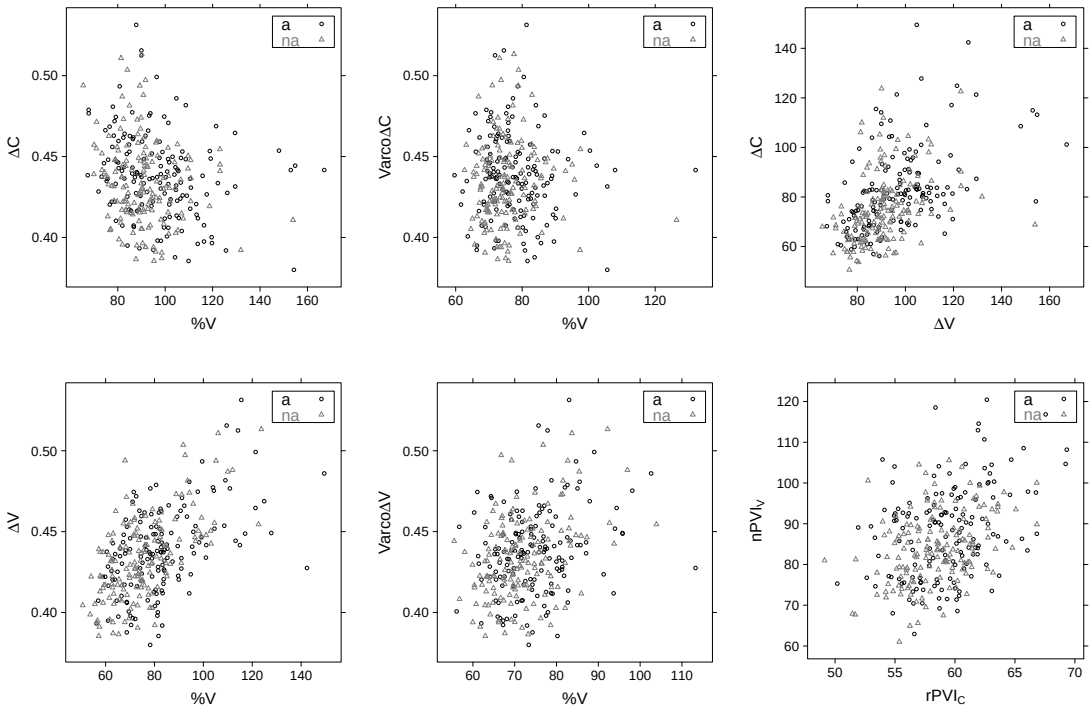


Abbildung 3.1: *$\%V$ und ΔC, $\%V$ und ΔV, $\%V$ und $Varco\Delta C$, $\%V$ und $Varco\Delta V$, ΔV und ΔC, $rPVI_C$ und $nPVI_V$. Schwarze Kreise a-Sprache, graue Dreiecke na-Sprache (alle Sprechstile, 150 Sprecher).*

Statistisch gesehen erhöhen sich praktisch alle Parameter bei Alkoholisierung der Sprecher, dennoch tritt auch bei einem großen Teil der Sprecher eine Verkleinerung der Werte auf. Es kann also keine allgemeingültige Aussage für alle Sprecher getroffen werden. Vielmehr verhalten sich die verschiedenen Sprecher individuell bzw. die Parameter sprecherabhängig. Dies ist auch aus Abbildung 3.2 ersichtlich, die beispielhaft für den Parameter $Varco\Delta V$ zeigt, bei wievielen Sprechern dieser sich von na-Sprache zu a-Sprache hin prozentual verringert bzw. erhöht.

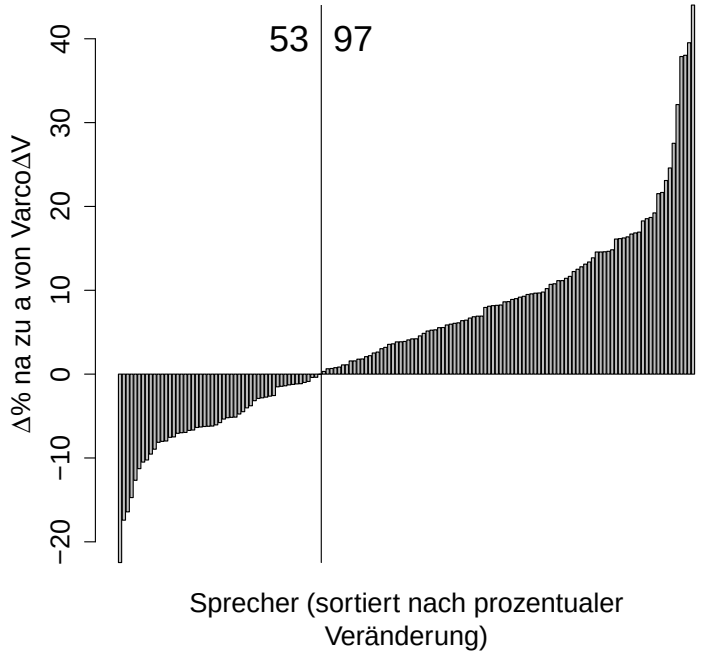


Abbildung 3.2: *Prozentuale Veränderung des Parameters $Varco\Delta V$ (na nach a) von 150 Sprechern (alle Sprechstile).*

Für die Änderungen aller Rhythmusparameter wurden Korrelationen mit dem Blutalkoholwert berechnet (162 Sprecher, alle Sprechstile zusammen). Die höchste aber dennoch geringe Korrelation zwischen der Änderung des Parameters und der BAK ergab sich für $rPVI_V$ (siehe Abbildung 3.3) mit $r = 0.28$ ($p < 0.001$). Im Allgemeinen lässt also der Grad der Veränderung der Rhythmusparameter keinen Rückschluss auf den Grad der Alkoholisierung

zu. Alkohol wirkt damit im Falle der Rhythmusparameter höchstwahrscheinlich individuell unterschiedlich. Korrelationen innerhalb der Daten eines Sprechers sind nicht möglich, da pro Sprecher und Aufnahme nur eine Messung der Blutalkoholkonzentration vorgenommen wurde.

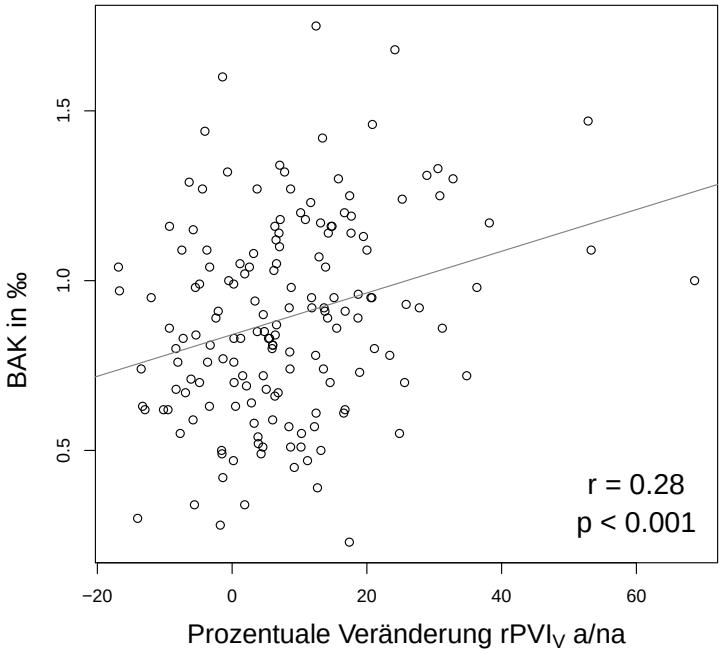


Abbildung 3.3: *Korrelation des Parameters $rPVI_V$ mit der BAK (162 Sprechern, alle Sprechstile).*

Die Berechnung der Parameter erfolgte auf Basis der automatischen Segmentierung. Auch wenn diese durch die vorgegebene orthographische Verschriftung idealerweise keine Fehler enthalten sollte, ist dies eher unwahrscheinlich. Des Weiteren kann nicht ausgeschlossen werden, dass die zeitlichen Informationen der einzelnen Phoneme nicht immer korrekt sind (Nach Kipp et al. [1997] erreicht MAUS eine Genauigkeit der Label von ca. 97% des Inter-Labeler Agreements für deutsche Dialogsprache und eine Genauigkeit der Segmentgrenzen [Abweichungen $<$ 20 ms] von ca. 90% des Inter-Labeler Agreements). Im statistischen Mittel sollten diese Fehler jedoch ausgeglichen werden und die Ergebnisse zu den berechneten Parametern aussagekräftig sein.

Die vorgestellte Methode ist nur bedingt vollautomatisch anwendbar. Nur für Kommandos, die in den meisten Fällen bzw. immer im selben Wortlaut verwendet werden, ist eine automatische Segmentierung theoretisch unüberwacht möglich.

Im folgenden Kapitel wird ein weiteres, diesmal voll automatisch arbeitendes Verfahren vorgestellt, welches Unterschiede in der rhythmischen Struktur zwischen Sprache unter Alkoholeinfluss und in nüchternem Zustand geäußerter Sprache anhand der Dynamik der Energiefunktion eines Sprachsignals aufdecken soll.

Kapitel 4

RMS Rhythmizitäts-Parameter

Die in Kapitel 3 vorgestellten Rhythmusparameter basieren auf zeitlichen Intervallen und beschreiben das Sprachsignal bzw. eine Gruppe von Sprachsignalen letztendlich mit jeweils nur einem Parameter. Dabei findet das zugrundeliegende Sprachsignal keine bzw. durch die automatische Segmentierung nur indirekte Beachtung. Weiterhin kann sich die rhythmische Struktur von Sprache nicht allein durch Dauermessungen und -berechnungen erfassen lassen. Ein zweites Experiment nimmt deshalb den *root mean square* (RMS), welcher die Wurzel aus dem arithmetischen Mittelwert der Summe der quadrierten Signalwerte ist, als Grundlage. Die Kurzzeit RMS- oder Energiekontur eines Sprachsignals beschreibt die Dynamik des Schalldrucks, der vereinfacht als Abfolge von leisen und lauten Abschnitten des Signals interpretiert werden kann. Die Energiekontur ist aber nicht nur eine geglättete Schätzung des Schalldruckpegels, sie kann auch zur Untersuchung von Sprechrhythmus im Allgemeinen (Tilsen and Johnson [2008]) und rhythmischen Parametern wie der Sprechgeschwindigkeit (Dekens et al. [2007], Morgan and Fosler-Lussier [1998], Pfau and Ruske [1998]), der Silbenposition (Xie and Niyogi [2006]) oder den im Folgenden vorgestellten *RMS Rhythmizitäts-Parametern* zur Unterscheidung von rhythmisch divergentem Sprachmaterial herangezogen werden. In den meisten der hier aufgeführten Studien wird das RMS-Signal zunächst durch Filterung oder ähnliche Verfahren vorverarbeitet und daraufhin durch eine regelbasierte Zählung der Maxima im Energieverlauf eine Schät-

zung der Silbenposition und damit der Silbenanzahl erreicht. Dadurch wird ein prosodisches Grundelement auf Basis des Sprachsignals markiert. Die Energiefunktion bietet somit die Möglichkeit, Untersuchungen zur rhythmischen Struktur von Sprache mittels automatischer Signalverarbeitung bei großen Datenmengen durchzuführen. Da sich schon durch die Rhythmusparameter gezeigt hat, dass Sprache unter Alkoholeinfluss gegenüber in nüchternem Zustand geäußerter Sprache rhythmische Unregelmäßigkeiten in Form von Daueunterschieden aufweist, die sich beispielsweise in einer Daueränderung von vokalischen Elementen oder einer Verlangsamung der Sprechgeschwindigkeit zeigen, sollten diese Unregelmäßigkeiten auch in der Energiefunktion eines Sprachsignals wiederzufinden sein, weil Energieminima und -maxima zu einem gewissen Teil die prosodische Struktur von Sprache widerspiegeln (siehe auch Kapitel 4.1). Damit lässt sich als Hypothese formulieren:

Hypothese 2: Veränderungen in der rhythmischen Struktur der Sprache in Form von Unregelmäßigkeiten und Lautstärkeschwankungen, die durch Alkoholisierung hervorgerufen werden, spiegeln sich auch in Veränderungen in der Dynamik der RMS- oder Energiefunktion des Sprachsignals wider.

Im nächsten Abschnitt (Kapitel 4.1) wird die Methode zur Extraktion der RMS Rhythmizitäts-Parameter beschrieben. Danach werden in Kapitel 4.2 die Auswertungsergebnisse zu den Parametern für die 150 Sprecher mit einer Blutalkoholkonzentration $\geq 0.5‰$ und in Kapitel 4.3 die Auswertungsergebnisse der Kontrollgruppenversuche präsentiert. Abschließend folgt die Diskussion zu den RMS Rhythmizitäts-Parametern (Kapitel 4.4).

4.1 RMS Rhythmizitäts-Parameter - Methode

Es wurde ein Verfahren entwickelt, das auf Basis der Kurzzeit Energiefunktion eines Sprachsignals Parameter zur Beschreibung der Rhythmizität extrahiert. Zunächst wurde im EMU Speech Database System mit Hilfe des im *ASSP* Modul integrierten Algorithmus *rmsana* eine RMS-Analyse für jedes Sprachsignal separat durchgeführt. Dabei berechnet ein gleitendes Fenster jeweils für ein kurzes Signalstück RMS-Werte. Der Algorithmus erlaubt die Festlegung verschiedener Fenstertypen, Fensterlängen und Verschiebungsdauern. Im Rahmen dieser Arbeit legte man sich auf die Analyse anhand eines *Blackman* Fensters fest und testete verschiedene Fensterlängen (50 ms, 100 ms und 200 ms) mit einer Fensterverschiebung von jeweils 20 ms. Je größer die gewählte Fensterlänge bei der Analyse, desto geglätteter ist die resultierende Kurve aus RMS-Werten. Wird eine kleinere Fensterlänge gewählt, wird die gewonnene Kurve detaillierter, d.h. feinere Fluktuationen im Energiesignal bleiben erhalten. Gleichzeitig wird die Kurve damit unruhiger. Von den getesteten Fensterlängen liefert das Fenster mit 100 ms den besten Kompromiss aus einer nicht zu unruhigen aber auch nicht zu geglätteten Kurve, die silbentypische Strukturen wie im Falle von reduzierten Silben beibehält und mit einer maximalen Anzahl von 10 Silben pro Sekunde auch herkömmliche Silbenraten erfasst. Sehr feine Bewegungen in der Energie, welche beispielsweise durch Verschlusslösungen bei Plosiven hervorgerufen werden, werden geglättet. Die Ergebnisse in Kapitel 4.2 basieren deshalb auf Berechnungen anhand des 100 ms Fensters.

Die so gewonnenen, logarithmisch skalierten RMS-Kurven wurden normalisiert, indem der durchschnittliche RMS-Wert jeder Kurve von der ursprünglichen Kurve subtrahiert wurde. Damit ergibt sich eine Funktion, die sich um 0 dB bewegt. Anhand dieser normalisierten Kurve wurden daraufhin lokale RMS-Minima und -Maxima ermittelt, die entweder unterhalb (Minima) oder oberhalb (Maxima) des RMS-Mittelwerts (0 dB nach Normalisierung) der jeweiligen Äußerung lie-

gen. Dadurch ergibt sich eine Sequenz aus einzelnen Minima und Maxima sowie Gruppen von Minima und Maxima. Ein Maximum in der Sonorität wird nach dem Silbenkonzept im Silbennukleus und Minima in der Sonorität an den Silbengrenzen erreicht (Kohler [1977], Wang and Narayanan [2007]). Weiterhin lies sich anhand psychoakustischer Experimente zeigen, dass Silbennuklei "lauter" perzipiert werden als ihre benachbarten Konsonanten (Pfau and Ruske [1998]). Dementsprechend repräsentiert ein Maximum in der Energiekurve im Groben einen Silbennukleus (lauter Abschnitt im Signal) und die angrenzenden Minima die Silbengrenzen (leise Abschnitte im Signal). Da eine Gruppe von Minima oder Maxima höchstwahrscheinlich nur je einer Silbengrenze beziehungsweise einem Silbenkern zugeordnet werden kann, wird aus einer solchen Gruppe repräsentativ jeweils ein Minimum bzw. Maximum ausgewählt. Das Minimum mit dem kleinsten RMS-Wert aus einer Gruppe von Minima sowie das Maximum mit dem größten RMS-Wert aus einer Gruppe von Maxima stellen jeweils den Scheitelpunkt der Gruppe dar und sind daher die logische Wahl. Das Ergebnis ist eine *min-max-min-max* Sequenz, die annähernd die Silbenabfolge einer Äußerung widerspiegelt.

Zu sehen ist eine solche Sequenz in der folgenden Grafik 4.1. Weiterhin dient diese Sequenz und die sich ergebende Kurve als Grundlage zur Ermittlung verschiedener Parameter zur Beschreibung der Rhythmizität eines Sprachsignals. Eine grafische Darstellung der Parameter befindet sich neben der *min-max* Kurve (schwarz) sowie der Kurve aus den ursprünglichen RMS-Analysewerten (grau) ebenfalls in Abbildung 4.1.

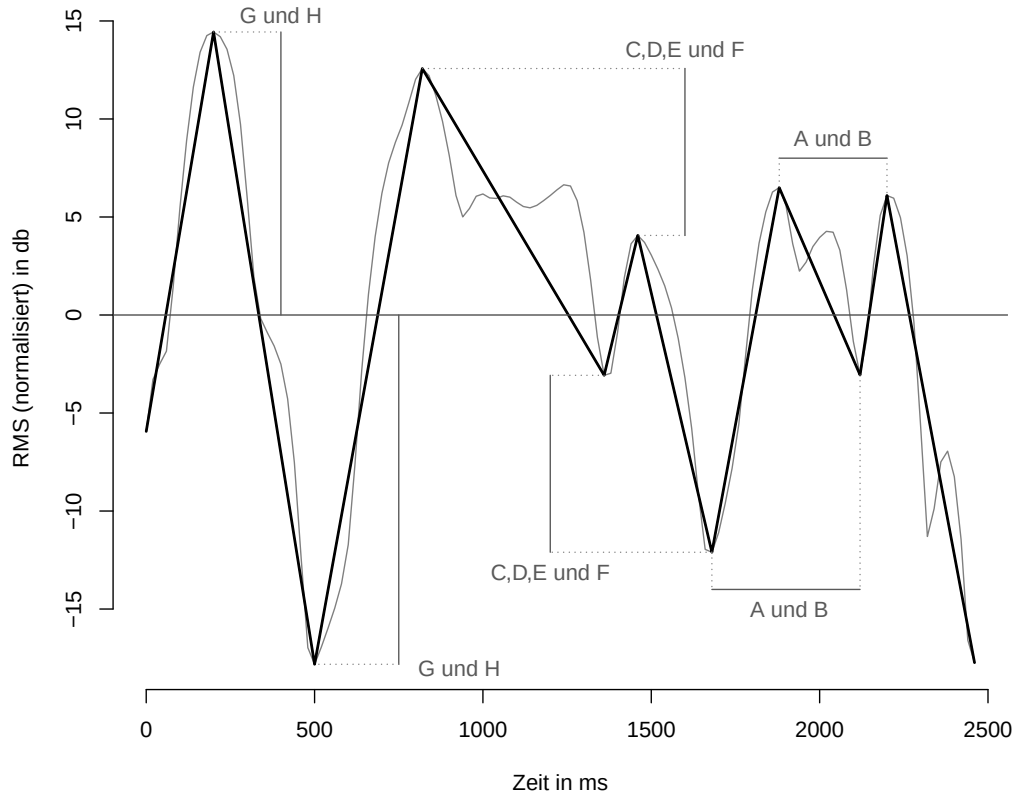


Abbildung 4.1: *Min-max Kurve (schwarz) über normalisierter RMS-Kontur (grau) sowie RMS Rhythmizitäts-Parameter A bis H.*

Anhand von Differenzmessungen zwischen den RMS-Minima bzw. -Maxima wurden einerseits Parameter auf zeitlicher Ebene berechnet, andererseits Parameter auf Ebene der RMS-Werte. Beim Vergleich von Sprache unter und ohne Alkoholeinfluss auf Basis der Energiefunktion lassen sich damit Aussagen über die Änderung der Dynamik in der Energiefunktion von Sprachsignalen treffen, die einen Rückschluss auf verschiedene Eigenschaften lautsprachlicher Äußerungen wie Sprechgeschwindigkeit oder Lautstärkeschwankungen zulassen.
Die einzelnen Parameter wurden jeweils für Folgen von RMS-Minima sowie analog für Folgen von RMS-Maxima berechnet und in den Tabellen der Auswertungsergebnisse 4.1, 4.2 und 4.3 jeweils mit *min* bzw. *max* markiert. Die Einzelmessungen der verschiedenen Differenzen zwischen den RMS-Minima und -Maxima

erfolgten dabei für jede Energiekurve separat, d.h. innerhalb jeder Äußerung, die Berechnung der Parameter äußerungsübergreifend für jeden Sprechstil. Messwerte A und B sind Mittelwert und Standardabweichung aller zeitlichen Abstände aufeinanderfolgender RMS-Minima bzw. -Maxima. Mit diesen zeitlichen Abständen wurde ebenfalls eine Adaption des in Kapitel 3.1 vorgestellten $nPVI$ (Grabe and Low [2002]) berechnet. Der *normalized Pairwise Variability Index* der *min-min* bzw. *max-max* Intervalle $nPVI_{min}$ bzw. $nPVI_{max}$ (Grabe and Low [2002]) ist der durchschnittliche Dauerunterschied aufeinanderfolgender *min-min* bzw. *max-max* Intervalle dividiert durch die Summe der Dauern derselben Intervalle. Weiterhin wurden der Median und der Interquartilsabstand der Differenzen zwischen den Werten aufeinanderfolgender RMS-Minima und Maxima berechnet (C, D) sowie der Median und der Interquartilsabstand für die Absolutwerte der Differenzen zwischen diesen Werten (E, F). G und H entsprechen dem Median und Interquartilsabstand der relativen Distanz von Minimum oder Maximum zum Mittelwert des normalisierten RMS. A im Besonderen, aber auch B können als Maße zur Bewertung der Sprechgeschwindigkeit und deren Dynamik verwendet werden (siehe Kapitel 5.2). C, D, E und F beschreiben die Energiedynamik des Signals, G und H die Änderung der Energiedynamik.

4.2 RMS Rhythmizitäts-Parameter - Ergebnisse

Die RMS Rhythmizitäts-Parameter wurden für die 150 Sprecher mit einer Blutalkoholkonzentration über $0.4\overline{9}‰$ (im alkoholisierten Zustand) und jeden Sprechstil (gelesene, spontane und Kommandosprache) der Sprachaufnahmen, die unter Alkoholeinfluss und in nüchternem Zustand durchgeführt wurden, gesondert berechnet. Auch hier wurde äußerungsübergreifend jeweils ein Wert für jeden Sprechstil ermittelt, damit also für jeden Sprecher 6 Werte (3 Sprechstile $\times$ 2 Sprecherzustände). Die Auswertung erfolgte ebenfalls per RM-MANOVA mit dem jeweiligen Parameter als abhängiger Variable, den 'within factors' Alkoholisierung (Stufen: alkoholisiert, nüchtern) und Sprechstil (Stufen: gelesen, spontan,

Kommando), dem 'between factor' Geschlecht (Stufen: weiblich, männlich) sowie dem 'random factor' Sprecher. Da sich größtenteils signifikante Interaktionen der Faktoren Alkoholisierung und Sprechstil ergeben, wurde gegebenenfalls ein Tukey HSD Post-hoc-Test durchgeführt, und nur die relevanten Sprechstile ermittelt. Die Ergebnisse der einzelnen Parameter mit den jeweiligen Signifikanzniveaus, den betroffenen Sprechstilen und den Prozentwerten der Sprecher, bei welchen sich der jeweilige Parameter bei a-Sprache gegenüber na-Sprache erhöht bzw. verringert, sind aus Tabelle 4.1 ersichtlich.

Parameter	% ⇑ na zu a	% ⇓ na zu a	p-level	Sprechstil(e)
A_{min}	66	34	$p < 0.001$	r,s
A_{max}	62.67	37.33	$p < 0.001$	r,s
B_{min}	65.33	34.67	$p < 0.001$	r,s
B_{max}	60	40	$p < 0.001$	r,s
$nPVI_{min}$	59.33	40.67	$p < 0.001$	r
$nPVI_{max}$	56.67	43.33	$p < 0.001$	r
C_{min}	61.33	38.67	n.s.	-
C_{max}	43.33	56.67	$p < 0.05$	c
D_{min}	54.67	45.33	n.s.	-
D_{max}	72.67	27.33	$p < 0.001$	r,s,c
E_{min}	54	46	n.s.	-
E_{max}	74	26	$p < 0.001$	r,s,c
F_{min}	58	42	$p < 0.01$	r
F_{max}	70	30	$p < 0.001$	r,s,c
G_{min}	53.33	46.67	n.s.	-
G_{max}	58	42	$p < 0.001$	r
H_{min}	66.67	33.33	$p < 0.001$	r
H_{max}	70.67	29.33	$p < 0.001$	r,s

Tabelle 4.1: *Auswertungsergebnisse zu den RMS Rhythmizitäts-Parametern mit Prozentwerten der Sprecher, bei welchen sich der jeweilige Parameter bei a-Sprache gegenüber na-Sprache erhöht (% ⇑ na zu a) bzw. verringert (% ⇓ na zu a). Sprechstile r (gelesen), s (spontan) und c (Kommando).*

Vor allem bei den die Maxima betreffenden Parametern zeigt sich ein klares Bild. Alle Werte bis auf den Parameter C_{max} zeigen eine signifikante Erhöhung bei a-Sprache (vor allem gelesen) an. Bei den Parametern die Minima betreffend stellt sich im statistischen Mittel zwar eine Erhöhung der Parameter bei a-Sprache gegenüber na-Sprache ein, jedoch ist der Unterschied hier nicht immer signifikant. Abbildung 4.2 verdeutlicht die Veränderungen der Parameter zwischen na- und a-Sprache.

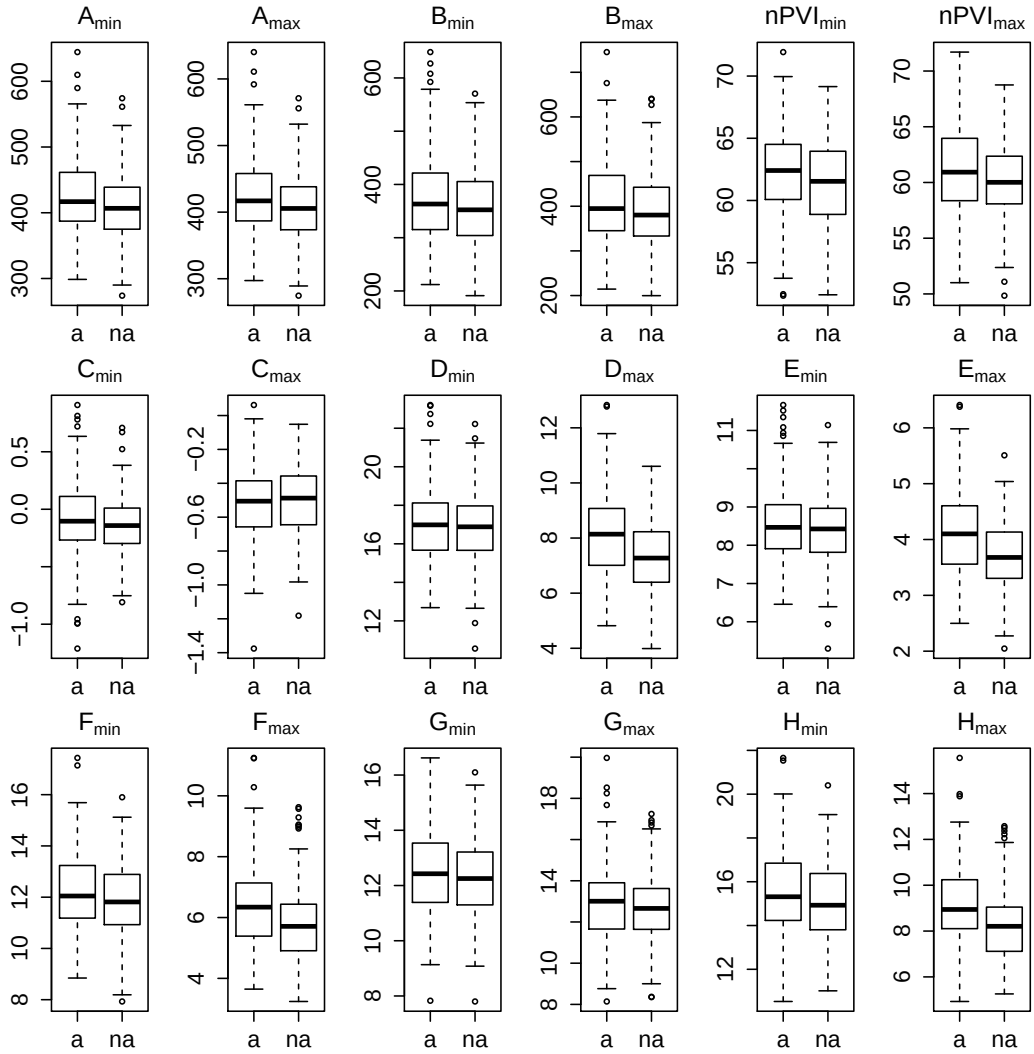


Abbildung 4.2: *Boxplots der RMS Rhythmizitäts-Parameter A bis H für Sprache unter Alkoholeinfluss (a) und in nüchternem Zustand geäußerter Sprache (na).*

4.3 RMS Rhythmizitäts-Parameter - Kontrollgruppenversuche - Ergebnisse

Um auszuschließen, dass es sich bei den gefundenen Effekten um Einflüsse verborgener Faktoren handelt, wurde auch im Falle der RMS Rhythmizitäts-Parameter eine Auswertungsreihe auf Basis der 20 Kontrollgruppensprecher vorgenommen. Dazu wurden die RMS Rhythmizitäts-Parameter für jeden der 20 Kontrollgruppensprecher und jeden Sprechstil (gelesene, spontane und Kommandosprache) der Sprachaufnahmen, die unter Alkoholeinfluss und in nüchternem Zustand durchgeführt wurden, sowie der Kontrollgruppenaufnahmen, gesondert berechnet. Hier wurde ebenfalls äußerungsübergreifend jeweils ein Wert für jeden Sprechstil ermittelt, damit also für jeden Sprecher 9 Werte (3 Sprechstile $\times$ 3 Sprecherzustände). Mit Hilfe einer RM-MANOVA, den 'within factors' Alkoholisierung (Stufen: alkoholisiert, nüchtern) und Sprechstil (Stufen: gelesen, spontan, Kommando), dem 'between factor' Geschlecht (Stufen: weiblich, männlich) sowie dem 'random factor' Sprecher und dem jeweiligen Parameter als abhängiger Variable, wurden zum Einen für a- und na-Sprache und zum Anderen für cna-Sprache und na-Sprache Signifikanztests durchgeführt. Da sich auch hier signifikante Interaktionen der Faktoren Alkoholisierung und Sprechstil ergeben, wurde gegebenenfalls ein Tukey HSD Post-hoc-Test vorgenommen, und nur die relevanten Sprechstile ermittelt.

Die Ergebnisse der einzelnen Parameter mit den jeweiligen Signifikanzniveaus, den betroffenen Sprechstilen und den Prozentwerten der Sprecher, bei welchen sich der jeweilige Parameter bei a-Sprache gegenüber na-Sprache erhöht bzw. verringert, sind aus den Tabellen 4.2 und 4.3 ersichtlich. Die Ergebnisse in Tabelle 4.3 lassen keinen signifikanten Unterschied zwischen den Parametern von na-Sprache und cna-Sprache erkennen, die Ergebnisse in Tabelle 4.2 hingegen zeigen für die Parameter größtenteils für gelesenes Sprachmaterial signifikante Unterschiede zwischen a- und na-Sprache an.

Parameter	% ⇑ na zu a	% ⇓ na zu a	p-level	Sprechstil(e)
A_{min}	80	20	$p < 0.001$	r
A_{max}	80	20	$p < 0.001$	r
B_{min}	85	15	$p < 0.001$	r
B_{max}	85	15	$p < 0.001$	r
$nPVI_{min}$	75	25	$p < 0.01$	r
$nPVI_{max}$	70	30	$p < 0.001$	r
C_{min}	65	35	n.s.	-
C_{max}	50	50	n.s.	-
D_{min}	75	25	$p < 0.1$	s
D_{max}	60	40	$p < 0.1$	r
E_{min}	75	25	$p < 0.1$	s
E_{max}	60	40	$p < 0.1$	r
F_{min}	80	20	$p < 0.05$	r
F_{max}	75	25	$p < 0.01$	r
G_{min}	75	25	$p < 0.1$	r
G_{max}	60	40	$p < 0.01$	r,s
H_{min}	80	20	$p < 0.05$	r
H_{max}	70	30	$p < 0.1$	r

Tabelle 4.2: *Auswertungsergebnisse zu den RMS Rhythmizitäts-Parametern Kontrollgruppe alkoholisiert-nüchtern mit Prozentwerten der Sprecher, bei welchen sich der jeweilige Parameter bei a-Sprache gegenüber na-Sprache erhöht (% ⇑ na zu a) bzw. verringert (% ⇓ na zu a). Sprechstile r (gelesen), s (spontan) und c (Kommando).*

Parameter	% ⇑ na zu cna	% ⇓ na zu cna	p-level	Sprechstil(e)
A_{min}	60	40	n.s.	-
A_{max}	70	30	n.s.	-
B_{min}	55	45	n.s.	-
B_{max}	55	45	n.s.	-
$nPVI_{min}$	55	45	n.s.	-
$nPVI_{max}$	70	30	n.s.	-
C_{min}	55	45	n.s.	-
C_{max}	70	30	n.s.	-
D_{min}	70	30	n.s.	-
D_{max}	45	55	n.s.	-
E_{min}	65	35	n.s.	-
E_{max}	45	55	n.s.	-
F_{min}	75	25	n.s.	-
F_{max}	65	35	n.s.	-

Parameter	% ⇑ na zu cna	% ⇓ na zu cna	p-level	Sprechstil(e)
G_{min}	70	30	n.s.	-
G_{max}	50	50	n.s.	-
H_{min}	65	35	n.s.	-
H_{max}	45	55	n.s.	-

Tabelle 4.3: *Auswertungsergebnisse zu den RMS Rhythmizitäts-Parametern Kontrollgruppe Kontrollgruppenaufnahmen-nüchtern mit Prozentwerten der Sprecher, bei welchen sich der jeweilige Parameter bei cna-Sprache gegenüber na-Sprache erhöht (% ⇑ na zu cna) bzw. verringert (% ⇓ na zu cna). Sprechstile r (gelesen), s (spontan) und c (Kommando).*

Die RMS Rhythmizitäts-Parameter, die auf Basis der Abfolge von Minima und Maxima in der Energiefunktion gewonnen werden können und deren Dynamik erfassen, scheinen demnach aus statistischer Sicht zur Unterscheidung von Sprache unter Alkoholeinfluss und in nüchternem Zustand geäußerter Sprache geeignet zu sein.

4.4 RMS Rhythmizitäts-Parameter - Diskussion

Der entwickelte Algorithmus zur Reduzierung der zugrundeliegenden RMS-Kurven auf Abfolgen von Minima und Maxima funktioniert schnell und zuverlässig. Er liefert eine robuste Methode zur Bewertung der Dynamik im Energieverlauf von sprachlichen Äußerungen. Insbesondere arbeitet er vollständig automatisch auf Basis des Sprachsignals und bedarf keiner Segmentierung. Die Ergebnisse einiger RMS Rhythmizitäts-Parameter bestätigen auch gefundene Effekte der segmentationsbasierten Rhythmusparameter aus Kapitel 3. Im Allgemeinen deuten die Ergebnisse hinsichtlich des ALC auf langsamere und unregelmäßigere Sprache im alkoholisierten Zustand der Sprecher hin, da im statistischen Mittel fast alle Parameter höhere Werte bei Sprache unter Alkoholeinfluss gegenüber Sprache ohne Alkoholeinfluss aufweisen (siehe auch Grafik 4.2). Damit konnte aus statistischer Sicht Hypothese 2 bestätigt werden. Veränderungen in der rhythmischen Struktur der Sprache in Form von Unregelmäßigkeiten und Lautstärkeschwankungen, die durch Alkoholisierung hervorgerufen werden, spiegeln sich zu einem

gewissen Teil in Veränderungen in der Dynamik der RMS- oder Energiefunktion der Sprache wider. Bei einigen Parametern wurden geringfügige geschlechtsspezifische Unterschiede gefunden, die jedoch unabhängig von der Alkoholisierung der Sprecher auftreten. Da sich für die Rhythmizitäts-Parameter sprecherspezifische Unterschiede ergeben (Vergrößerung aber auch Verkleinerung der Werte), ist wiederum keine allgemeingültige Aussage möglich. Wäre eine solche allgemeingültige Aussage möglich, würde zur automatischen Klassifikation von Sprache eine einzelne Sprachaufnahme eines Sprechers nicht ausreichen. Zusätzlich müssten Referenzdaten zur Bewertung vorhanden sein. Beispielsweise könnte eine sprachgesteuerte Applikation (z.B. Navigationssystem in einem Kraftfahrzeug) Sprachbefehle, die wiederholt verwendet werden, im Hintergrund speichern. Daraufhin könnte die in Kapitel 4.1 beschriebene Methode verwendet werden, um die Parameter speziell für Sprache ohne Alkoholeinfluss (dabei wird vorausgesetzt, dass der Sprecher zum Zeitpunkt der Sprachbefehlseingabe nüchtern ist) zu berechnen. Ein Vergleich mit den Parametern, die zu aktuell gesprochenen Sprachbefehlen ermittelt werden, würde dann eine Bewertung des Sprecherzustandes (alkoholisiert oder nüchtern) erlauben.

Für alle RMS Rhythmizitäts-Parameter wurden Korrelationen ihrer Änderungen mit dem Blutalkoholwert berechnet (162 Sprecher, alle Sprechstile zusammen). Die höchste aber dennoch geringe Korrelation zwischen der Änderung des Parameters und der BAK ergab sich für A_{max} (Mittelwert aller zeitlichen Abstände aufeinanderfolgender RMS-Maxima) mit $r = 0.22$ ($p < 0.01$). Im Allgemeinen lässt also der Grad der Veränderung der RMS Rhythmizitäts-Parameter keinen Rückschluss auf den Grad der Alkoholisierung zu. Alkoholkonsum wirkt sich damit im Falle der RMS Rhythmizitäts-Parameter vermutlich individuell unterschiedlich aus. Korrelationen innerhalb der Daten eines Sprechers sind auch hier nicht möglich, da pro Sprecher und Aufnahme nur eine Messung der Blutalkoholkonzentration vorgenommen wurde.

Kapitel 5

Sprechgeschwindigkeit

Wie bereits in Kapitel 1 angeführt, wurde in den bisherigen Studien zu Sprache unter Alkoholeinfluss über eine Verlangsamung der Sprechgeschwindigkeit bei Sprache unter Alkoholeinfluss gegenüber in nüchternem Zustand geäußerter Sprache berichtet. Um diese Befunde statistisch zu bestätigen bzw. zu widerlegen, wurde die Sprechgeschwindigkeit wie im Folgenden beschrieben anhand verschiedener Verfahren ermittelt, und es wurde untersucht, ob diese Verfahren zum selben Ergebnis führen bzw. welche Beziehungen untereinander bestehen. Zum Einen wurde dafür eine Schätzung der Sprechgeschwindigkeit auf Basis der phonetischen Segmentierung vorgenommen, zum Anderen auf Basis der Energiefunktion des Sprachsignals. Beide Ansätze werden hinsichtlich ihrer Vor- und Nachteile kritisch beleuchtet und ihre Beziehung untereinander sowie ihre Beziehung zu perzeptiven Daten dargelegt. Als Hypothese gilt:

Hypothese 3: Sprache alkoholisierter Personen weist eine verlangsamte Sprechgeschwindigkeit gegenüber in nüchternem Zustand geäußerter Sprache auf. Dies zeigt sich

a) in einer Änderung der Dauern phonetischer Einheiten und kann

b) auf Basis der Energiefunktion des Sprachsignals, bei der eine Änderung der Abstände zwischen Energiemaxima auftritt, bestätigt werden.

c) Verlangsamte Sprechgeschwindigkeit ist weiterhin ein Indiz für die Alkoholisierung eines Sprechers bei der Perzeption von Sprache.

Als Maß für die Sprechgeschwindigkeit dienen meist linguistische Einheiten pro Zeiteinheit, wie beispielsweise Silben pro Sekunde oder Sprachlaute pro Sekunde. Da, wie im Falle des ALC, bei größeren Sprachkorpora eine manuelle Zählung der linguistischen Einheiten auf Grund des großen zeitlichen und personellen Aufwandes relativ impraktikabel und unökonomisch ist, wird eine Schätzung der Anzahl linguistischer Einheiten und gleichzeitig eine Schätzung der Sprechgeschwindigkeit meist auf automatischer Ebene mit Hilfe von Spracherkennungsverfahren vorgenommen. Diese Vorgehensweise birgt Nachteile. Automatische Spracherkenner arbeiten zwar relativ schnell, jedoch sind ihre Schätzungen bei manchem Sprachmaterial wie beispielsweise Dialogen mit hohem Sprechtempo nicht zuverlässig genug, da die akustischen Modelle zur Erkennung meist nur für normales Sprechtempo ausgelegt sind. Das Erkennungsergebnis ist damit bei hoher Sprechgeschwindigkeit oftmals fehlerhaft. Eine adäquate Schätzung der Sprechgeschwindigkeit lässt sich in diesem Fall deshalb schwer erreichen.
Besteht vorab Kenntnis über die Sprechgeschwindigkeit, kann das Spracherkennungsergebnis erheblich verbessert werden. Gelingt beispielsweise eine Klassifikation in langsame, normale oder schnelle Sprache, ist es möglich, vor dem eigentlichen Erkennungsprozess der Sprechgeschwindigkeit angepasste akustische Modelle auszuwählen. Dies ist vor allem bei höheren Sprechgeschwindigkeiten vorteilhaft, da die Erkennungsleistung gewöhnlich mit zunehmender Sprechgeschwindigkeit abnimmt und durch die Vorauswahl der passenden Modelle verbessert werden kann. Die Verwendung von sprechgeschwindigkeitsabhängigen Modellen zusammen mit einem Sprechgeschwindigkeitsklassifikator führte in der Studie von Martinez et al. [1998] zu einer Reduzierung der Wortfehlerquote um 32%.

Zunächst wird das auf der phonetischen Segmentierung beruhende Sprechgeschwindigkeitsmaß SR nochmals aufgegriffen (Kapitel 5.1), bevor das neu ent-

wickelte Sprechgeschwindigkeitsmaß $SRRP$ vorgestellt wird (Kapitel 5.2 bzw. 5.2.1). In Kapitel 5.2.2 werden die Ergebnisse zum Sprechgeschwindigkeitsmaß $SRRP$ kurz zusammengefasst. Daraufhin folgt in Kapitel 5.3 eine Evaluierung von $SRRP$ durch Korrelationsberechnungen mit dem Sprechgeschwindigkeitsmaß SR_P (siehe auch Kapitel 5.2.1). Um festzustellen, ob die in der Akustik gefundenen Effekte auch perzeptiv nachzuweisen sind, wurde außerdem ein Perzeptionsexperiment durchgeführt, welches in Kapitel 5.4 beschrieben und die Ergebnisse dargestellt werden. In der abschließenden Diskussion (Kapitel 5.5) werden die Ergebnisse der auf akustischen Analyseverfahren beruhenden Sprechgeschwindigkeitsschätzungen und die Ergebnisse zur Perzeption kritisch beleuchtet.

5.1 Sprechgeschwindigkeitsmaß SR

Eine möglichst akkurate Schätzung der Sprechgeschwindigkeit lässt sich wie in Kapitel 3.1 beschrieben mit Hilfe einer vorab verfügbaren zeitlichen Segmentierung des Sprachmaterials erreichen, wobei hier zu beachten ist, welche linguistische Einheit ausgewählt wird und wie diese im Rahmen der Ermittlung der Sprechgeschwindigkeit definiert wird. Für das Sprechgeschwindigkeitsmaß SR wurde die Anzahl der Silben, und damit die Anzahl der vokalischen Elemente V (Nuklei) pro Sekunde berechnet (siehe Kapitel 3.1). Der Vollständigkeit halber seien die Ergebnisse von SR und SR_P (inklusive äußerungsinterner Pausen) in Tabelle 5.1 nochmals kurz aufgeführt.

Parameter	% ⇑ na zu a	% ⇓ na zu a	p-level	Sprechstile
SR	25.33	74.67	$p < 0.001$	r,s
SR_P	25.33	74.67	$p < 0.001$	r,s

Tabelle 5.1: *Auswertungsergebnisse Sprechgeschwindigkeitsmaß SR (ohne Pausen) und SR_P (inklusive Pausen) mit Prozentwerten der Sprecher, bei welchen sich der jeweilige Parameter bei a-Sprache gegenüber na-Sprache erhöht (% ⇑ na zu a) bzw. verringert (% ⇓ na zu a). Sprechstile r (gelesen) und s (spontan).*

5.2 Sprechgeschwindigkeitsmaß $SRRP$

Verfahren, die eine Sprechgeschwindigkeitsschätzung unabhängig von einer Segmentierung vornehmen, evaluieren vorwiegend akustische Eigenschaften eines Sprachsignals. Die Evaluierung beruht größtenteils auf der Zählung markanter Ereignisse wie Maxima im Energieverlauf. Eigens dafür entwickelte Algorithmen helfen, die relevanten Maxima nach verschiedenen Kriterien auszuwählen und erlauben damit letztendlich eine Schätzung der Sprechgeschwindigkeit (Kitaazawa et al. [1997], Morgan et al. [1997] und Morgan and Fosler-Lussier [1998]). Pfau and Ruske [1998] bedienten sich einer modifizierten und geglätteten Energiekurve, um Vokale beziehungsweise Vokalcluster, und somit Silbennuklei zu finden, deren Anzahl der Anzahl der Silben in einer Äußerung entspricht. Sie gingen hierbei vereinfacht davon aus, dass ein Silbennukleus jeweils einem Vokalcluster zuzuordnen ist. Die Methode *tcssbc* (*temporal correlation and selected sub-band correlation*) nach Narayanan und Wang (Narayanan and Wang [2005], Wang and Narayanan [2007]) beinhaltet neben einer spektralen und temporalen Korrelation zusätzlich einen Glättungsmechanismus und berücksichtigt einen Schwellenwert bei der Bestimmung relevanter Maxima im Energieverlauf. Diese komplexe Methode wurde auch von Dekens et al. [2007] im Vergleich mit anderen als zuverlässigste eingestuft. Eine akkuratere Schätzung der Sprechgeschwindigkeit verglichen mit der *tcssbc* Methode kann laut Zhang and Glass [2009] erreicht werden, indem nach Bestimmung der Energiefunktion mit Hilfe eines kurzen Energiesignalstücks eine Schätzung der Intervalldauer zwischen aufeinanderfolgenden Silbennuklei bzw. Energiemaxima erfolgt, und diese Intervalldauer daraufhin zur Bestimmung der restlichen Nuklei bzw. Energiemaxima des jeweiligen Signals herangezogen wird.

5.2.1 Sprechgeschwindigkeitsmaß $SRRP$ - Methode

Im Rahmen dieser Arbeit wurde ein neuer Ansatz zur Beurteilung der Sprechgeschwindigkeit auf Basis der in Kapitel 4 eingeführten RMS Rhythmizitäts-Parameter entwickelt und das Resultat mit dem in Kapitel 3.1 definierten segmen-

tationsbasierten Sprechgeschwindigkeitsmaß SR verglichen. SR entspricht der Anzahl der Silben pro Sekunde und damit dem Verhältnis aus der Anzahl der vokalischen Elemente V und der Gesamtdauer jeder Äußerung (siehe Formel 3.15). Zur Evaluierung des im Folgenden beschriebenen Rhythmizitäts-Parameters der Sprechgeschwindigkeit $SRRP$ (*Speech Rate Rhythmicity Parameter*) wurde der SR Wert inklusive äußerungsinterner Pausen SR_P als Referenz verwendet, da auch zur Berechnung von $SRRP$ die gesamte Energiekurve herangezogen wurde, ohne dabei gegebenenfalls auftretende Pausen bzw. Stilleintervalle auszuschließen.

Das vorgeschlagene Maß zur Beschreibung der Sprechgeschwindigkeit anhand der Energiekurve eines Sprachsignals $SRRP$ ist der aus Abbildung 4.1 ersichtliche zeitliche Parameter A der Maxima A_{max}, der Mittelwert aller zeitlichen Abstände zwischen aufeinanderfolgenden Maxima. Bezugnehmend auf das Silbenkonzept, nach welchem ein Maximum in der Sonorität im Silbennukleus und Minima in der Sonorität an den Silbengrenzen erreicht werden (Kohler [1977], Wang and Narayanan [2007]), stehen auch hier die Maxima in der Energiekurve für die Silbennuklei und die Minima für die Silbengrenzen. Die Zeitintervalle zwischen aufeinanderfolgenden Maxima entsprechen damit grob den Zeitintervallen zwischen Silbennuklei und können als Maß zur Schätzung der Sprechgeschwindigkeit betrachtet werden. Zur Verdeutlichung dient folgende Grafik 5.1, die denselben Teil der *min-max* Sequenz einer Äußerung zeigt wie Abbildung 4.1. Zusätzlich werden die durch $d_1 \dots d_4$ markierten zeitlichen Intervalle zwischen den Maxima dargestellt. $SRRP$ ist der Mittelwert aller N zeitlichen Abstände d_n zwischen den Maxima:

$$SRRP = \frac{1}{N} \sum_{n=1}^{N} d_n \tag{5.1}$$

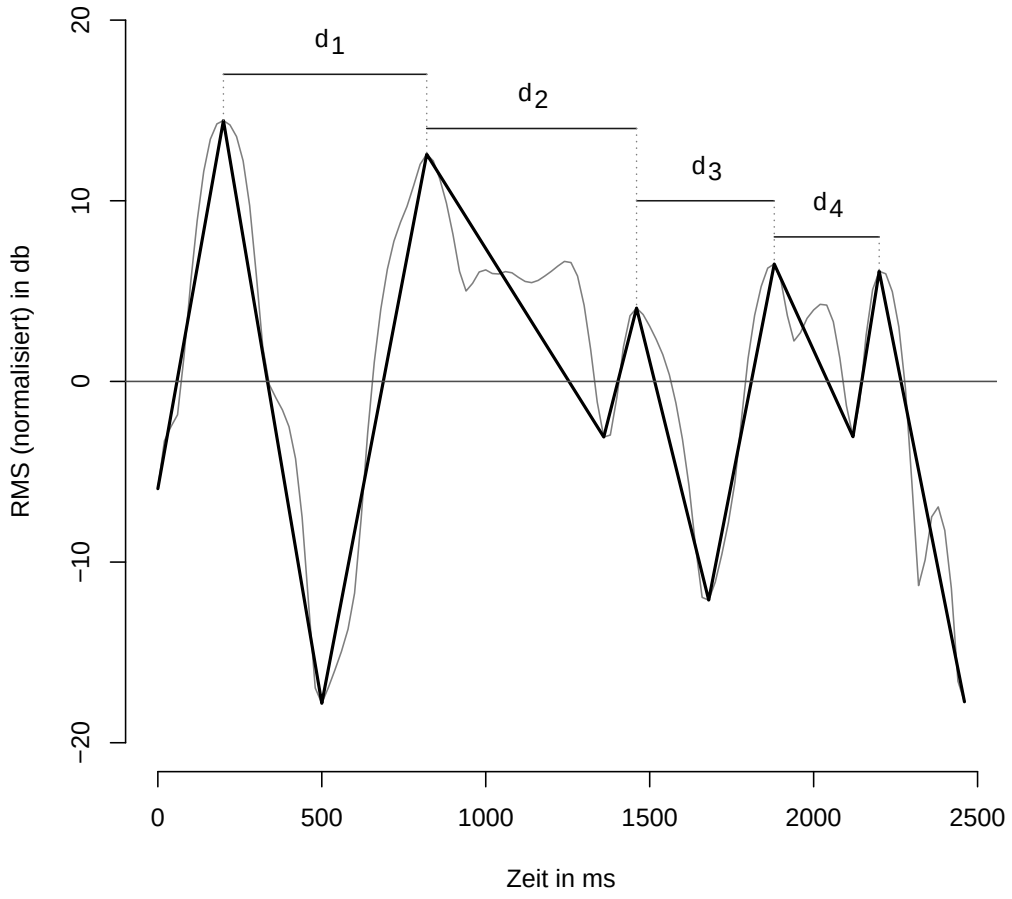


Abbildung 5.1: *Min-max Kurve (schwarz) über normalisierter RMS-Kontur (grau) mit Zeitintervallen d_1 bis d_4 zwischen aufeinanderfolgenden Maxima.*

5.2.2 Sprechgeschwindigkeitsmaß $SRRP$ - Ergebnisse

Der Vollständigkeit halber sind die Ergebnisse des RMS Rhythmizitäts-Parameters A_{max} bzw. Sprechgeschwindigkeitsmaßes $SRRP$ in folgender Tabelle 5.2 nochmals aufgeführt.

Parameter	% ⇑ na zu a	% ⇓ na zu a	p-level	Sprechstile
$SRRP$	62.67	37.33	$p < 0.001$	r,s

Tabelle 5.2: *Auswertungsergebnisse zum Sprechgeschwindigkeitsmaß SRRP (bzw. RMS Rhythmizitäts-Parameter A_{max}) mit Prozentwerten der Sprecher, bei welchen sich der Parameter bei a-Sprache gegenüber na-Sprache erhöht (% ⇑ na zu a) bzw. verringert (% ⇓ na zu a). Sprechstile r (gelesen) und s (spontan).*

Wie aus Tabelle 5.2 ersichtlich sind die Abstände zwischen den RMS-Maxima im Mittel ($SRRP$) bei a-Sprache signifikant größer ($p < 0.001$) als bei na-Sprache. Entsprechen die Maxima grob den Silbenkernen und wird die durchschnittliche Distanz zwischen ihnen größer, erhöht sich damit auch die durchschnittliche Silbendauer. Dies ist gleichbedeutend mit einer Verlangsamung der Sprechgeschwindigkeit. Auch die Ergebnisse des segmentationsbasierten Sprechgeschwindigkeitsmaßes SR_P zeigen eine Verlangsamung an. Korrelieren beide Maße hinreichend miteinander, würde dies darauf hindeuten, dass $SRRP$ eine einfache und schnelle Schätzung der Sprechgeschwindigkeit ermöglicht.

5.3 Korrelationen $SRRP$ - SR_P

Zur Evaluierung des Sprechgeschwindigkeitsmaßes $SRRP$ benötigt man wie bereits angeführt eine verlässliche Referenz, welche durch das Sprechgeschwindigkeitsmaß SR_P auf Basis der phonetischen Segmentierung zur Verfügung steht. SR_P entspricht der durchschnittlichen Sprechgeschwindigkeit einer kompletten Äußerung. Im Kontext einer Äußerung in ihrer Gesamtheit sind auch äußerungsinterne Pausen von Bedeutung, da sie sprechgeschwindigkeitsspezifische Informationen tragen und damit Auswirkungen auf die globale Sprechgeschwindigkeit haben können. Diese werden deshalb sowohl in der Berechnung von SR_P als auch in der Berechnung von $SRRP$ mit eingeschlossen.
Beide vorgestellten Sprechgeschwindigkeitsmaße (SR_P und $SRRP$) wurden für die 150 Sprecher des ALC mit einer Blutalkoholkonzentration über $0.4\overline{9}‰$ (im alkoholisierten Zustand) für a-Sprache sowie na-Sprache des jeweiligen Sprechers und für jeden Sprechstil einzeln berechnet. Damit ergeben sich jeweils 900 Werte (150 Sprecher $\times$ 3 Sprechstile $\times$ 2 Sprecherzustände) für SR_P und $SRRP$. Tabelle 5.3 zeigt die Mittelwerte von SR_P und $SRRP$ für alle 150 Sprecher getrennt nach Sprechstilen und für a- und na-Sprache.

Sprechstil	r	r	s	s	c	c
Alkoholisierung	a	na	a	na	a	na
SR_P	2.94	3.18	3.0	3.2	3.66	3.71
$SRRP$	349	316	496	468	319	324

Tabelle 5.3: *Mittelwerte von SR_P (in Silben pro Sekunde) und $SRRP$ (in ms) für 150 Sprecher, Sprache unter Alkoholeinfluss (a) und in nüchternem Zustand geäußerter Sprache (na) und die drei Sprechstile r (gelesen), s (spontan) und c (Kommando).*

Die Werte zeigen für beide Maße reduzierte Sprechgeschwindigkeiten bei a-Sprache an, außer bei der Kommandosprache. Letzteres ist nicht ungewöhnlich, da bei der Kommunikation mit einer Maschine Sprache entsprechend bewusst oder unbewusst angepasst wird, um möglichst deutlich zu artikulieren und damit der Maschine die Erkennung zu erleichtern. Diese Anpassung findet gleichermaßen alkoholisiert wie nüchtern statt, was in Kombination mit den überwiegend sehr kurzen Sprachbefehlen dazu führt, dass die Sprechgeschwindigkeitsunterschiede zwischen den Einzelnennungen der Kommandos bei a- und na-Sprache nur marginal sind. Wie aus den Ergebnissen ersichtlich, ergibt sich weder für SR_P noch $SRRP$ im Falle der Sprachkommandos eine nennenswerte Änderung in der Sprechgeschwindigkeit zwischen a- und na-Sprache. Die hohen Werte für $SRRP$, vor allem bei der Spontansprache, legen die Vermutung nahe, dass die offensichtlich vorkommenden großen Zeitintervalle zwischen Maxima oftmals Pausen bzw. Stilleintervalle beinhalten. Dasselbe Bild zeigt sich auch in Abbildung 5.2, welche die Korrelationen zwischen $SRRP$ und $1/SR_P$ für a- und na-Sprache aufgeteilt nach Sprechstilen darstellt. Wie zu erkennen, erreicht $SRRP$ bei Spontansprache Werte von bis zu 900 ms für a-Sprache und fast 750 ms für na-Sprache. Im Gegensatz zu Spontansprache kommen bei gelesener Sprache und Kommandosprache deutlich weniger und kürzere Pausen vor, da sich die Sprecher durch den vorgegebenen Lesetext vor Sprechbeginn bis zu einem gewissen Grad vorbereiten können und dadurch im Redefluss weniger

Defizite auftreten sollten. Deshalb ist auch der Korrelationskoeffizient von SR_P und $SRRP$ bei Spontansprache erwartungsgemäß am niedrigsten ($r = 0.74$ [$p < 0.001$] für a-Sprache und $r = 0.72$ [$p < 0.001$] für na-Sprache). Der höchste Korrelationskoeffizient ergibt sich für Kommandosprache ohne Alkoholeinfluss ($r = 0.87$ [$p < 0.001$]).

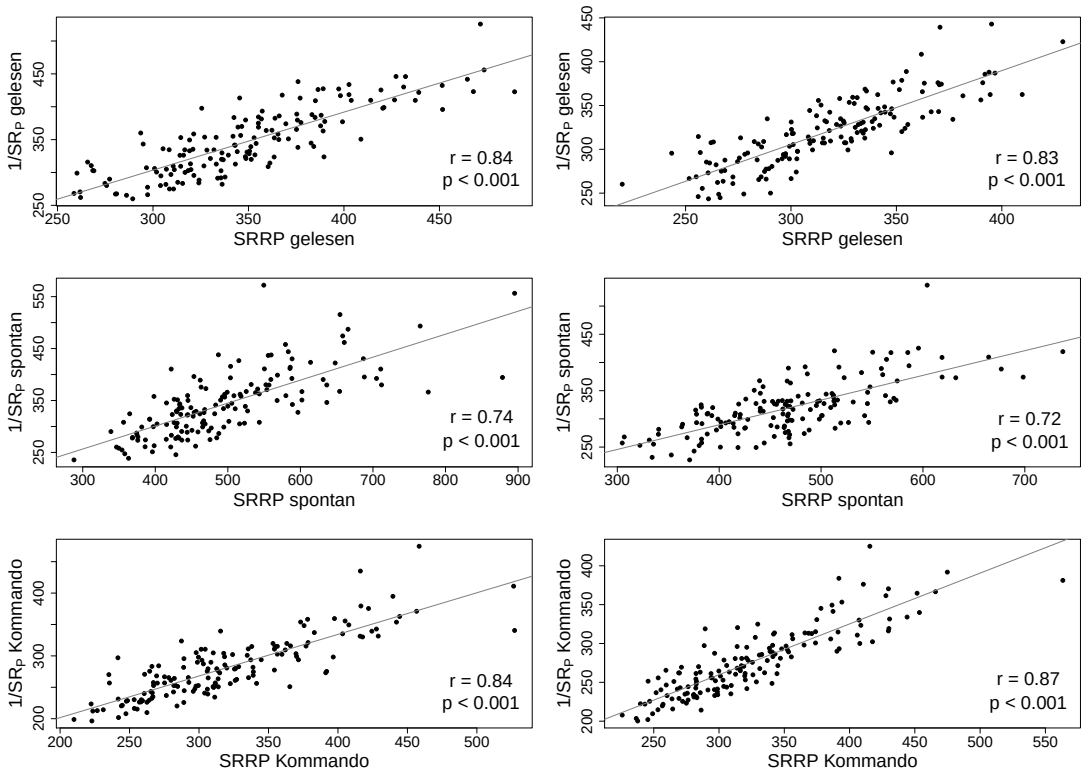


Abbildung 5.2: *Korrelationen zwischen SRRP und 1/SR_P für Sprache unter Alkoholeinfluss (links) und in nüchternem Zustand geäußerte Sprache (rechts) getrennt nach Sprechstilen.*

Die Streudiagramme lassen erkennen, dass sich $SRRP$ auch bei verschiedenen Sprechstilen konsistent verhält. Die Korrelationen mit dem segmentationsbasierten Sprechgeschwindigkeitsmaß SR_P zeigen, dass das Maß basierend auf der Kurzzeit Energiefunktion des Sprachsignals $SRRP$ die Sprechgeschwindigkeit zuverlässig widerspiegelt und damit zur Sprechgeschwindigkeitsschätzung bei divergentem Sprachmaterial herangezogen werden kann.

5.4 Sprechgeschwindigkeit - Perzeptionsexperiment

Wie aus den Ergebnissen zu den Sprechgeschwindigkeitsmaßen ersichtlich, stellt sich für einen Großteil der Sprecher eine Verlangsamung der Sprechgeschwindigkeit bei Sprache unter Alkoholeinfluss ein. Deshalb sollte ein Perzeptionsexperiment Aufschluss darüber geben, ob verlangsamte Sprache gleichzeitig ein Indiz für Alkoholisierung bei der Perzeption von Sprache sein kann.

5.4.1 Perzeptionsexperiment - Methode

Um diese Frage zu beantworten, wurde zunächst Datenmaterial der Kontrollgruppe synthetisch manipuliert. Dafür wurden aus den 19 in nüchternem Zustand gelesenen Äußerungen[1] und denselben Äußerungen der Kontrollgruppenaufnahmen automatisch jeweils 8 bzw. insgesamt 16 kurze möglichst übereinstimmende Teilstücke von durchschnittlich 2.3 Sekunden Audio extrahiert. Die Audioteilstücke der Kontrollgruppe wurden mit Hilfe des Praat[2] (Boersma [2001]) Algorithmus *lengthen (overlap-add)* pauschal um 5% gelängt, um eine künstliche Verlangsamung der Sprache zu erreichen (als Referenz diente die errechnete durchschnittliche Änderung der Silbenrate SR auf Basis der phonetischen Segmentierung. Diese zeigte eine Reduzierung der Sprechgeschwindigkeit um 4.42% bei a-Sprache im Vergleich zu na-Sprache an). Insgesamt ergeben sich damit 160 Stimuluspaare (20 Sprecher mit je 8 Stimuluspaaren, welche jeweils den Stimulus der na-Sprache und den dazugehörigen synthetisch gelängten Kontrolldatenstimulus vom gleichen Sprecher umfassen).

Die Hörer wurden gebeten, in einem 'forced-choice' Diskriminationstest aus jedem dieser Stimuluspaare, also jeweils einem Stimulus der na-Sprache und dem

[1] Es sind 19 Aufnahmeelemente sowohl unter Alkoholeinfluss als auch in nüchternem Zustand der Sprecher und weiterhin der Kontrollgruppe in gleichem Wortlaut vorhanden. Nur Aufnahmeelemente, die in den Sprecherzuständen in gleichem Wortlaut vorhanden sind, können im Perzeptionsexperiment verwendet werden. Eine Liste dieser Aufnahmeelemente findet sich in Anhang C.

[2] http://www.praat.org/, Version 5.2.18

dazugehörigen manipulierten Stimulus, denjenigen auszuwählen, bei dem sie den Sprecher als alkoholisiert einstufen. Dabei war es erlaubt, jeden Stimulus bis zu 5 mal anzuhören. Jeder Hörer hört im Laufe des Experiments genau jeden Sprecher einmal, dementsprechend insgesamt 20 Stimuluspaare (randomisiert). Damit wird das Problem der statistischen Abhängigkeit der Stimuli von einem Sprecher vermieden. Nach 8 Hörern sind damit alle 160 Stimuluspaare bereits 1 mal bewertet worden. Bei 48 Hörern (24 weiblich, 24 männlich, 20-36 Jahre, 24.6 Jahre Durchschnitt) wurde jedes Stimuluspaar demnach insgesamt 6 mal bewertet.

5.4.2 Perzeptionsexperiment - Ergebnisse

In 55.5% der Fälle wurde der manipulierte Stimulus als derjenige ausgewählt, bei dem der Sprecher vermeintlich alkoholisiert war, was bei 533 aus 960 Positivantworten signifikant über Zufall (50%) liegt ($\chi^2 = 11.5, p < 0.001$). Die Antworten fallen also nicht rein zufällig aus. Es besteht eine signifikant höhere Wahrscheinlichkeit, dass der Hörer denjenigen Stimulus eines Stimuluspaares als den a-Sprachstimulus auswählt, der synthetisch gelängt wurde und damit gleichzeitig, dass die Sprechgeschwindigkeit im Rahmen der Perzeption tatsächlich ein Faktor bei der Klassifikation von Sprache unter Alkoholeinfluss und in nüchternem Zustand geäußerter Sprache zu sein scheint. Mit Hilfe von MEM Analyse (Baayen [2008]) wurde der Einfluss von Sprecher- sowie Hörergeschlecht ('fixed factors', einzeln und kombiniert) auf die Antworten untersucht (R-Funktion lmer() mit family=binomial). Zudem wurden auch Sprecher und Hörer in den verschiedenen Modellen als 'random factors' kombiniert oder einzeln modelliert. Es zeigen sich keinerlei signifikante Effekte der Faktoren und keine signifikanten Unterschiede zwischen den verschiedenen Modellen.

5.5 Sprechgeschwindigkeit - Diskussion

Die Ergebnisse der verschiedenen Experimente zur Sprechgeschwindigkeit, die auf akustischen Analysen des Sprachsignals beruhen, deuten gleichermaßen auf eine Verlangsamung der Sprechgeschwindigkeit bei Sprache unter Alkoholeinfluss gegenüber in nüchternem Zustand geäußerter Sprache hin und bestätigen statistisch damit Teilhypothesen 3 a) und b). Ein Konzentrationsdefizit bzw. eine Einschränkung kognitiver Fähigkeiten und eventuell auch Störungen der Artikulatoren durch Alkoholeinfluss scheinen die Sprecherin bzw. den Sprecher dazu zu zwingen, langsamer zu sprechen, um Einbußen bei der Verständlichkeit zu vermeiden.

Korrelationen mit dem segmentationsbasierten Sprechgeschwindigkeitsmaß SR zeigen, dass das Sprechgeschwindigkeitsmaß, das auf der Kurzzeit Energiefunktion des Sprachsignals beruht ($SRRP$), Sprechgeschwindigkeit in hinreichendem Maße für verschiedene Sprechstile widerspiegelt und allein anhand des Sprachsignals ohne Kenntnis der phonetischen Segmente eine adäquate Schätzung ermöglicht.

Einfache Korrelationen mit dem Blutalkoholwert lassen für verschiedene Sprecher zwar keinen Zusammenhang zwischen einer Verlangsamung der Sprechgeschwindigkeit und steigender Alkoholisierung erkennen, es wäre jedoch interessant, ob ein Sprecher, bei dem zunehmende Alkoholisierung mit mehreren Messungen des Blutalkoholwerts dokumentiert wurde, auch eine mit steigendem Blutalkoholwert zunehmende Verlangsamung der Sprechgeschwindigkeit aufweisen würde. Leider liegen diese Daten für den ALC nicht vor.

Auch wenn das Ergebnis nicht sehr eindeutig ausfällt, zeigt das Perzeptionsexperiment, dass verlangsamte Sprache für Hörer tatsächlich ein Indiz für eine vorliegende Alkoholisierung zu sein scheint. Teilhypothese 3 c) konnte statistisch bestätigt werden. Die Ergebnisse der akustischen Analysen werden damit auch auf perzeptiver Ebene affirmiert.

Kapitel 6

Konturen

Im Rahmen dieser Arbeit wurden rhythmische Eigenschaften von Grundfrequenz- und Energiekonturen von Sprachaufnahmen, die unter Alkoholeinfluss und in nüchternem Zustand der Sprecher durchgeführt wurden, untersucht. Die Grundfrequenz- oder Tonhöhenkontur ist die gegebenenfalls interpolierte F0-Kurve zu einem Sprachsignal oder einer Musikaufnahme. Sie spiegelt rhythmische Eigenschaften des Sprachsignals zu einem Teil wider. Dies ergibt auch die Befragung von Versuchspersonen, die als Merkmal von unter Alkoholeinfluss geäußerter Sprache oftmals eine Veränderung des Rhythmus angeben (Schiel [2011]). Hierbei stellt sich die Frage, ob diese offensichtlich perzeptiv wahrnehmbare Veränderung des Rhythmus auch in der Grundfrequenzkontur zu finden ist. Laut Niebuhr [2009] spielt die Grundfrequenz eine tragende Rolle bei der Zuweisung von Prominenzpositionen und damit grundlegenden Rhythmuselementen.
Weil auch die RMS- oder Energiekontur eines Sprachsignals wie in Kapitel 4 angeführt die rhythmische Struktur gesprochener Sprache zu einem gewissen Grad widerspiegelt, wurden nicht nur die Grundfrequenzkonturen, sondern auch die RMS-Konturen, hier aber in ihrer Gesamtheit, näher untersucht. Als Hypothesen gelten:

Hypothese 4: Veränderungen in der rhythmischen Struktur der Sprache, die durch Alkoholisierung hervorgerufen werden, spiegeln sich auch in charakteristischen Veränderungen der F0-Kontur des Sprachsignals wider.

Hypothese 5: Veränderungen in der rhythmischen Struktur der Sprache, die durch Alkoholisierung hervorgerufen werden, spiegeln sich auch in charakteristischen Veränderungen der RMS-Kontur des Sprachsignals wider.

Dieses Kapitel umfasst die Erläuterung der nötigen Vorverarbeitungsschritte für F0- und RMS-Konturen (Kapitel 6.1), die Konturanalyse durch verschiedene Distanzwerte (Kapitel 6.2), die Konturanalyse durch Parametrisierung (Kapitel 6.3) und die funktionale Datenanalyse (*Functional Data Analysis* bzw. FDA, siehe z.B. Ramsay et al. [2009]) von F0- und RMS-Konturen (Kapitel 6.4). Abschließend folgt in Kapitel 6.5 eine Diskussion der verschiedenen Verfahren.

6.1 Konturen - Vorverarbeitung

Für jedes Sprachsignal der 19 relevanten, gelesenen Aufnahmeelemente (a-Sprache, na-Sprache und ggf. cna-Sprache)[1] wurde mit Hilfe des *Schaefer-Vincent Algorithmus* (Schaefer-Vincent [1983]) alle 5 ms ein F0-Wert berechnet. Dabei wurde der Suchbereich für Frauen auf 100 bis 500 Hz und bei Männern auf 50 bis 250 Hz begrenzt. Der Algorithmus findet stimmhafte Bereiche im Sprachsignal, indem er nach periodischen Signalanteilen sucht. Die stimmlosen Bereiche der gewonnenen F0-Kontur sind als Lücken im eigentlichen Verlauf erkennbar (der Schaefer-Vincent Algorithmus liefert hier $F0 = 0$). Diese wurden durch *lineare Interpolation* für die darauffolgenden Analyseschritte geschlossen, um eine lückenlose Kontur (d.h. ohne 0-Werte) für die gesamte Äußerung zu erhalten. Die lineare Interpolation ist ein Zwischenwert-Berechnungsverfahren und verbindet zwei Datenpunkte, hier den Anfangs- und Endpunkt der Lücke, durch eine Strecke miteinander. Sie wurde einer Interpolation mit Splines oder Polynomen

[1] Es sind 19 Aufnahmeelemente sowohl unter Alkoholeinfluss als auch in nüchternem Zustand der Sprecher und weiterhin der Kontrollgruppe in gleichem Wortlaut vorhanden. Nur Aufnahmeelemente, die in den Sprecherzuständen in gleichem Wortlaut vorhanden sind, können für einen direkten Konturvergleich (Distanzwerte) verwendet werden. Die Analyse durch Parametrisierung und die funktionale Datenanalyse beschränkte sich ebenfalls auf diese Aufnahmeelemente. Eine Liste der Aufnahmeelemente findet sich in Anhang C.

(Grad > 1) vorgezogen, da diese bei Extremwerten (z.B. durch Berechnungsfehler des Algorithmus) in der zu interpolierenden Kontur direkt vor oder nach den Lücken oftmals dazu neigen, die Extremwerte durch ebenfalls extreme und damit unpassende Schwingungen in die Interpolante zu integrieren.
Abbildung 6.1 zeigt beispielhaft für eine Sprachaufnahme die zugehörige F0-Kontur in ihrer Originalform mit Lücken (definiert durch die vom Algorithmus ausgegebenen Werte; stimmlose Bereiche bei $F0 = 0$) in grau sowie die linear interpolierte F0-Kontur in schwarz.

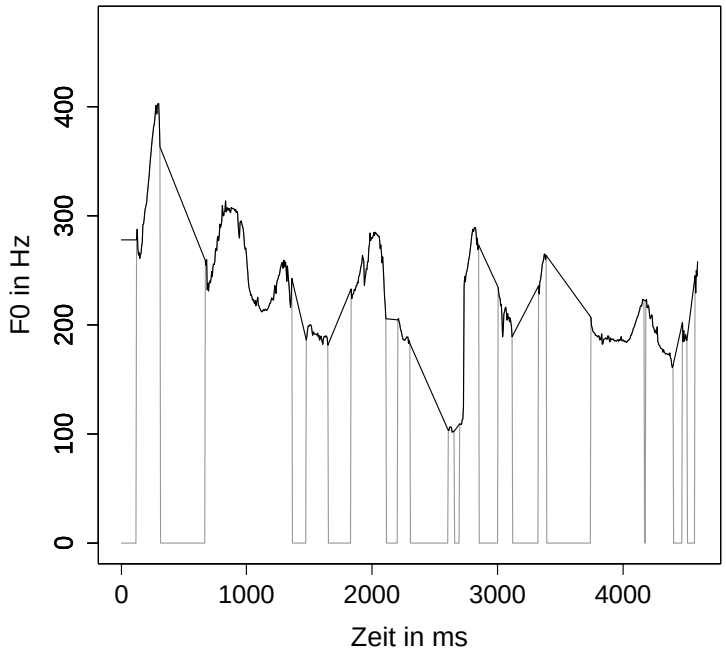


Abbildung 6.1: *F0-Kontur einer Äußerung mit Lücken (stimmlose Bereiche) in grau und linear interpoliert (schwarz).*

Die Vorverarbeitung der RMS-Konturen entspricht im Grunde der in Kapitel 4.1 vorgestellten Methode, jedoch wird die vom Algorithmus ausgegebene Kurve in ihrer Originalform beibehalten und nicht auf lokale Minima und Maxima reduziert. Wie zuvor wurde mit Hilfe des im *ASSP* Modul des EMU Speech Database Systems integrierten Algorithmus *rmsana* eine RMS-Analyse für jedes Sprachsignal der 19 relevanten, gelesenen Aufnahmeelemente (siehe Anhang C)

separat durchgeführt. Den besten Kompromiss aus Glättung und Erhaltung der Feinstruktur liefert ein *Blackman* Fenster mit 100 ms Länge und einer Fensterverschiebung von 20 ms. Neben der Normalisierung der logarithmisch skalierten RMS-Kontur durch Subtraktion des durchschnittlichen RMS-Wertes jeder Kurve wurde hier auch eine lineare Normalisierung auf Ebene der Zeit vorgenommen. Dafür wurden alle Konturen durch Abtastratenkonvertierung (verlustfreier Interpolation mit Hilfe der R-Funktion *resamp*) paarweise angeglichen, wobei die Anzahl von Abtastwerten der Kontur der in nüchternem Zustand geäußerten Sprachaufnahme (im Folgenden auch als na-Kontur bezeichnet, Konturen von Sprachaufnahmen unter Alkoholeinfluss als a-Konturen) als Referenz diente. Bei der funktionalen Datenanalyse (siehe Kapitel 6.4) wurden alle Konturen ebenfalls mit Hilfe einer Abtastratenkonvertierung (R-Funktion *resamp*) auf die gleiche Anzahl von $N = 200$ Abtastwerten abgebildet, da hier für alle der Analyse zugeführten Konturen die gleiche Länge vorausgesetzt wird.
Dieser jeweils lineare Normalisierungsprozess wurde sowohl auf die RMS- als auch die F0-Konturen angewandt, um einen direkten Vergleich von unterschiedlich langen Konturen zu ermöglichen. Eine lineare Normalisierung wurde einer nicht linearen Normalisierung wie *dynamic time warping* (DTW [Sakoe and Chiba [1978]]) vorgezogen, um eine Verzerrung der Zeitstruktur in der Dynamik des Signals auszuschließen. Laut Hypothese beinhaltet diese zeitliche Struktur gegebenenfalls Informationen über den Zustand des Sprechers hinsichtlich Alkoholisierung und sollte deshalb vor der eigentlichen Analyse nicht verändert werden.

6.2 Konturen - Distanzwerte

Eine Reihe von Untersuchungen beschäftigt sich mit der Ähnlichkeit von Tonhöhenkonturen bei Musik. Ein gegebenes Teilstück einer Tonhöhenkontur wird dabei mit einer hinterlegten Tonhöhenkontur eines Musikstückes verglichen (Francu and Nevill-Manning [2000], Lu et al. [2001], Shmulevich [2004], Zhu and Kankanhalli [2002]). Es kommen meist verschiedene dynamische Alignierungsalgorithmen wie

beispielsweise DTW zum Einsatz, um das Teilstück der Tonhöhenkontur mit der hinterlegten Kontur abzugleichen. Dafür wird das Teilstück in seiner zeitlichen Struktur sowie der Tonhöhe analysiert und gegebenenfalls so verändert, dass die daraufhin ermittelte Distanz zwischen den alignierten Konturen minimal wird.
In der automatischen Sprachsynthese spielt die Manipulation der Intonation und damit der Tonhöhe bzw. F0-Kontur eine wichtige Rolle. Bei der Evaluierung verschiedener Sprachsynthesesysteme und zum Vergleich des durch sie erzeugten Sprachsignals werden gefundene Unterschiede in der Grundfrequenz oftmals zur Beurteilung der Ähnlichkeit der Konturen herangezogen. Die Methode von Latsch and Netto [2011] kombiniert einen Mechanismus zum Angleichen der Tonhöhe (*time domain pitch synchronous overlap and add* [TD-PSOLA, Moulines and Charpentier [1990]]) mit einer zeitlichen Alignierung der Sprache mittels DTW. Clark and Dusterhoff [1999] berechneten Distanzen zwischen synthetisierten Grundfrequenzkonturen und den Originalkonturen derselben gesprochenen Sätze durch zeitliche Alignierung und verglichen die Ergebnisse mit einer Perzeptionsstudie zu denselben Daten, um zu sehen, inwieweit die objektiven Maße die perzipierten Unterschiede widerspiegeln. In ähnlichen Studien von Hermes [1998a,b] wurden Phonetiker gebeten, die Ähnlichkeit von Grundfrequenzkonturen visuell und auditiv zu bewerten. Die Ergebnisse wurden auch hier mit verschiedenen automatisch berechneten Distanzmaßen verglichen. Barlow and Wagner [1988] benutzten die mit Hilfe von DTW ermittelten Distanzen zwischen verschiedenartigen Konturen wie Energie- oder Grundfrequenzkonturen in einem Experiment zur Sprecheridentifikation. In seiner Doktorarbeit zur Modellierung von Intonationsverläufen in der Sprachsynthese stellte Moehler [1998] verschiedene Evaluierungsmethoden zur Ähnlichkeit von Konturen vor. Darunter befinden sich neben perzeptiven Evaluierungsmethoden durch Perzeptionsexperimente auch Methoden, die eine automatische Berechnung verschiedener Distanzwerte beinhalten. Zwei dieser Distanzwerte wurden im Rahmen der vorliegenden Arbeit verwendet und werden im folgenden Kapitel 6.2.1 vorgestellt.

Viele der genannten Verfahren bedienen sich einer Verzerrung der zeitlichen Struktur der Konturen, damit der Distanzwert möglichst klein ausfällt. Eine solche Verzerrung würde wie oben angesprochen im Falle der Konturen des ALC gegebenenfalls rhythmische Informationen beseitigen und relevante feine rhythmische Unterschiede zwischen den Konturen verschwinden lassen, welche aber beim Vergleich von Sprache unter Alkoholeinfluss und in nüchternem Zustand geäußerter Sprache von Bedeutung sein können. Deshalb wurden die hier vorgestellten Untersuchungen zum Konturvergleich nur mit linear zeitnormalisierten Konturen wie bei Moehler [1998] oder durch Parametrisierung der Konturen durchgeführt.

6.2.1 Distanzwerte - Methode

Zwischen zwei zeitnormalisierten Konturen können verschiedene Distanzwerte berechnet werden. Grafik 6.2 illustriert beispielhaft für F0 die physische Distanz zwischen zwei zeitnormalisierten Konturen.

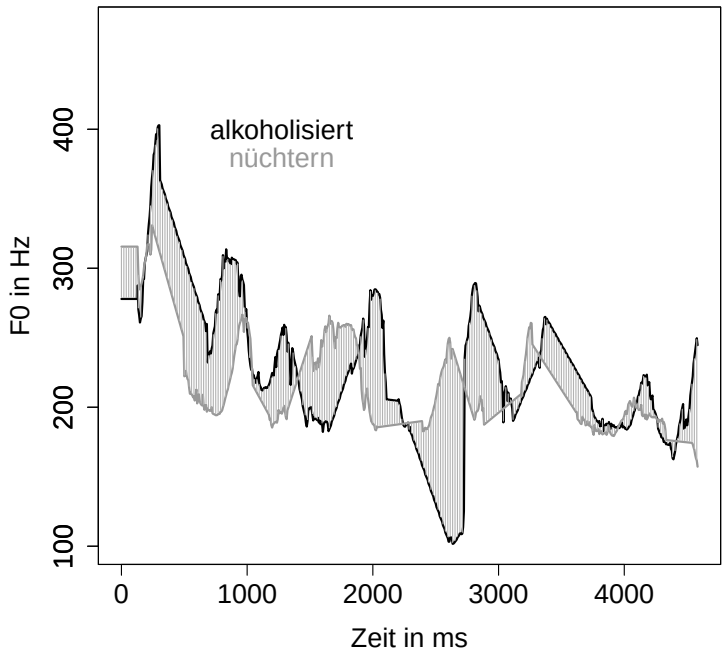


Abbildung 6.2: *Beispielhafte Darstellung der physischen Distanz zwischen zwei Konturen (F0).*

Die Hypothese im Falle von Sprache unter Alkoholeinfluss besagt für die Distanzwerte, dass diese größer sind zwischen a- und na-Konturen (im Folgenden als a-Distanz bezeichnet) als zwischen na- und cna-Konturen (im Folgenden als na-Distanz bezeichnet):

$$D(\text{a}) > D(\text{na}) \tag{6.1}$$

Im Rahmen dieser Arbeit wurden drei verschiedene Distanzwerte für F0- und RMS-Konturen berechnet und untersucht.

1. Der *root mean square error* (RMSE), der der Euklidischen Distanz zwischen zwei Vektoren derselben Länge entspricht. Hier spiegelt er die physische Distanz zwischen zwei zeitnormalisierten Konturen x und y entlang der Zeitachse wider. Ein höherer Wert deutet auf eine größere Distanz und ein niedrigerer Wert auf eine kleinere Distanz zwischen den Konturen hin (Moehler [1998]).

$$D_{\text{rmse}} = \sqrt{\frac{\sum_{t=1}^{N} \big(x(t) - y(t)\big)^2}{N}} \tag{6.2}$$

2. Das zweite Maß basiert auf dem Korrelationskoeffizienten (Moehler [1998]), der hierbei die Synchronizität der Auf- und Abbewegungen der beiden Konturen beschreibt. Die Korrelationsdistanz berechnet sich durch Subtraktion des Korrelationskoeffizienten von 1. x und y sind zeitnormalisierte Konturen, $\bar{x}$ und $\bar{y}$ ihre Mittelwerte und sd_x und sd_y ihre Standardabweichungen (Klabbers and van Santen [2004]):

$$D_{\text{corr}} = 1 - \left(\frac{1}{N-1} \sum_{t=1}^{N} \Big(\frac{x(t) - \bar{x}}{sd_x} \Big) \Big(\frac{y(t) - \bar{y}}{sd_y} \Big) \right) \tag{6.3}$$

3. Der dritte Wert ist die Distanz in einem niedrigdimensionalen spektralen Raum, der durch Parametrisierung erzeugt wird. Dafür werden beide Konturen in diesen Raum transformiert und dann die Euklidische Distanz zwi-

schen den Konturen berechnet. Die Diskrete Cosinus Transformation (DCT) zerlegt eine gerade Wellenform x in Faktoren ihrer inhärenten Cosinus Wellen $\Psi_x(\nu)$, die im Folgenden als *DCT Koeffizienten* bezeichnet werden (Harrington [2010]). DCT Koeffizienten $\Psi_x(\nu)$ mit niedrigeren Indizes ν repräsentieren dabei Bewegungen niedrigerer Frequenzen in der transformierten Wellenform, Koeffizienten mit höheren Indizes Bewegungen höherer Frequenzen. Der erste DCT Koeffizient $\Psi_x(\nu = 1)$ bleibt bei vorangegangener Normalisierung für alle Konturen derselbe und wird in der Auswertung nicht in Betracht gezogen. Durch experimentelle Variation der Anzahl niedrigerer DCT Koeffizienten zeigte sich, dass die Indizes 2 bis 7 eine bestmögliche Unterscheidung zwischen a- und na-Konturen zulassen.

$$D_{\text{dct}} = \sqrt{\sum_{i=2}^{7} \big(\Psi_x(i) - \Psi_y(i)\big)^2} \tag{6.4}$$

6.2.2 Distanzwerte - Ergebnisse

Die a- sowie na-Distanzen D_{rmse}, D_{corr} und D_{dct} wurden für jedes F0- und RMS-Konturpaar der 19 inhaltsgleichen gelesenen Äußerungen (siehe Anhang C) aller 20 Kontrollgruppensprecher berechnet. Distanzwerte können nur für diese 20 Sprecher und nicht für alle Sprecher des ALC berechnet werden, da ein Distanzwert ohne Referenz nichts aussagt. Als Referenz diente hier die Distanz zwischen der cna- und na-Kontur, die im optimalen Falle minimal sein sollte, da sich zwei Konturen von in nüchternem Zustand getätigten inhaltsgleichen Äußerungen im Idealfall nicht unterscheiden. Erwartungsgemäß müsste dagegen die Distanz zwischen den Konturen von Sprache unter Alkoholeinfluss und in nüchternem Zustand geäußerter Sprache größer sein. Die statistische Auswertung wurde anhand einer MEM Analyse (Baayen [2008]) mit Alkoholisierung und Geschlecht als 'fixed factors' sowie Aufnahmeelement und Sprecher als 'random factors' durchgeführt. Im Kontext der Distanzwerte bezeichnet der Faktor Alkoholisierung die a- und na-Distanzen.

6.2.2.1 F0

Alle drei Distanzen D_{rmse}, D_{corr} und D_{dct} sind signifikant größer zwischen a- und na-F0-Konturen (a-Distanz) als zwischen cna- und na-F0-Konturen (na-Distanz), wie aus Abbildung 6.3 ersichtlich. Es traten sehr schwache geschlechtsspezifische Effekte auf, wobei sich bei getrennten Auswertungen für beide Geschlechter ebenso signifikante Unterschiede zeigten.

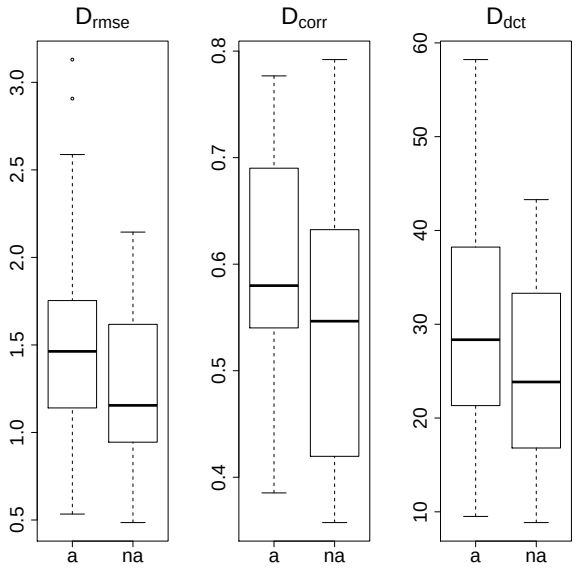


Abbildung 6.3: *Boxplots der Distanzwerte der F0-Konturen.*

Tabelle 6.1 beinhaltet neben den Signifikanzniveaus die prozentualen Anteile der Sprecher, bei welchen die a-Distanz größer ist als die na-Distanz bzw. kleiner.

Parameter	% ⇑ na zu a	% ⇓ na zu a	F-Wert	p-level
D_{rmse}	85	15	$F = 42.27$	$p < 0.001$
D_{corr}	70	30	$F = 9.34$	$p < 0.01$
D_{dct}	75	25	$F = 23.43$	$p < 0.001$

Tabelle 6.1: *Auswertungsergebnisse zu den F0-Kontur-Distanzwerten mit Prozentwerten der Sprecher, bei welchen die a-Distanz größer ist als die na-Distanz (% ⇑ na zu a) bzw. kleiner (% ⇓ na zu a).*

6.2.2.2 RMS

Auch für die Energiekonturen sind alle drei Distanzen D_{rmse}, D_{corr} und D_{dct} signifikant größer zwischen a- und na-Konturen (a-Distanz) als zwischen cna- und na-Konturen (na-Distanz) - siehe Abbildung 6.4. Es zeigten sich keine geschlechtsspezifischen Effekte.

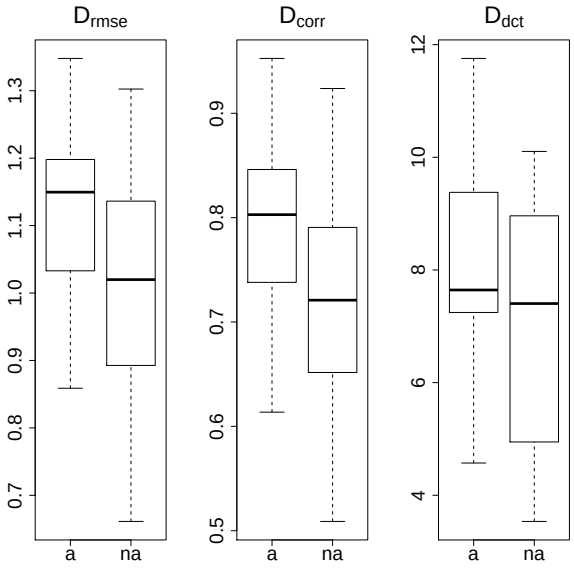


Abbildung 6.4: *Boxplots der Distanzwerte der RMS-Konturen.*

Die prozentualen Anteile der Sprecher, bei welchen eine Vergrößerung bzw. Verkleinerung der Distanzwerte auftritt, sowie das jeweilige Signifikanzniveau, sind aus Tabelle 6.2 ersichtlich.

Parameter	% ⇑ na zu a	% ⇓ na zu a	F-Wert	p-level
D_{rmse}	75	25	$F = 27.31$	$p < 0.001$
D_{corr}	70	30	$F = 14.59$	$p < 0.001$
D_{dct}	80	20	$F = 12.45$	$p < 0.001$

Tabelle 6.2: *Auswertungsergebnisse zu den RMS-Kontur-Distanzwerten mit Prozentwerten der Sprecher, bei welchen die a-Distanz größer ist als die na-Distanz (% ⇑ na zu a) bzw. kleiner (% ⇓ na zu a).*

6.2.3 Distanzwerte - Diskussion

Im Allgemeinen bestätigen sowohl die Ergebnisse zu den F0- als auch zu den RMS-Konturen statistisch die Hypothese, dass sich Konturen von a-Sprache und Konturen von na-Sprache physisch unterscheiden. $D_{rmse}(a)$ ist im statistischen Mittel größer als $D_{rmse}(na)$, demnach sind sich die cna- und na-Konturen physisch gesehen ähnlicher als die a- und na-Konturen. In beiden Fällen sind die Auf- und Abbewegungen von Konturen der Sprachaufnahmen nüchterner Sprecher und Konturen der Kontrollgruppenaufnahmen synchroner als jene der unter Alkoholeinfluss und in nüchternem Zustand geäußerten Sprachaufnahmen $\big(D_{corr}(\mathrm{a}) > D_{corr}(\mathrm{na})\big)$. Auch die Euklidische Distanz im 6-dimensionalen DCT Raum, D_{dct}, ist bei den F0- sowie den RMS-Konturen größer zwischen a- und na-Konturen als zwischen cna- und na-Konturen $\big(D_{dct}(\mathrm{a}) > D_{dct}(\mathrm{na})\big)$.
Auf Grund der Notwendigkeit von Referenzdaten konnten die Abstandsmaße nur für die 20 Sprecher mit Kontrolldaten berechnet werden. Es wurde gezeigt, dass die Abstandsmaße zwischen a- und na-Konturen sprecherabhängig größere aber auch kleinere Werte gegenüber der Referenzdistanz annehmen können und damit kein allgemeingültiges sprecherunabhängiges Modell für das Verhalten der Abstandsmaße bei Sprache unter Alkoholeinfluss gebildet werden kann. Deshalb, und weil Referenzdaten im Allgemeinen nicht zur Verfügung stehen, ist eine solche Technik in der Praxis, d.h. zur automatischen Detektion von Alkoholisierung, nur bedingt anwendbar. Trotzdem könnten bei der Benutzung einer sprachgesteuerten Applikation wie einem Navigationssystem wiederholt geäußerte Befehle eines Individuums aufgezeichnet und mit Hilfe der Abstandsmaße ein Modell für deren Verhalten bei Sprache ohne Alkoholeinfluss erstellt werden (es wird davon ausgegangen, dass die zur Erstellung des Modells verwendeten Sprachbefehle in nüchternem Zustand des Sprechers geäußert wurden), das dann zur Bewertung einer aktuellen sprachlichen Äußerung, ebenfalls mit Hilfe einer Abstandsberechnung, herangezogen werden kann.

Zusätzlich stellt sich die Frage, welche konkreten Änderungen an der Grundfrequenz- bzw. Energiekontur, wie z.B. ein globales Gefälle, die signifikanten Unterschiede bei den Abstandsmaßen hervorrufen. Zur Untersuchung dieser Fragestellung wurde im folgenden Kapitel 6.3 vor der vergleichenden Analyse eine Parametrisierung der Konturen vorgenommen. Änderungen der Parameter können gegebenenfalls als Änderungen von Basisformstrukturen interpretiert werden.

6.3 Konturen - Parametrisierung

Durch eine Parametrisierung ist es möglich, potentielle Merkmale für jede beliebige Kontur zu ermitteln, diese Merkmale bezüglich des Faktors Alkoholisierung zu testen und eine mögliche Korrelation zu den gemessenen BAK Werten aufzudecken. Dabei muss bei der Parametrisierung keine Relation wie bei den Distanzwerten berechnet werden. Die Kontrolldaten werden deshalb nicht als Referenzen benötigt. Aus diesem Grund kann eine Parametrisierung für die 150 Sprecher des ALC mit einer Blutalkoholkonzentration über $0.4\overline{9}‰$ (im alkoholisierten Zustand) durchgeführt werden.

6.3.1 Parametrisierung - Methode

Die DCT Koeffizienten $\Psi(\nu)$ mit $\nu = 2 \dots 7$ wurden wie in Kapitel 6.2.1 beschrieben für jede na- und a-F0- und RMS-Kontur als Merkmale berechnet (150 Sprecher, 19 gelesene Äußerungen). Weiterhin wurden durch erneute Parametrisierung des DCT Spektrums das erste (m_1) und zweite Moment der DCT Koeffizienten (m_2) berechnet. Die ersten beiden Momente beschreiben grundlegende Eigenschaften der Form des DCT Spektrums, m_1 den Schwerpunkt und m_2 die Varianz. Die statistischen Momente $m_{1,2}$ werden wie folgt berechnet, wobei $m_0 = 0$ und $k = 1, 2$ (siehe zum Beispiel Harrington [2010], S. 298):

$$m_k = \frac{\sum_\nu |\Psi(\nu - m_{k-1})|^k}{\sum_\nu \Psi(\nu)} \tag{6.5}$$

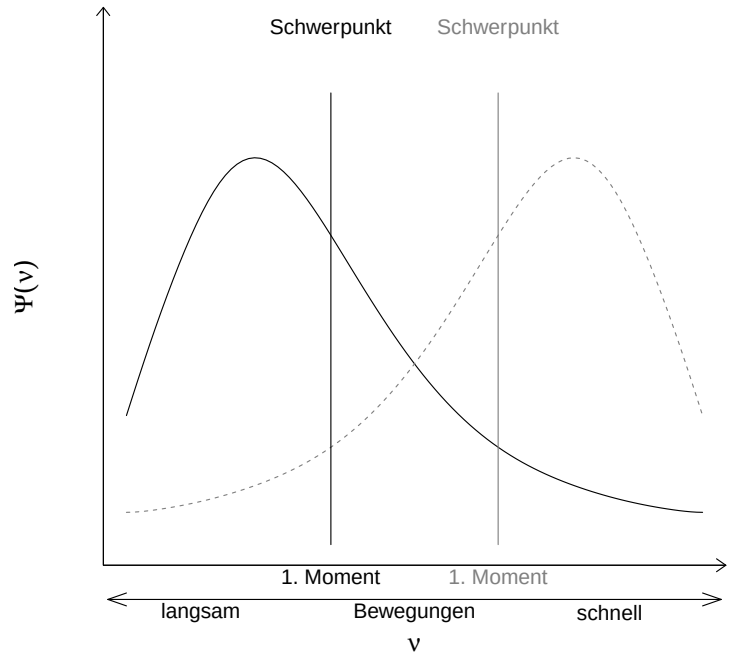


Abbildung 6.5: *Schematische Darstellung des Effekts einer Änderung des Schwerpunkts (1. Moment) im DCT Spektrum der Konturbewegungen.*

Ein niedriger Wert des ersten Moments m_1, der einen niedrigen Schwerpunkt des DCT Spektrums bezeichnet, wird für Konturen mit dominanten Langzeitbewegungen und weniger Kurzzeitbewegungen erwartet. Ein größerer Wert von m_1 deutet auf eine dynamischere Kontur hin (siehe Abbildung 6.5). Das zweite Moment m_2 wird durch die Energiekonzentration (Varianz) im DCT Spektrum bestimmt. Konturen mit einer gleichmäßigeren Form, zum Beispiel einer gleichmäßigen Abfolge von Maxima, weisen eine geringere Varianz im DCT Spektrum und damit ein niedrigeres zweites Moment m_2 auf, wohingegen unregelmäßigere Konturen ein höheres zweites Moment m_2 zur Folge haben. Zur Berechnung von $m_{1,2}$ wurde das DCT Spektrum über die Frequenzindizes $\nu = 2 \ldots 51$ verwendet. Damit ist die kleinste in Betracht gezogene Wellenlänge der Bewegung in den Konturen:

$$\frac{4}{51}L = 0.078L \tag{6.6}$$

wobei L die Dauer der Aufnahme in Sekunden ist. Da die durchschnittliche Silbenanzahl der Testäußerungen bei 12.3 liegt, was einer Wellenlänge von $0.081L$ entspricht, sollten die DCT Koeffizienten $\Psi(\nu)$ mit $\nu = 2 \ldots 51$ ausreichend sein, um alle Konturbewegungen bis hin zur Silbenrate zu erfassen.

6.3.2 Parametrisierung - Ergebnisse und Diskussion

Berechnet wurden die DCT Koeffizienten $\Psi(\nu)$ mit $\nu = 2 \ldots 7$ sowie das 1. und 2. Moment des DCT Spektrums $m_{1,2}$ aller na- und a-F0- und RMS-Konturen der 19 relevanten, gelesenen Äußerungen (siehe Anhang C) von 150 Sprechern (BAK $\geq 0.5‰$).

6.3.2.1 F0

Signifikanztests mit Hilfe von MEM Analyse (Baayen [2008]), den 'fixed factors' Alkoholisierung und Geschlecht sowie den 'random factors' Aufnahmeelement und Sprecher ergaben für die DCT Koeffizienten $\Psi(\nu)$ mit $\nu = 2 \ldots 7$ der F0-Konturen keinen signifikanten Effekt für die Alkoholisierung (siehe Tabelle 6.3). Alkoholisierung bezeichnet hier wieder den Zustand des Sprechers.

Parameter	% ⇑ na zu a	% ⇓ na zu a	F-Wert	p-level
$\Psi(\nu = 2)$	56	44	$F = 1.27$	n.s.
$\Psi(\nu = 3)$	54.67	45.33	$F = 0.34$	n.s.
$\Psi(\nu = 4)$	48	52	$F = 0.71$	n.s.
$\Psi(\nu = 5)$	44	56	$F = 3.20$	n.s.
$\Psi(\nu = 6)$	56	44	$F = 0.02$	n.s.
$\Psi(\nu = 7)$	45.33	54.67	$F = 2.03$	n.s.
m_1	64	36	$F = 22.05$	$p < 0.001$
m_2	55.33	44.67	$F = 5.13$	$p < 0.05$

Tabelle 6.3: *Auswertungsergebnisse zu den DCT Koeffizienten $\Psi(\nu)$ mit $\nu = 2 \ldots 7$ und Momenten des DCT Spektrums m_1 und m_2 bei F0 mit Prozentwerten der Sprecher, bei welchen sich der jeweilige Parameter bei a-Sprache gegenüber na-Sprache erhöht (% ⇑ na zu a) bzw. verringert (% ⇓ na zu a).*

Die festgestellte globale Veränderung der Distanz im 6 dimensionalen DCT Raum (siehe Kapitel 6.2.2) kann deshalb keinen speziellen DCT Frequenzen und damit keiner Veränderung in der Gesamtform der Konturen zugeordnet werden. Für das erste Moment m_1 ergibt sich ein signifikanter Anstieg ($p < 0.001$) bezüglich des Faktors Alkoholisierung. Ein Anstieg im ersten Moment m_1, und damit eine Verlagerung des Schwerpunkts im DCT Spektrum hin zu größeren Werten, deutet auf schnellere Bewegungen der a-Grundfrequenzkonturen gegenüber den na-Konturen hin. Beim zweiten Moment m_2 weist eine nur schwache Erhöhung ($p < 0.05$) von na- zu a-Sprache auf geringfügig unregelmäßigere Grundfrequenzkonturen bei Sprache unter Alkoholeinfluss hin. Korrelationen der Änderungen der DCT Koeffizienten $\Psi(\nu)$ mit $\nu = 2 \ldots 7$ und Momente $m_{1,2}$ mit dem Blutalkoholwert ergeben einen maximalen Korrelationskoeffizienten von $r = 0.11$ (nicht signifikant) für das erste Moment m_1. Daher lässt im Falle der F0-Konturen der Grad der Veränderung der DCT Koeffizienten $\Psi(\nu)$ mit $\nu = 2 \ldots 7$ und Momente $m_{1,2}$ keinen Rückschluss auf den Grad der Alkoholisierung zu. Korrelationen innerhalb der Daten eines Sprechers sind auch hier nicht möglich, da pro Sprecher und Aufnahme nur eine Messung der Blutalkoholkonzentration vorgenommen wurde.

6.3.2.2 RMS

Eine MEM Analyse (Baayen [2008]) mit den 'fixed factors' Alkoholisierung und Geschlecht sowie den 'random factors' Aufnahmeelement und Sprecher wurde analog zur Analyse der Grundfrequenzkonturen auch für die RMS-Konturen durchgeführt. Für die DCT Koeffizienten $\Psi(\nu)$ mit $\nu = 2, 4, 5, 6$ ergeben sich hinsichtlich des Faktors Alkoholisierung signifikante Unterschiede (siehe Tabelle 6.4), insbesondere für die DCT Koeffizienten $\Psi(\nu)$ mit $\nu = 2$ und $\nu = 4$ (beide sinken). Der DCT Koeffizient $\Psi(\nu = 2)$ spiegelt die Neigung der Kontur wider (Harrington [2010]), $\Psi(\nu = 4)$ die Schiefe. Die Ergebnisse deuten auf RMS-Konturen mit geringerer globaler Neigung und geringerer Schiefe bei

Parameter	% ⇑ na zu a	% ⇓ na zu a	F-Wert	p-level
$\Psi(\nu = 2)$	40.67	59.33	$F = 9.13$	$p < 0.01$
$\Psi(\nu = 3)$	53.33	46.67	$F = 0.06$	n.s.
$\Psi(\nu = 4)$	37.33	62.67	$F = 25.00$	$p < 0.001$
$\Psi(\nu = 5)$	62	38	$F = 5.76$	$p < 0.05$
$\Psi(\nu = 6)$	40.67	59.33	$F = 5.12$	$p < 0.05$
$\Psi(\nu = 7)$	55.33	44.67	$F = 1.28$	n.s.
m_1	40	60	$F = 6.80$	$p < 0.05$
m_2	56	44	$F = 9.87$	$p < 0.01$

Tabelle 6.4: *Auswertungsergebnisse zu den DCT Koeffizienten $\Psi(\nu)$ mit $\nu = 2 \ldots 7$ und Momenten des DCT Spektrums m_1 und m_2 bei RMS mit Prozentwerten der Sprecher, bei welchen sich der jeweilige Parameter bei a-Sprache gegenüber na-Sprache erhöht (% ⇑ na zu a) bzw. verringert (% ⇓ na zu a).*

Sprachsignalen unter Alkoholeinfluss hin. Abbildung 6.6 zeigt die absoluten Änderungen beider Koeffizienten und den Momenten $m_{1,2}$ von na- zu a-Sprache aller 150 Sprecher. Für das erste Moment m_1 (siehe auch Abbildung 6.6) liefert der Test eine schwach signifikante Verminderung ($p < 0.05$) von na- zu a-Sprache beziehungsweise eine Verschiebung des Schwerpunkts im DCT Spektrum nach links hin zu kleineren Werten, was darauf schließen lässt, dass RMS-Konturen von Sprachaufnahmen unter Alkoholeinfluss aus mehr langsamen Bewegungen zu bestehen scheinen als Konturen von in nüchternem Zustand getätigten Sprachaufnahmen. Dieses Ergebnis unterstützt die Aussagen von Behne et al. [1991], Hollien et al. [1999, 2001], Künzel and Braun [2003] und Sobell et al. [1982] sowie die Ergebnisse zur Sprechgeschwindigkeit aus Kapitel 5, wonach ein Großteil der Sprecher die Sprechgeschwindigkeit im alkoholisierten Zustand reduziert. Die gefundene signifikante Erhöhung des 2. Moments m_2 ($p < 0.01$) von na-Sprache hin zu a-Sprache (siehe auch Abbildung 6.6) zeigt, dass Konturen von Sprachaufnahmen unter Alkoholeinfluss signifikant unregelmäßiger sind als Konturen von Sprachaufnahmen ohne Alkoholeinfluss. Dies bestätigt wiederum die Ergebnisse zu den RMS Rhythmizitäts-Parametern (Kapitel 4.2 und Tabelle 4.1), die eine Veränderung vieler Parameter anzeigen und damit ebenfalls auf

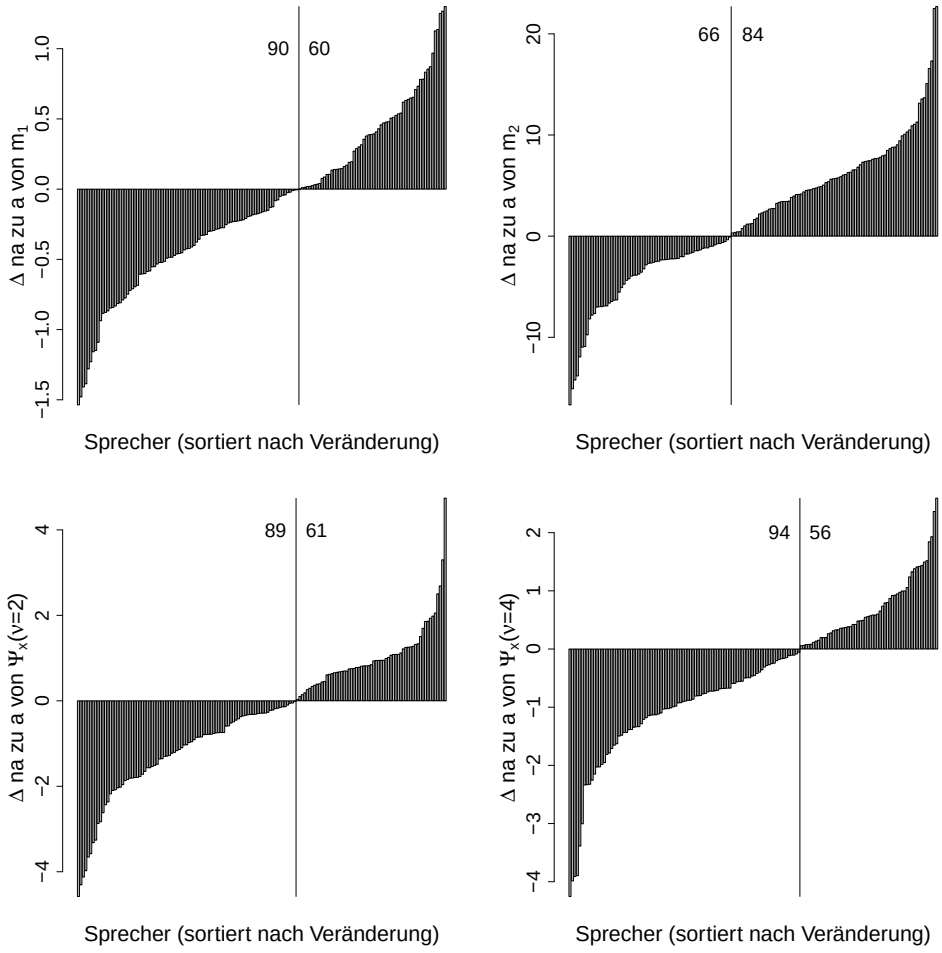


Abbildung 6.6: *Änderungen der Momente des DCT Spektrums $m_{1,2}$ und DCT Koeffizienten $\Psi(\nu)$ mit $\nu = 2, 4$ (RMS) von na- zu a-Sprache aller 150 Sprecher.*

unregelmäßigere Konturen der Sprachaufnahmen unter Alkoholeinfluss hinweisen.

Korrelationen der Änderungen der DCT Koeffizienten $\Psi(\nu)$ mit $\nu = 2 \ldots 7$ und Momente $m_{1,2}$ mit dem Blutalkoholwert ergeben einen maximalen Korrelationskoeffizienten von $r = -0.15$ (nicht signifikant) für das erste Moment m_1. Daher lässt auch im Falle der RMS-Konturen der Grad der Veränderung der DCT Koeffizienten $\Psi(\nu)$ mit $\nu = 2 \ldots 7$ und Momente $m_{1,2}$ keinen Rückschluss auf den Grad der Alkoholisierung zu. Korrelationen innerhalb der Daten eines Sprechers sind auch hier nicht möglich, da pro Sprecher und Aufnahme nur eine Messung der Blutalkoholkonzentration vorgenommen wurde.

6.3.2.3 Kontrollgruppenversuche Parametrisierung - Ergebnisse F0 und RMS

Kontrollgruppenversuche wurden auch im Falle der parametrisierten Daten vorgenommen, um auszuschließen, dass es sich bei den gefundenen Effekten um Einflüsse verborgener Faktoren handelt. Die im Hauptexperiment gefundenen Effekte konnten hier nicht nachgewiesen werden, vermutlich auf Grund der geringen Sprecherzahl der Kontrollgruppe. Der Vollständigkeit halber seien die Ergebnisse aber trotzdem aufgeführt, jedoch ohne Erklärungsansatz.
Berechnet wurden die DCT Koeffizienten $\Psi(\nu)$ mit $\nu = 2 \dots 7$ sowie das 1. und 2. Moment des DCT Spektrums $m_{1,2}$ der na-, a- und cna-F0- und RMS-Konturen der 19 relevanten, gelesenen Äußerungen (siehe Anhang C) der 20 Kontrollgruppensprecher. Die Auswertung fand ebenfalls durch MEM Analyse (Baayen [2008]) mit den 'fixed factors' Alkoholisierung und Geschlecht sowie den 'random factors' Aufnahmeelement und Sprecher statt.

Für die DCT Koeffizienten $\Psi(\nu)$ mit $\nu = 2 \dots 7$ der F0-Konturen ergeben sich weder zwischen a- und na-Sprache noch zwischen cna- und na-Sprache signifikante Unterschiede. Betrachtet man das 1. Moment m_1, so zeigt sich ein signifikanter Unterschied zwischen a- und na-Konturen ($F = 7.6495$, $p < 0.01$), nicht aber zwischen na- und cna-Konturen. Die Unterschiede für das 2. Moment m_2 sind in beiden Fällen nicht signifikant.

Im Falle der RMS-Konturen sind die Unterschiede zwischen a- und na-Sprache für die DCT Koeffizienten $\Psi(\nu)$ mit $\nu = 4, 5$ signifikant, für die restlichen DCT Koeffizienten und die Momente zeigt sich jedoch kein signifikanter Effekt. Zwischen den na- und cna-Konturen ergibt sich ein signifikanter Unterschied für den DCT Koeffizienten $\Psi(\nu = 7)$.

6.4 Funktionale Datenanalyse von F0- und RMS-Konturen

Die bisher vorgestellten Analysen und Verfahren reduzieren die Konturen und damit deren Informationsgehalt auf nur wenige Werte. Die eigentliche Form der Kontur findet dabei keine beziehungsweise nur wenig Beachtung. Einzig die Parametrisierung gibt einen Hinweis auf vorhandene Grundformen, jedoch lässt sich damit nicht sagen, ob diese Grundformen für die vorliegenden Daten optimal und phonetisch bzw. prosodisch interpretierbar sind. Elementare prosodische Merkmale gelesener Sprache wie fallende Grundfrequenzverläufe (F0) oder ein periodischer Silbenrhythmus (RMS) zeigen sich erst beim Betrachten der Konturform in ihrer Gesamtheit. Für die im Folgenden beschriebene Analyse diente deshalb die gesamte Kontur als Grundlage.

Die funktionale Datenanalyse (siehe z.B. Ramsay et al. [2009]) stellt eine Anzahl von statistischen Werkzeugen zur Verfügung, mit deren Hilfe es möglich ist, Konturen oder im Allgemeinen Kurven anstelle von Vektoren aus Parametern zu analysieren. Um die eigentliche FDA anwenden zu können, müssen die Rohdaten, die auf Grund ihrer Digitalisierung lediglich zeitdiskrete Abtastwerte enthalten, vorverarbeitet und damit in eine funktionale Form überführt werden. Im Rahmen dieser Arbeit wurden die so erstellten funktionalen Daten mit Hilfe der funktionalen Erweiterung der Hauptkomponentenanalyse (*functional Principal Components Analysis* bzw. fPCA) ausgewertet. Die FDA bei Untersuchungen in der Phonetik wurde von Gubian und anderen (Cheng and Gubian [2011], Gubian et al. [2009, 2010, 2011], Turco and Gubian [2012], Turco et al. [2011], Zellers et al. [2010]) eingeführt. Im Normalfall wurde dabei phonetisches Datenmaterial (zeitdiskretes Kontursignal) in funktionale Daten überführt, um daraufhin eine Analyse der Daten mit Hilfe einer fPCA durchführen zu können. Die FDA stellt weitaus mehr Werkzeuge zur Verfügung als das der fPCA (eine detaillierte Beschreibung findet sich in Ramsay et al. [2009]), ihr Potential ist damit

sehr umfangreich. Hinsichtlich der F0- und RMS-Konturen gelesener Äußerungen des ALC kam jedoch nur die funktionale Hauptkomponentenanalyse zur Anwendung. Die Diversität der Konturen der 19 gelesenen Äußerungen (siehe Anhang C) spricht eigentlich gegen die Verwendung der FDA bzw. fPCA bei einer Analyse aller Konturen gemeinsam, da die FDA im Grunde nur für gleichartige Konturen (z.B. Temperaturkurven mehrerer Jahre, siehe Ramsay et al. [2009]) beschrieben und empfohlen wird. Weil sich mit ihr aber auf funktionaler Ebene eine Hauptkomponentenanalyse durchführen lässt und es das Ziel dieser Analyse war, fundamentale Formen (wie z.B. ein globales Gefälle) in den Konturen von Sprachdaten, die von Sprechern in alkoholisiertem und nüchternem Zustand erhoben wurden, zu finden, wurde eine funktionale Datenanalyse dennoch vorgenommen.

6.4.1 FDA - Methode

Für die FDA wurden die Konturen der 19 relevanten, gelesenen Äußerungen (siehe Anhang C) der 150 Sprecher des ALC, die eine Blutalkoholkonzentration von mehr als $0.4\overline{9}‰$ (im alkoholisierten Zustand) aufweisen, durch Abtastratenkonvertierung auf $N = 200$ Abtastwerte abgebildet, da die Konturen alle miteinander analysiert wurden und die FDA Konturen mit gleicher Länge voraussetzt. Zur Umwandlung der zeitdiskreten Rohdaten in ihre funktionale Form wird zunächst ein passendes System aus Basisfunktionen benötigt. Ramsay et al. [2009] bezeichnet dieses System als Basissystem. Im Gegensatz zu zeitdiskreten Daten können funktionale Daten beziehungsweise Konturen zu jedem Zeitpunkt evaluiert werden. Um eine detailgetreue funktionale Repräsentation vor allem für komplexe Kurven zu erhalten, eignet sich am besten eine B-Spline Basis oder eine Fourier Basis. Hier ist es möglich, die Anzahl an Basisfunktionen hoch genug zu wählen, um auch kleine Bewegungen in den Konturen zu modellieren. Reduziert man die Anzahl an Basisfunktionen, so resultiert dies in einer Glättung der Konturen. Oftmals findet dieses Prozedere in der FDA Anwendung, um Störungen im Signal herauszufiltern. Um verschiedene Stufen der Glättung miteinzubeziehen, wurden

mehrere Anzahlen von Basisfunktionen (5, 11, 21, 51, 75, 101, 151, 197) für beide Basissysteme (B-Spline und Fourier) getestet. Durch die Vorgaben der von der FDA zur Verfügung gestellten Funktionen in R war die maximal mögliche Anzahl von Basisfunktionen 197 für beide Systeme bei einer Anzahl von $N = 200$ Werten pro Kontur sowie eine minimal mögliche Anzahl von 5 Basisfunktionen. Eine ungerade Anzahl ist deswegen notwendig, da das Fourier Basissystem eine ungerade Anzahl an Basisfunktionen voraussetzt. Weiterhin war eine Modellierung mit mehr als 197 beziehungsweise weniger als 5 Basisfunktionen nicht für beide Systeme gleichzeitig realisierbar. Da sich der Glättungsgrad unterhalb einer Grenze von 101 Basisfunktionen deutlicher bemerkbar macht als darüber, wurden mehr Stufen zwischen 5 und 101 Basisfunktionen als zwischen 101 und 197 Basisfunktionen integriert. Beim B-Spline Basissystem ist zu beachten, dass eine zu hohe Anzahl an Basisfunktionen an Anfang und Ende der Kontur ungenaue Näherungen an die Rohdaten bei der Modellierung verursacht, weil die Werte der Spline Funktionen an den Rändern nur durch einen einzigen Koeffizienten bestimmt werden (Ramsay et al. [2009]). Deshalb wurden im Falle des B-Spline Basissystems die funktionalen Daten ausgeschlossen, die auf 151 und 197 Basisfunktionen beruhen. Diese Problematik ergibt sich beim Fourier Basissystem nicht, da es periodisch ist und zur Bestimmung der Werte die Daten wie in einer Schleife behandelt. Damit sind auch an den Grenzen immer genug Daten zur Modellierung vorhanden (Ramsay et al. [2009]). Aus diesen Gründen belaufen sich die Anzahlen der Basisfunktionen im Rahmen dieser Untersuchung auf 5 - 197 für das Fourier Basissystem sowie 5 - 101 für das B-Spline Basissystem. Die erzeugten funktionalen Repräsentationen der Konturdaten werden in sogenannten funktionalen Datenobjekten abgelegt.

Unter den Werkzeugen der FDA befinden sich auch solche, die ein sogenanntes 'registering' der Daten erlauben. Dieses richtet die Konturen anhand von Ankerpunkten durch nicht lineare Alignierung aufeinander aus und macht die Daten für viele Anwendungsbereiche interpretierbarer. Soll beispielsweise ein betonungsbe-

dingtes lokales Maximum in der Grundfrequenz innerhalb eines Wortes (das Maximum stellt damit einen Ankerpunkt dar) untersucht werden, wobei viele Wiederholungen desselben Wortes unterschiedlicher Länge vorhanden sind, so findet das 'registering' durch die Alignierung in den Konturen der einzelnen Nennungen jeweils dieses lokale Maximum und passt sie zeitlich so an, dass die Maxima der verschiedenen Nennungen möglichst zusammenfallen. Ein 'registering' der Daten ist bei der FDA im Normalfall sinnvoll und gegebenenfalls auch notwendig. Da dieses Verfahren aber nicht linear arbeitet und darüber hinaus im Falle des ALC die gelesenen Äußerungen in ihrer Länge und ihrem Inhalt verschieden sind, was eine Bestimmung von übereinstimmenden Ankerpunkten in den verschiedenen Äußerungen unmöglich macht, wurde hier auf 'registering' verzichtet.
Daraufhin wurde auf die funktionalen B-Spline und Fourier Datenobjekte aller Konturen fPCA angewendet. Wie bei einer normalen Hauptkomponentenanalyse transformiert die fPCA die analysierten Daten in einen neuen Raum, der bei der fPCA durch orthogonale 'Eigenfunktionen' (bei der normalen PCA 'Eigenvektoren') bzw. fPCs definiert ist. Die transformierten Daten ermöglichen eine Näherung an die ursprünglichen Konturen durch eine kleinere Anzahl von Linearkombinationen der funktionalen Hauptkomponenten. Dabei erklärt die erste Hauptkomponente den Großteil der Varianz im transformierten Raum, die zweite Hauptkomponente den zweitgrößten Anteil der Varianz usw. Hinsichtlich der F0- und RMS-Konturen des ALC wurde erwartet, dass die funktionalen Hauptkomponenten elementare prosodische Merkmale gelesener Sprache wie fallende Grundfrequenzverläufe (F0) oder einen periodischen Silbenrhythmus (RMS) widerspiegeln.

Jede ursprüngliche Kontur $x_i(t)$ kann durch eine Näherung aus der Summe von C funktionalen Hauptkomponenten $\phi_c(t)$, deren Gewichtungen (Scores) $\delta_{c,i}$, wobei $C <= T$ ist, und der durchschnittlichen Kontur $\mu(t)$, beschrieben werden.

$$x_i(t) \approx \mu(t) + \sum_{c=1}^{C} \left(\phi_c(t) \cdot \delta_{c,i} \right) \tag{6.7}$$

Die fPC Scores $\delta_{c,i}$ können als Merkmale der ursprünglichen Konturen betrachtet werden, weil sie, genau wie die DCT Koeffizienten, Anteile von zugrundeliegenden Funktionen an der zu rekonstruierenden Kontur repräsentieren (sie können dabei positive und negative Werte annehmen). Bei der DCT sind es die Anteile an Cosinusfunktionen, bei der fPCA die Anteile der jeweiligen Hauptkomponenten. Ein hoher (Absolut-)Wert des fPC Scores $\delta_{c,i}$ spricht der jeweiligen funktionalen Hauptkomponente ϕ_c bei der Rekonstruktion der ursprünglichen Kontur eine höhere Gewichtung zu. Die erste funktionale Hauptkomponente ϕ_1 erklärt idealerweise den Großteil der Varianz in den Daten und repräsentiert elementare Konturformen. Falls ein Teil dieser Varianz auf dem Unterschied zwischen alkoholisiert und nüchtern beruht, oder mit anderen Worten, falls sich Konturen von Sprachaufnahmen unter Alkoholeinfluss und ohne Alkoholeinfluss in den zugehörigen Scores unterscheiden, dann sollten die Scores, die diese funktionalen Hauptkomponenten bzw. elementaren Formen repräsentieren, als potentielle Merkmale zur Kennzeichnung des Faktors Alkoholisierung in den Konturdaten betrachtet werden. Somit ist es möglich, sie als abhängige Variablen einer MEM Analyse zu unterziehen, um den auf sie ausgeübten Einfluss von Alkoholisierung und anderen Faktoren untersuchen zu können.

6.4.2 FDA - Ergebnisse

Jedes funktionale Datenobjekt enthält die Konturen der 19 relevanten, gelesenen Äußerungen (siehe Anhang C) aller 150 Sprecher (BAK $\geq 0.5‰$), sowohl im alkoholisierten als auch im nüchternen Zustand, also $19 \times 150 \times 2 = 5700$ Konturen. Diese wurden alle zusammen mit Hilfe der fPCA auf ihre Gruppierungstauglichkeit hin analysiert, da nach allgemein vorhandenen grundlegenden Mustern in a- und na-Konturen gesucht wurde.

Die funktionalen Hauptkomponenten $\phi_c, c = 1 \ldots 9$ und die zugehörigen Scores $\delta_{c,i},\ c = 1 \ldots 9$ wurden einerseits für alle F0-Konturen und andererseits für alle RMS-Konturen (jeweils a und na zusammen) und die oben genannten Anzahlen

von Basisfunktionen (Fourier: 5-197, B-Splines: 5-101) berechnet. Die Darstellung der Ergebnisse beschränkt sich größtenteils auf die Datenobjekte, welche die Rohdaten detailliert widerspiegeln (197 Basisfunktionen) sowie die Datenobjekte, deren Konturdaten einen hohen Glättungsgrad aufweisen (11 Basisfunktionen).

6.4.2.1 F0

Die ersten 9 funktionalen Hauptkomponenten des funktionalen F0-Datenobjekts basierend auf 197 Fourier Basisfunktionen sind aus Abbildung 6.7 ersichtlich.

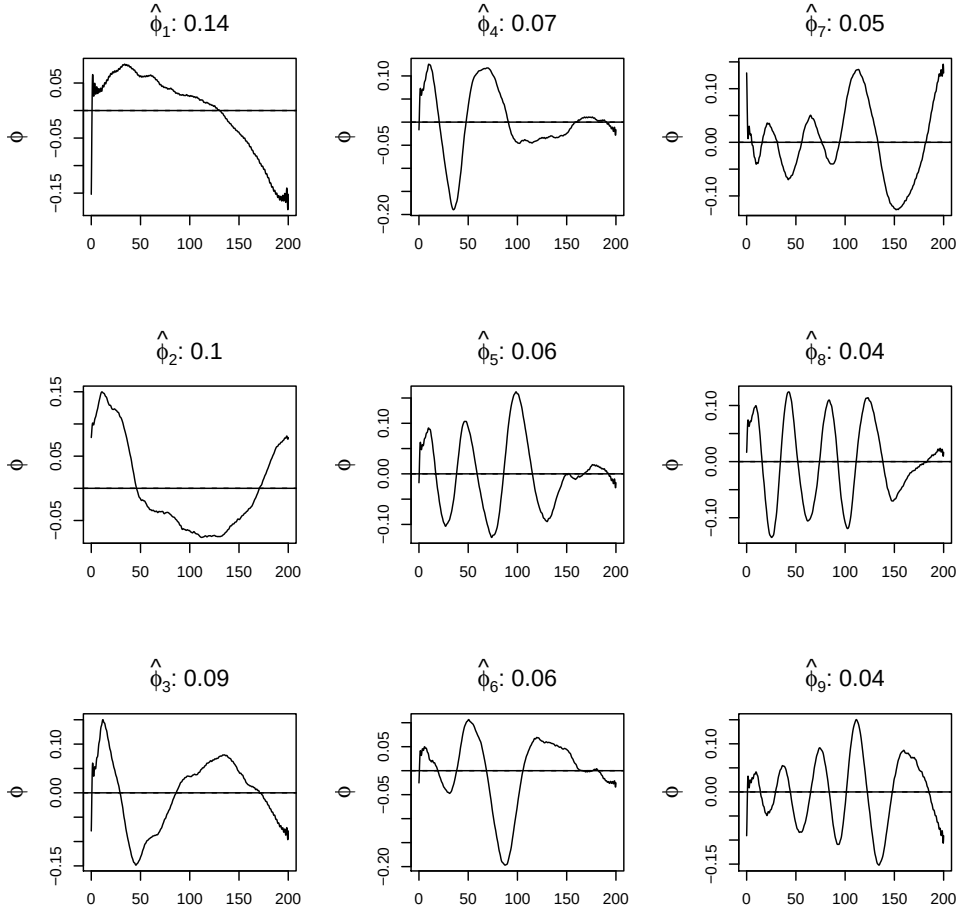


Abbildung 6.7: *Funktionale Hauptkomponenten 1 bis 9 des funktionalen F0-Datenobjekts basierend auf 197 Fourier Basisfunktionen.*

Den größten Anteil erklärter Varianz trägt hierbei die erste funktionale Hauptkomponente mit $\hat{\Phi}_1 = 0.14$, kann aber hinsichtlich einer Gruppierung der Da-

ten trotzdem nicht als allzu aussagekräftig bewertet werden. Das heißt, dass die erste fPC ϕ_1 zwar eine in den Konturdaten vorhandene elementare Grundform widerspiegelt, aber zu selten vorkommt, um eine hinsichtlich der Alkoholisierung interpretierbare Gruppierung zu ermöglichen. Dieselbe Aussage lässt sich auch für die fPCs $\phi_c, c = 2 \ldots 9$ treffen, deren Anteile erklärter Varianz bei nur $\hat{\Phi}_c = 0.1, 0.09, 0.07, 0.06, 0.06, 0.05, 0.04, 0.04$ liegen. Wird die Anzahl der Fourier Basisfunktionen auf 11 reduziert, werden die Konturen zwangsläufig geglättet und auch ähnlicher. Dabei verbessert sich das Bild der fPCA nicht erheblich. Abbildung 6.8 zeigt die ersten neun funktionalen Hauptkomponenten für das funktionale F0-Datenobjekt basierend auf 11 Fourier Basisfunktionen.

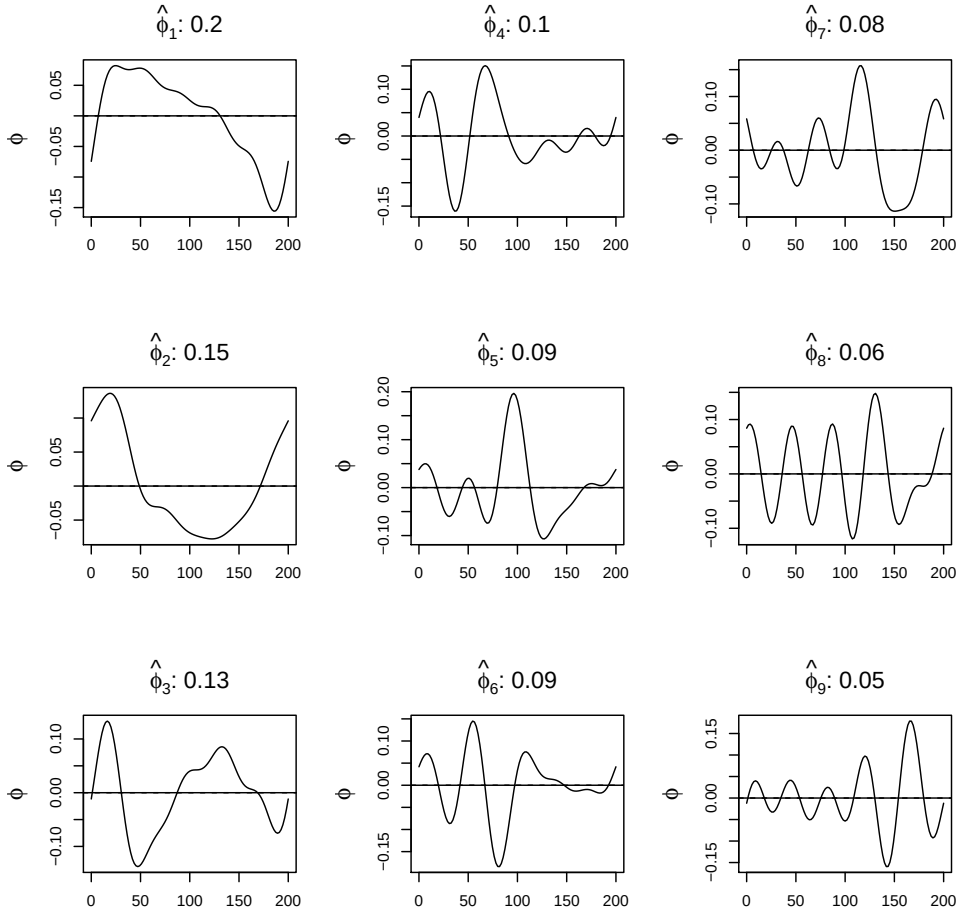


Abbildung 6.8: *Funktionale Hauptkomponenten 1 bis 9 des funktionalen F0-Datenobjekts basierend auf 11 Fourier Basisfunktionen.*

Der Anteil der erklärten Varianz liegt hier für die erste funktionale Hauptkomponente bei $\hat{\Phi}_1 = 0.2$ und ist damit ebenfalls sehr niedrig. Dennoch ergeben sich in beiden Fällen interessante Formen vor allem der ersten drei Hauptkomponenten. So könnten die erste Hauptkomponente ein globales Gefälle und die zweite Hauptkomponente die Intonationskurve einer Frage als vorhandene elementare Grundformen der F0-Konturen widerspiegeln, die dritte Hauptkomponente eventuell wichtige Akzentpositionen der Konturen markieren. Da die Anteile erklärter Varianz aber so gering sind, ist ihre Bedeutung im Kontext der gesammelten Konturen aller 19 relevanten, gelesenen Äußerungen des ALC (siehe Anhang C) eher marginal. Ähnliche fPCs wie bei den funktionalen Datenobjekten basierend auf Fourier Basisfunktionen ergeben sich für alle auf B-Splines basierenden funktionalen Datenobjekte. Der Anteil erklärter Varianz für die erste funktionale Hauptkomponente des funktionalen F0-Datenobjekts basierend auf 11 B-Spline Basisfunktionen liegt bei $\hat{\Phi}_1 = 0.22$.

Die fPC Scores $\delta_{c,i}$, $c = 1 \ldots 9$ wurden mit MEM Analyse (Baayen [2008]) und den 'fixed factors' Alkoholisierung und Geschlecht sowie den 'random factors' Aufnahmeelement und Sprecher getestet. Für die Scores der ersten fPC $\delta_{c,i}$, $c = 1$ ergibt sich für alle Anzahlen von Basisfunktionen und beide Basissysteme ein hochsignifikanter Effekt für die Alkoholisierung ($F > 18.3$, $p < 0.001$). Die Scores der fPCs $\delta_{c,i}$, $c = 2 \ldots 9$ unterscheiden sich nicht signifikant zwischen a- und na-Konturen. Es wurde keine Interaktion mit dem Geschlecht festgestellt.

6.4.2.2 RMS

Abbildung 6.9 zeigt die ersten 9 funktionalen Hauptkomponenten des funktionalen RMS-Datenobjekts basierend auf 197 Fourier Basisfunktionen. Die geringen Anteile erklärter Varianz $\hat{\Phi}_c = 0.06, 0.05, 0.05, 0.04, 0.04, 0.04, 0.04, 0.03, 0.03$ lassen nicht leicht zu gruppierende Daten vermuten. Die Hypothese, dass die fPCA elementare Konturmuster findet, die einen Großteil der durch Alkoholisierung verursachten Varianz erklären, muss damit abgelehnt werden. Wird die Anzahl der

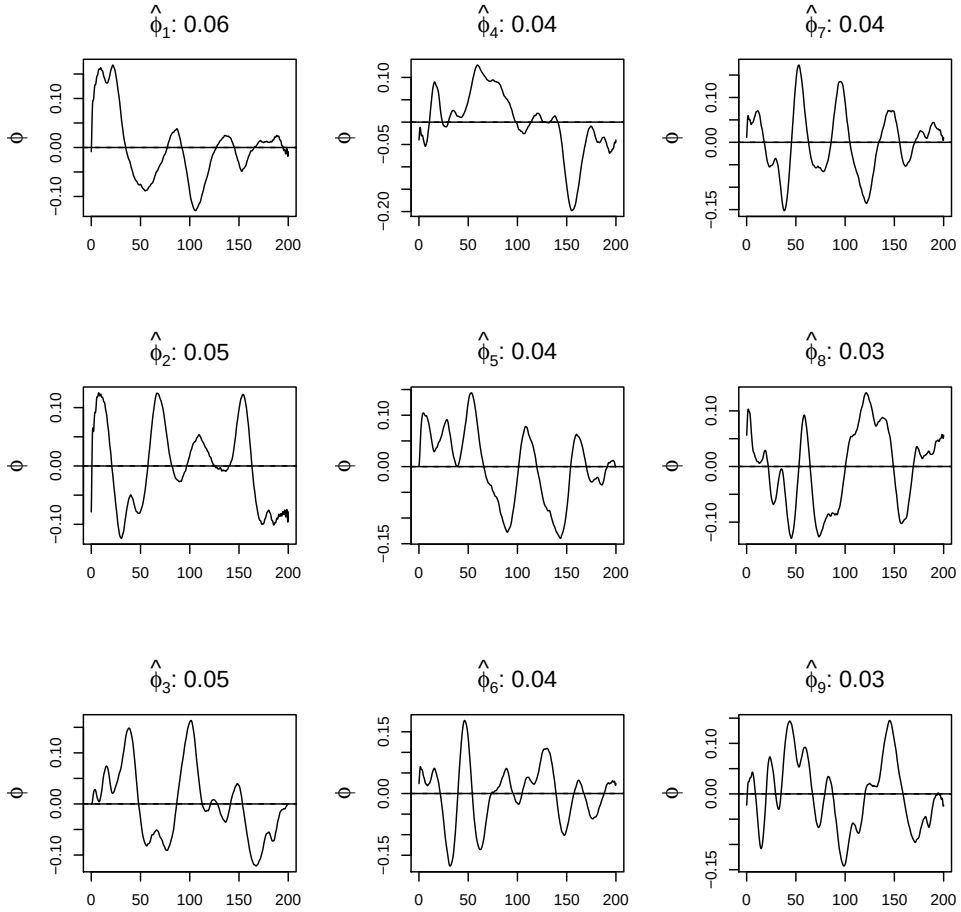


Abbildung 6.9: *Funktionale Hauptkomponenten 1 bis 9 des funktionalen RMS-Datenobjekts basierend auf 197 Fourier Basisfunktionen.*

Fourier Basisfunktionen auf 11 reduziert, verbessert sich das Bild der fPCA nicht erheblich. Abbildung 6.10 zeigt die ersten neun funktionalen Hauptkomponenten für das funktionale Datenobjekt basierend auf 11 Fourier Basisfunktionen.

Der Anteil der erklärten Varianz liegt hier für die erste funktionale Hauptkomponente bei nur $\hat{\Phi}_1 = 0.15$. Die einzelnen Hauptkomponenten erlauben keine sinnvollen Interpretationen hinsichtlich vorhandener elementarer prosodischer Grundformen der RMS-Konturen. Ähnliche Anteile erklärter Varianz ergeben sich für die auf B-Splines basierenden funktionalen Datenobjekte. Für die erste funktionale Hauptkomponente des funktionalen Datenobjekts basierend auf 11 B-Spline Basisfunktionen beträgt der Anteil erklärter Varianz $\hat{\Phi}_1 = 0.16$.

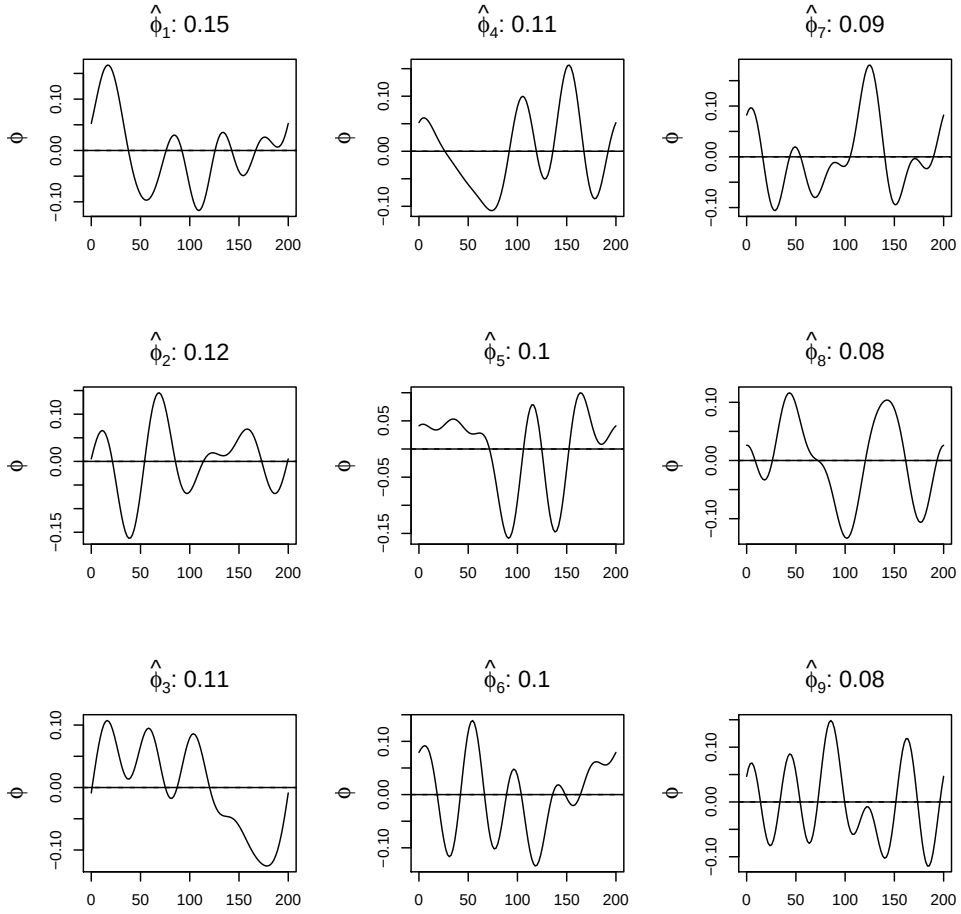


Abbildung 6.10: *Funktionale Hauptkomponenten 1 bis 9 des funktionalen RMS-Datenobjekts basierend auf 11 Fourier Basisfunktionen.*

Die fPC Scores $\delta_{c,i}$, $c = 1 \ldots 9$ wurden auch hier mit MEM Analyse (Baayen [2008]) und den 'fixed factors' Alkoholisierung und Geschlecht sowie den 'random factors' Aufnahmeelement und Sprecher getestet. Für die Scores der ersten fPC $\delta_{c,i}$, $c = 1$ ergibt sich bis auf das funktionale Datenobjekt basierend auf 5 Fourier Basisfunktionen ($F = 5.25$, $p < 0.05$) für alle Anzahlen von Basisfunktionen und beide Basissysteme ein hochsignifikanter Effekt für die Alkoholisierung ($F > 14.86$, $p < 0.001$). Die Scores der fPCs $\delta_{c,i}$, $c = 2 \ldots 9$ unterscheiden sich nicht signifikant zwischen a- und na-Konturen. Auch hier wurde keine Interaktion mit dem Geschlecht festgestellt.

6.4.2.3 Kontrollgruppenversuche Scores - Ergebnisse F0 und RMS

Da schon die Auswertungen zu den Scores der fPCs $\delta_{c,i}$, $c = 1 \dots 9$ auf Basis aller Sprecher nur für die Scores der ersten fPC $\delta_{c,i}$, $c = 1$ signifikante Ergebnisse lieferten, waren für eine Analyse der Scores die Kontrollgruppe betreffend keine signifikanten und aufschlussreichen Ergebnisse zu erwarten. Diese seien der Vollständigkeit halber dennoch aufgeführt. Auf eine Interpretation wurde verzichtet, da sich hier vermutlich nur durch die geringe Datenmenge erkennbare Effekte ergeben. Die fPCs $\phi_c, c = 1 \dots 9$ und die zugehörigen Scores $\delta_{c,i}$, $c = 1 \dots 9$ wurden einerseits für alle F0- und andererseits für alle RMS-Konturen (a, na und cna) und die oben genannten Anzahlen von Basisfunktionen (Fourier: 5-197, B-Splines: 5-101) aller 20 Kontrollgruppensprecher berechnet. Die Scores wurden auch hier mit MEM Analyse (Baayen [2008]) und den 'fixed factors' Alkoholisierung und Geschlecht sowie den 'random factors' Aufnahmeelement und Sprecher getestet. Die Kontrollgruppenversuche lieferten für die Scores im Falle der F0 weder zwischen a- und na-Konturen noch zwischen na- und cna-Konturen und für kein funktionales Datenobjekt signifikante Ergebnisse.

Bei den RMS-Konturen ergeben sich signifikante Unterschiede zwischen Sprache unter Alkoholeinfluss und in nüchternem Zustand geäußerter Sprache für die Scores der 4. Hauptkomponente der funktionalen Datenobjekte (B-Splines und Fourier) basierend auf 5 Basisfunktionen und für die Scores der 8. Hauptkomponente des funktionalen Datenobjekts basierend auf 11 B-Spline Basisfunktionen ($F > 9.04$, $p < 0.001$). Zwischen na- und cna-Sprache ergeben sich für die Scores der 7. Hauptkomponente für nahezu alle Datenobjekte basierend auf Fourier Basisfunktionen (21-197 Basisfunktionen) schwach signifikante Unterschiede ($F > 3.96$, $p < 0.1$) und für die funktionalen Datenobjekte basierend auf 51 und 75 B-Spline Basisfunktionen ebenfalls ($F > 3.76$, $p < 0.1$).

6.5 Konturen - Diskussion

Sowohl die Ergebnisse zu den F0-Konturen als auch die zu den RMS-Konturen der gelesenen Äußerungen des ALC lassen einen signifikanten Unterschied zwischen Sprachaufnahmen in alkoholisiertem und nüchternem Zustand eines Sprechers erkennen. F0- und RMS-Werte können auf algorithmischem Wege direkt aus dem Sprachsignal berechnet werden. Damit sind die entsprechenden Konturen eine einfach zu ermittelnde Basis zur Extraktion verschiedener Parameter auf Ebene der Konturen in ihrer Gesamtheit. In Kombination mit anderen Merkmalen in einer sprecherabhängigen Klassifikationsaufgabe könnten sie eine etwaige Alkoholisierung des Sprechers aufdecken. Globale Maße wie der Euklidische Abstand zwischen zwei Konturen, der Distanz, die auf Korrelation beruht und der Distanz im DCT Raum, sind deutlich größer zwischen Konturen von Sprachaufnahmen in alkoholisiertem und nüchternem Zustand der Sprecher als zwischen Konturen von in nüchternem Zustand geäußerten Aufnahmen und Kontrollgruppenaufnahmen. Dies ist ein Hinweis darauf, dass sowohl F0- als auch RMS-Konturen von Sprache unter Alkoholeinfluss und in nüchternem Zustand geäußerter Sprache grundsätzlich verschieden sind. Wie genau sich diese Unterschiede aber zeigen, ist damit nicht geklärt. Beispielsweise könnte ein vermehrtes Auftreten von Pausen oder längere Pausen Wortverschiebungen hervorrufen, oder es könnten Längungen und Kürzungen von linguistischen Einheiten auftreten. Eine Isolierung der verantwortlichen Faktoren würde zur Verbesserung des Klassifikationsmodells beitragen.
Deshalb wurden die Konturen mit Hilfe einer DCT parametrisiert und damit interpretierbare DCT Koeffizienten sowie das erste und zweite Moment des DCT Spektrums berechnet. Bei den F0-Konturen ergeben sich nur für das erste Moment signifikant höhere Werte für Konturen von Sprachaufnahmen unter Alkoholeinfluss als für Konturen von Sprachaufnahmen, die in nüchternem Zustand der Sprecher durchgeführt wurden, was auf schnellere Bewegungen in der Grundfrequenz bei Konturen von Sprachaufnahmen unter Alkoholeinfluss als bei Sprachaufnah-

men ohne Alkoholeinfluss hinweist (Frühere Studien wie Klingholz et al. [1988], Watanabe et al. [1994] oder Künzel and Braun [2003] berichten für Sprache unter Alkoholeinfluss gegenüber Sprache ohne Alkoholeinfluss von einer Vergrößerung des Bereichs, in dem sich die Grundfrequenz bewegt, was ebenfalls schnellere Bewegungen in den Konturen vermuten lässt). Bei den RMS-Konturen decken vor allem die DCT Koeffizienten 2 und 4 einen signifikanten Effekt zwischen den Konturen von Sprachaufnahmen unter Alkoholeinfluss und Sprachaufnahmen, die in nüchternem Zustand der Sprecher durchgeführt wurden, auf. Die Werte sind signifikant niedriger im alkoholisierten Fall. Damit deuten die Ergebnisse auf RMS-Konturen mit geringerer globaler Neigung und geringerer Schiefe bei Sprachsignalen unter Alkoholeinfluss hin. Die signifikant niedrigeren Werte bei Sprache unter Alkoholeinfluss für das erste Moment deuten auf langsamere Bewegungen in den Energiekonturen von Sprache unter Alkoholeinfluss gegenüber den Energiekonturen von in nüchternem Zustand geäußerter Sprache hin. Dies bestätigt Ergebnisse aus früheren Untersuchungen, wo Alkoholisierung mit einer Reduzierung der Sprechgeschwindigkeit einhergeht (Behne et al. [1991], Hollien et al. [1999, 2001], Künzel and Braun [2003], Sobell et al. [1982]) und ebenso die Ergebnisse zur Sprechgeschwindigkeit aus Kapitel 5.

Dennoch haben DCT Koeffizienten den Nachteil, dass die Basisfunktionen der Transformation fixe Cosinuswellen und nicht datengesteuert sind. Um elementar vorhandene Kurvenformen zu finden, mit deren Hilfe eine Unterscheidung von a-Konturen und na-Konturen möglich ist, wurde deshalb FDA in Form einer funktionalen Hauptkomponentenanalyse auf die Konturdaten angewandt. Jedoch sind die funktionalen Hauptkomponenten weder im Falle der F0-Konturen noch der RMS-Konturen in der Lage, alkoholisierungsbedingte Unterschiede aufzudecken. Der Anteil erklärter Varianz der ersten fPC der transformierten Konturdaten liegt in beiden Fällen und auch bei hohem Glättungsgrad bei maximal 22% (funktionales F0-Datenobjekt basierend auf 11 B-Spline Basisfunktionen), was ein Hinweis darauf ist, dass sowohl F0- als auch RMS-Konturdaten verschiedener

Äußerungen für eine Analyse per fPCA ungeeignet sind. Zwar erkennt eine MEM Analyse in beiden Fällen einen signifikanten Effekt der Alkoholisierung auf die Scores der ersten fPC, trotzdem zeigen sich jeweils bei der ersten fPC selbst keine erkennbaren alkoholisierungsabhängigen spezifischen elementaren Konturformen, die hinsichtlich der Konturdaten aller 19 relevanten, gelesenen Äußerungen des ALC (siehe Anhang C) Hilfe bei der Klassifizierung in a- und na-Konturen leisten oder linguistisch interpretiert werden könnten. Die erste fPC bei den F0-Konturen (Fourier Basis) könnte zwar auf ein globales Gefälle (positive Scores) oder im Falle negativer Scores auf einen globalen Anstieg als elementare Grundform in den Konturdaten hindeuten, doch ist ihr Anteil erklärter Varianz mit 20% auch bei hohem Glättungsgrad (funktionales F0-Datenobjekt basierend auf 11 Fourier Basisfunktionen) so gering, dass diese Grundformen nur in einem kleinen Teil der Konturdaten wiederzufinden sind. Die Vermutung liegt nahe, dass Konturen von verschiedenen gelesenen Aussagesätzen zu unterschiedlich für eine funktionale Hauptkomponentenanalyse sind. Dennoch könnte gesondert untersucht werden, in welchen der Konturen der 19 gelesenen Äußerungen die erste fPC als Grundform auftaucht. Wenn ein solcher Teil der Konturdaten Konturpaare gleicher Äußerungen (eines Sprechers) in beiden Sprecherzuständen von idealerweise allen Sprechern oder einem Großteil der Sprecher beinhaltet und gleichzeitig eine Neuberechnung der fPCs und Scores auf Basis dieses Teils der Daten ebenfalls einen signifikanten Unterschied für die erste fPC hinsichtlich der Alkoholisierung ergibt, würde dieser Teil der Konturdaten als gemeinsame Eigenschaft höchstwahrscheinlich ein globales Gefälle oder einen globalen Anstieg in der Grundfrequenz aufweisen und auf Basis der Scores der ersten fPC in Konturen der Sprachsignale von alkoholisierten und nüchternen Sprechern gruppierbar sein. Somit wäre es gegebenenfalls möglich, durch eine Eingrenzung der Konturdaten (im Idealfall beispielsweise auf die Konturen einer bestimmten Art von Äußerungen wie Adressen) eine elementar vorhandene Grundform in diesen Daten nachzuweisen, die eine maximale Trennbarkeit in Konturdaten alkoholisierter

und nüchterner Sprecher gestattet.

Auch eine einfache PCA (Karhunen-Loève-Transformation) auf den Rohdaten außerhalb der FDA liefert sehr ähnliche Werte und zeigt, dass eine Analyse per fPCA nicht in der Lage ist, aufschlussreichere Aussagen über die Daten zu erlauben. Abbildung 6.11 zeigt beispielhaft die Hauptkomponenten 1-9 für die RMS-Rohdaten. Der Unterschied zum funktionalen RMS-Datenobjekt basierend auf 197 Fourier Basisfunktionen (siehe Abbildung 6.9), welches die Rohdaten am präzisesten widerspiegelt, ist wie zu erkennen nur marginal.

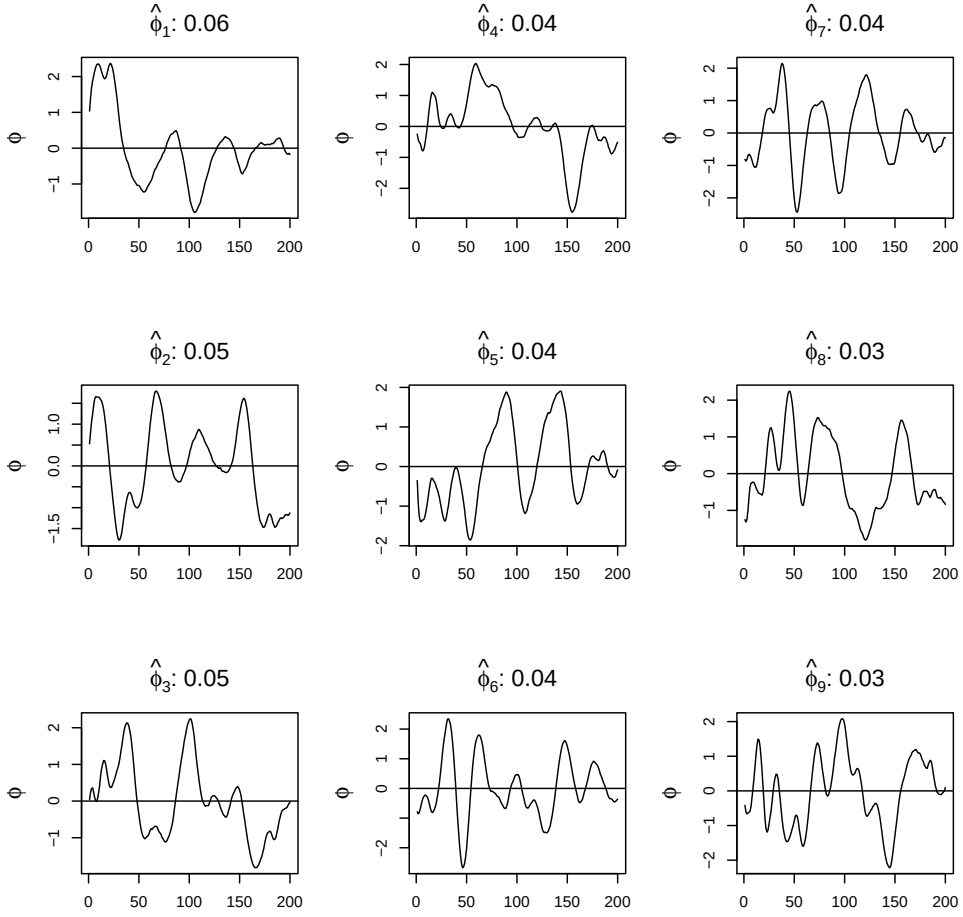


Abbildung 6.11: *Hauptkomponenten 1 bis 9 der RMS-Rohdaten.*

Kapitel 7

Diskussion und Zusammenfassung

Auf Basis des Alcohol Language Corpus, der in alkoholisiertem und nüchternem Zustand geäußerte Sprache von 77 weiblichen und 85 männlichen deutschen Sprecherinnen und Sprechern beinhaltet, wurden Untersuchungen zur rhythmischen Struktur der Sprache durchgeführt. Dabei wurde die Auswirkung von Alkohol auf das Verhalten einzelner aus dem Sprachsignal automatisch extrahierbarer Parameter untersucht, und ob sich elementare Grundformen bei Grundfrequenz- und Energiekonturen von Sprachsignalen nüchterner und alkoholisierter Sprecher finden lassen, die hinsichtlich der Alkoholisierung der Sprecher interpretierbar sind. Eine Vorhersage der Blutalkoholkonzentration mit Hilfe eines Parameters oder einer Kombination mehrerer Parameter bzw. eine Klassifikation von unbekanntem Sprachmaterial in Sprache unter Einfluss von Alkohol oder ohne Einfluss von Alkohol wurde nicht versucht. Diese Arbeit beschränkte sich auf Grundlagenforschung.

Die Ergebnisse der einzelnen Experimente zu Sprache unter Einfluss von Alkohol deuten darauf hin, dass im Vergleich zu in nüchternem Zustand geäußerter Sprache keine einheitlich auf alle Sprecher zutreffende Aussage über die Veränderung einzelner anhand des Sprachsignals extrahierbarer Parameter möglich ist. Vielmehr zeigte sich, dass sich die untersuchten Parameter sprecherindividuell verhalten. Rein statistisch ergeben sich aber dennoch signifikante Unterschiede.

7.1 Rhythmusparameter

Im ersten Experiment (Kapitel 3) wurden sogenannte *Rhythmusparameter*, die aus den Dauern und Dauerverhältnissen von vokalischen und konsonantischen Elementen bestimmt werden können, dahingehend untersucht, ob sich diese Dauern und Dauerverhältnisse zwischen Sprache unter Einfluss von Alkohol und in nüchternem Zustand geäußerter Sprache unterscheiden, und ob dieser Unterschied auf die Alkoholisierung zurückzuführen ist. Des Weiteren wurde untersucht, ob das Geschlecht des Sprechers und der Sprechstil ebenfalls einen Einfluss haben und ob die Faktoren sich gegenseitig beeinflussen. Als Basis zur Bestimmung der vokalischen und konsonantischen Elemente diente die automatisch erstellte phonetische Segmentierung durch MAUS (siehe auch Kapitel 3.1). Bei einem Teil der Parameter beinhaltet die Berechnung eine Normalisierung hinsichtlich der Sprechgeschwindigkeit. Der andere Teil wird ohne eine Normalisierung berechnet und muss damit als weniger aussagekräftig eingestuft werden, da die Sprechgeschwindigkeiten der beiden Sprachaufnahmen (Sprecher unter Alkoholeinfluss und nüchtern) und damit die absoluten Dauern der Elemente voneinander abweichen können. Es wurde festgestellt, dass die meisten Parameter unabhängig von der Normalisierung im statistischen Mittel bei Sprache unter Einfluss von Alkohol höhere Werte aufweisen als bei Sprache ohne Einfluss von Alkohol. Die vokalischen Elemente sind im Mittel länger bei Sprache unter Alkoholeinfluss, die Daueunterschiede zwischen den Elementen werden größer, es treten mehr kurze Pausen auf und die Sprechgeschwindigkeit sinkt. Demnach ist Sprache unter Alkoholeinfluss unregelmäßiger und langsamer als in nüchternem Zustand geäußerte Sprache. Anhand einer Kontrollgruppe wurden rein zufällige Effekte ausgeschlossen und nachgewiesen, dass die Veränderungen der Parameter tatsächlich auf die Alkoholisierung der Sprecher zurückzuführen sind. Hypothese 1 konnte damit statistisch bestätigt werden. Veränderungen in der rhythmischen Struktur der Sprache in Form von Unregelmäßigkeiten, die durch Alkoholisierung hervor-

gerufen werden, spiegeln sich in Veränderungen der Dauern und Dauerverhältnisse von phonetischen Einheiten wider. Die gefundenen Veränderungen betreffen nicht alle Sprechstile gleichermaßen, am deutlichsten zeigen sie sich bei gelesener Sprache. Das Sprechergeschlecht hat keinen Einfluss auf die Veränderungen der Parameter. Eine der Gruppierung verschiedener Sprachen ähnliche Trennbarkeit anhand paarweiser Darstellung bestimmter Rhythmusparameter im kartesischen Koordinatensystem (siehe Kapitel 3.4) lässt sich im Bezug auf Sprache unter Alkoholeinfluss und Sprache ohne Alkoholeinfluss nicht erreichen.
Die betäubende Wirkung von Alkohol zeigt sich hier vermutlich in einer Einschränkung des Artikulationsprozesses hinsichtlich Planung und Steuerung, was insgesamt in einer unbeständigen und verlangsamten Artikulation resultiert. Es könnte sein, dass die Sprecher auf Grund des erhöhten Konzentrationsaufwandes, der durch die Betäubung entsteht, unbewusst langsamer sprechen, um keine Einbußen bei der Verständlichkeit durch undeutliche Artikulation zu riskieren. Ebenfalls eingeschränkt könnte auch die Koordination der Bewegungssteuerung der Artikulatoren sein, wodurch Defizite im Bewegungsablauf hervorgerufen werden. Ob dabei auch direkte physiologische Auswirkungen von Alkohol auf die Artikulatoren eine Rolle spielen, bleibt unklar. Es ist jedoch bekannt, dass die Aufnahme von Alkohol die physiologische Beschaffenheit der Stimmlippen verändern kann (Watanabe et al. [1994]).

7.2 Rhythmizitäts-Parameter

Ein weiteres Experiment (Kapitel 4) basierte auf der Energiekontur, deren zugrundeliegende Werte sich aus dem Sprachsignal berechnen lassen. Anhand dieser Kontur wurden auf algorithmischem Wege lokale Minima und Maxima der Energie einer Äußerung bestimmt, deren sequenzielle Abfolge mehr oder weniger die Silbenabfolge der Äußerung widerspiegelt. Die Maxima entsprechen hierbei grob den Silbenkernen und die Minima den Silbengrenzen. Diese Sequenz bildete die Grundlage für die Berechnung von sogenannten *Rhythmizitäts-Parametern*,

welche die Dynamik im Energieverlauf beschreiben. Es wurde untersucht, ob sich unterschiedliche Werte der Parameter für Sprache unter Alkoholeinfluss und in nüchternem Zustand geäußerter Sprache ergeben, und ob diese Unterschiede auf die Alkoholisierung zurückzuführen sind. Des Weiteren wurde untersucht, ob auch das Geschlecht des Sprechers und der Sprechstil einen Einfluss haben und ob die Faktoren sich gegenseitig beeinflussen. Für alle Parameter ergeben sich im statistischen Mittel höhere Werte für Sprache unter Alkoholeinfluss gegenüber Sprache ohne Alkoholeinfluss. Der Dynamikumfang steigt, die Differenzen im RMS-Wert zwischen den ermittelten Minima und Maxima im Energieverlauf werden größer, ebenfalls die zeitlichen Differenzen. Die RMS-Kontur wird im Allgemeinen unruhiger. Damit deuten die Ergebnisse auf langsamere und unregelmäßigere Sprache in alkoholisiertem Zustand der Sprecher gegenüber Sprache in nüchternem Zustand hin. Diese generelle Aussage bestätigt die Ergebnisse zu den Rhythmusparametern des ersten Experiments. Vor allem auf zeitlicher Ebene lassen sich die gefundenen Dauerunterschiede der Rhythmusparameter in den Dauerunterschieden der Rhythmizitäts-Parameter wiederfinden. In beiden Fällen zeigen die Parameter im statistischen Mittel eine langsamere Sprechgeschwindigkeit für Sprache unter Alkoholeinfluss gegenüber Sprache ohne Alkoholeinfluss an. Auch hier sind vermutlich Einschränkungen im Artikulationsprozess bei Planung und Steuerung durch die betäubende Wirkung von Alkohol die Ursache. Die einzelnen Sprechstile sind nicht gleichermaßen betroffen. Wiederum scheint gelesene Sprache eine gute Basis für die Berechnung hinsichtlich der Alkoholisierung aussagekräftiger Parameter zu sein. Bei einigen Parametern wurden geringfügige geschlechtsspezifische Unterschiede gefunden, die jedoch unabhängig von der Alkoholisierung der Sprecher auftreten. In diesen Fällen unterscheiden sich die Parameter zwischen Sprache unter Alkoholeinfluss und Sprache ohne Alkoholeinfluss und zusätzlich geringfügig zwischen den Geschlechtern. Mit Hilfe der Kontrollgruppenversuche wurden zufällige Effekte ausgeschlossen und nachgewiesen, dass die Unterschiede der Parameter tatsächlich auf die Alkoholisierung der Sprecher zurückzuführen

sind. Hypothese 2 konnte somit statistisch bestätigt werden. Veränderungen der rhythmischen Struktur der Sprache in Form von Unregelmäßigkeiten und Lautstärkeschwankungen, die durch Alkoholisierung hervorgerufen werden, spiegeln sich in Veränderungen in der Dynamik der RMS- oder Energiefunktion eines Sprachsignals wider.

7.3 Sprechgeschwindigkeit

In einer gesonderten Untersuchung wurden die Sprechgeschwindigkeitsunterschiede zwischen Sprache unter Alkoholeinfluss und in nüchternem Zustand geäußerter Sprache behandelt. Kapitel 5 beschreibt und vergleicht dazu zwei Ansätze zur Schätzung der Sprechgeschwindigkeit. Ein Ansatz basiert auf der phonetischen Segmentierung durch MAUS und liefert durch Zählung der vokalischen Elemente, welche vereinfacht angenommen die Silbennuklei repräsentieren, die Anzahl der Silben pro Sekunde (SR_P, siehe auch Kapitel 3.1). Unabhängig von einer vorhandenen Segmentierung wurde in einem zweiten Ansatz die Sprechgeschwindigkeit aus dem durchschnittlichen zeitlichen Abstand der Maxima im Energieverlauf geschätzt ($SRRP$). Auch die Maxima repräsentieren im Groben die Silbennuklei und wurden wie in Kapitel 4 beschrieben auf algorithmischem Wege auf Basis der Energiekontur, deren zugrundeliegende Werte sich wiederum aus dem Sprachsignal berechnen lassen, ermittelt. Beide Sprechgeschwindigkeitsmaße zeigen im statistischen Mittel eine verlangsamte Sprechgeschwindigkeit bei Sprache unter Alkoholeinfluss gegenüber Sprache ohne Alkoholeinfluss an. Korrelationen von SR_P, welches als Referenzwert dient, und $SRRP$ signalisieren, dass $SRRP$ hinreichend zur Abschätzung der Sprechgeschwindigkeit auch bei verschiedenen Sprechstilen herangezogen werden kann (niedrigster Korrelationskoeffizient $r = 0.72$ $[p < 0.001]$ bei Spontansprache ohne Alkoholeinfluss, höchster Korrelationskoeffizient $r = 0.87$ $[p < 0.001]$ bei Kommandosprache ohne Alkoholeinfluss). Wie bereits angeführt, scheint Alkohol eine Verlangsamung des kompletten Artikulationsprozesses zu bewirken. Teilhypothesen 3 a) und b)

konnten statistisch bestätigt werden. Sprache alkoholisierter Personen weist eine verlangsamte Sprechgeschwindigkeit gegenüber in nüchternem Zustand geäußerter Sprache auf. Des Weiteren wurde ein Perzeptionsexperiment durchgeführt, das zur Klärung der Frage dienen sollte, ob die gefundene Verlangsamung der Sprache unter Alkoholeinfluss gleichzeitig ein Indiz für Alkoholisierung bei der Perzeption von Sprache sein kann. Dafür wurden Hörern je zwei Stimuli von in nüchternem Zustand geäußerter Sprache zur Auswahl geboten, wobei einer der beiden Stimuli synthetisch um 5% gelängt war, um eine Verlangsamung wie bei Sprache unter Alkoholeinfluss zu erwirken. Der Hörer musste sich trotz der Tatsache, dass im Grunde nur Sprachmaterial ohne Alkoholeinfluss dargeboten wurde, entscheiden, welcher der beiden Stimuli dem Anschein nach Sprache unter Alkoholeinfluss enthält. Es stellte sich heraus, dass die Antworten nicht rein zufällig ausfallen. In 55.5% aller Fälle wurde der manipulierte Stimulus als derjenige ausgewählt, bei dem der Sprecher vermeintlich alkoholisiert war (signifikant über Zufall [50%], $\chi^2 = 11.5, p < 0.001$). Demnach scheint bei der Perzeption von Sprache eine Verringerung der Sprechgeschwindigkeit ein Indiz für eine etwaige Alkoholisierung des Sprechers zu sein $\big($Teilhypothese 3 c)$\big)$. Dieses Ergebnis bestätigt die auf Basis des akustischen Sprachsignals gefundenen Effekte zur Sprechgeschwindigkeit.

7.4 Konturen

Im letzten Experiment (Kapitel 6) wurden rhythmische Eigenschaften von Grundfrequenz- und Energiekonturen von Sprachaufnahmen, die unter Alkoholeinfluss und in nüchternem Zustand der Sprecher durchgeführt wurden, untersucht. Genau wie Energiekonturen lassen sich auch Grundfrequenzkonturen bzw. deren zugrundeliegende Werte auf algorithmischem Wege aus dem Sprachsignal ermitteln. Die gewonnenen Konturen wurden auf zeitlicher Ebene in ihrer Länge linear aneinander angeglichen und dann auf verschiedene Weisen miteinander verglichen. Da ein Vergleich nur bei Konturen von Äußerungen, die bei Sprache unter Alkoholeinfluss und Sprache ohne Alkoholeinfluss inhaltlich

übereinstimmen, sinnvoll ist, wurde nur gelesenes Sprachmaterial in Betracht gezogen, da hier für beide Sprecherzustände inhaltsgleiche Äußerungen vorliegen.

Anhand verschiedener Distanzmaße wurde zunächst festgestellt, dass unabhängig vom Distanzmaß der jeweilige Wert zwischen Konturen von Sprachaufnahmen unter Alkoholeinfluss und Konturen von Sprachaufnahmen in nüchternem Zustand der Sprecher größer ist als zwischen Konturen von Sprachaufnahmen in nüchternem Zustand der Sprecher und den Konturen der ebenfalls in nüchternem Zustand der Sprecher durchgeführten Kontrollgruppenaufnahmen. Die Ergebnisse besagen, dass Grundfrequenz- und Energiekonturen von Sprachaufnahmen unter Alkoholeinfluss und Konturen von Sprachaufnahmen ohne Alkoholeinfluss generell verschieden sind. Wodurch die Unterschiede in den Distanzwerten aber hervorgerufen werden und auf welche Weise sich die Konturen unterscheiden, ist damit nicht geklärt. Allein durch Pausen, Längungen oder Kürzungen könnten sich linguistische Einheiten verschieben und damit eine Veränderung des jeweiligen Distanzwertes hervorrufen.

Deshalb wurden die Konturen im weiteren Verlauf anhand einer Diskreten Cosinus Transformation parametrisiert und damit interpretierbare DCT Koeffizienten, sowie das erste und zweite Moment des DCT Spektrums berechnet. Bei den Grundfrequenzkonturen lassen die Ergebnisse schnellere Bewegungen im Konturverlauf von Sprachaufnahmen unter Alkoholeinfluss gegenüber Sprachaufnahmen ohne Alkoholeinfluss und damit eine unruhigere Grundfrequenzkontur erkennen. Im Gegensatz dazu deuten die Ergebnisse zu den Energiekonturen auf langsamere Bewegungen im Konturverlauf von Sprachaufnahmen unter Alkoholeinfluss gegenüber Sprachaufnahmen ohne Alkoholeinfluss hin. Langsamere Bewegungen lassen auf eine verlangsamte Sprechgeschwindigkeit schließen, was die Ergebnisse zur Sprechgeschwindigkeit in Kapitel 5 bestätigt.

Die oben genannten Verfahren haben den Nachteil, dass sie den Informationsgehalt der Konturen auf wenige Werte reduzieren. Die eigentliche Form der Kontur findet dabei keine bzw. nur geringfügige Beachtung. Einzig die Parametrisierung gibt einen Hinweis auf vorhandene Grundformen, jedoch lässt sich damit nicht sagen, ob diese Grundformen für die vorliegenden Daten optimal und phonetisch bzw. prosodisch interpretierbar sind. Um elementar vorhandene Formen in den Konturverläufen aufzudecken und dabei einen Unterschied zwischen den Konturen von Sprache unter Alkoholeinfluss und Sprache ohne Alkoholeinfluss zu finden, wurden die Grundfrequenz- und Energiekonturen deshalb mit Hilfe der funktionalen Datenanalyse und im speziellen der funktionalen Hauptkomponentenanalyse untersucht. Jedoch sind die funktionalen Hauptkomponenten weder bei den F0- noch den RMS-Konturen in der Lage, grundsätzlich vorhandene Unterschiede in den gesammelten Konturdaten aufzudecken. Interessante Grundformen könnten unter Umständen die ersten drei funktionalen Hauptkomponenten bei den Grundfrequenzkonturen widerspiegeln. Ein globales Gefälle als Grundform in den Konturen, eine frageähnliche Intonationskontur und eine Markierung wesentlicher Akzentpositionen. Allerdings sind diese Grundformen nur in einem relativ kleinen Teil der Daten wiederzufinden. Deshalb lässt sich durch sie für die kollektiven Konturdaten der 19 relevanten, gelesenen Äußerungen des ALC (siehe Anhang C) keine prosodisch interpretierbare Trennung in Konturen von Sprachaufnahmen unter Alkoholeinfluss und Sprachaufnahmen ohne Alkoholeinfluss erreichen. Bei den Energiekonturen ergeben sich keine interessanten Grundformen und auch hier ist durch sie nur ein sehr geringer Teil der Daten erklärbar. Die Ergebnisse deuten darauf hin, dass Grundfrequenz- und Energiekonturen von verschiedenen gelesenen Aussagesätzen zu unterschiedlich für eine funktionale Hauptkomponentenanalyse sind. Es stellte sich damit im Nachhinein heraus, dass die Anwendung der funktionalen Datenanalyse auf die Daten des ALC als nicht sehr erfolgreich eingestuft werden kann. Ebenfalls durchgeführt wurde eine einfache Hauptkom-

ponentenanalyse auf den Rohdaten, welche sehr ähnliche Ergebnisse wie die funktionale Hauptkomponentenanalyse liefert. Somit zeigt sich außerdem, dass die funktionale Hauptkomponentenanalyse keine aufschlussreicheren Aussagen über die Daten zulässt als eine einfache Hauptkomponentenanalyse.
Die Hypothesen 4 und 5 konnten auf Basis der Distanzwerte und Parametrisierung größtenteils statistisch bestätigt werden. Veränderungen in der rhythmischen Struktur der Sprache, die durch Alkoholisierung hervorgerufen werden, spiegeln sich auch in charakteristischen Veränderungen der F0- und RMS-Kontur des Sprachsignals wider. Die einfache und auch die funktionale Hauptkomponentenanalyse dagegen tragen durch ihre Ergebnisse kaum zur Aufklärung bei.

Wie in Ramsay et al. [2009] anhand verschiedener Beispiele beschrieben, sind für die funktionale Datenanalyse vergleichbare Datensätze (z.B. über mehrere Jahre dokumentierte Temperaturdaten einer Ortschaft mit jeweils einer Temperaturkurve pro Jahr [siehe Ramsay et al. [2009]] oder verschiedene Äußerungen desselben Satztyps [siehe Gubian et al. [2011]]) im Grunde eine zwar nicht notwendige, aber empfohlene Voraussetzung, damit eine sinnvolle Interpretation der Ergebnisse möglich ist. Die 19 analysierten gelesenen Äußerungen des ALC sind relativ unterschiedlich, sowohl in Struktur, Länge als auch Inhalt. Konturen eines solchen Datensatzes scheinen für die funktionale Datenanalyse ungeeignet zu sein. Trotzdem war ein Versuch, die Daten mittels funktionaler Datenanalyse zu untersuchen, durchaus gerechtfertigt. Denn so konnte auf funktionaler Ebene eine Hauptkomponentenanalyse durchgeführt werden, um damit eventuell vorhandene prosodisch interessante fundamentale Grundformen in den Konturen von Sprache unter Alkoholeinfluss und Sprache ohne Alkoholeinfluss aufzuspüren, auch wenn sich am Ende herausstellte, dass keine wirklich aussagekräftige Interpretation möglich war.
Mit Hilfe der durch die funktionale Datenanalyse bereitgestellten Werkzeuge kön-

nen Konturen durch 'registering' aligniert werden. Hierbei werden die Konturen der Äußerungen durch verschiedene Ankerpunkte auf zeitlicher Ebene aneinander angeglichen bzw. aligniert und damit vergleichbar gemacht. Diese Alignierung macht aber nur Sinn, wenn sich in den Konturen gleiche oder zumindest vergleichbare Ankerpunkte finden lassen, was bei 19 verschiedenen Äußerungen nahezu unmöglich ist. Des Weiteren wird beim 'registering' die zeitliche Struktur der einzelnen Äußerungen und damit auch der Konturen nicht linear verändert bzw. verzerrt. Bei der Untersuchung der rhythmischen Struktur von Sprache läuft man dabei Gefahr, relevante, in der Zeitstruktur gespeicherte, Informationen zu entfernen, die gegebenenfalls einen Hinweis auf eine vorliegende Alkoholisierung des Sprechers liefern könnten und deshalb erhalten bleiben müssen. Aus diesem Grund wurden die Konturen im vorliegenden Fall nur durch lineare Zeitnormalisierung in ihrer Länge angepasst.

Ein weiterer Schritt ist die funktionale Datenanalyse der einzelnen Konturen einer Äußerung für 150 Sprecher. Dabei werden nur Konturen von inhaltsgleichen Äußerungen miteinander verglichen. Exemplarisch wurde ein solcher Versuch für die Konturen (ohne 'registering') einiger Äußerungen durchgeführt. Auch hier ergaben sich keine interessanten funktionalen Hauptkomponenten und nur sehr geringe Anteile erklärter Varianz, vermutlich weil auch inhaltsgleiche Äußerungen durch Pausen, Längungen oder Kürzungen Konturen zur Folge haben können, die letztendlich zu unterschiedlich für eine Hauptkomponentenanalyse sind. Um sinnvolle interpretierbare Ergebnisse zu erhalten, wurde deshalb darüber hinaus versuchsweise eine Alignierung durch 'registering' durchgeführt, auch wenn dabei die erwähnte Verzerrung auf zeitlicher Ebene auftritt. Die Ergebnisse sind aber trotz 'registering' und der Analyse von Konturen einer Äußerung verschiedener Sprecher nicht sehr aufschlussreich. Es ergeben sich ebenfalls nur sehr geringe Anteile erklärter Varianz für die funktionalen Hauptkomponenten. Selbst wenn im Idealfall einige funktionale Hauptkomponenten die Varianz in den Konturdaten einer Äußerung zum Großteil erklären könnten, ließen sie sich dennoch nur

hinsichtlich jeweils einer Äußerung sinnvoll interpretieren, was nicht Zweck dieser Untersuchung war. Vielmehr sollten allgemein vorhandene fundamentale Grundformen von Sprachaufnahmen unter Alkoholeinfluss und Sprachaufnahmen ohne Alkoholeinfluss gefunden werden.

7.5 Prognosemodelle

Im Rahmen dieser Arbeit wurde kein Prognosemodell zur Alkoholisierung von Sprechern vorgestellt. Es wäre aber durchaus interessant, die hier untersuchten und gegebenenfalls weitere Parameter zu kombinieren und ein solches Prognosemodell zu entwickeln. Die Interspeech Speaker State Challenge (Schuller et al. [2011]) stellte ihre Teilnehmer vor ein binäres Klassifikationsproblem. Der Sprecherzustand (alkoholisiert oder nüchtern) sollte anhand des Sprachsignals vorhergesagt werden. Sprachmaterial von Sprechern mit einer Blutalkoholkonzentration kleiner oder gleich 0.5‰ wurde als in nüchternem Zustand geäußerte Sprache gewertet, Sprachmaterial von Sprechern mit einer Blutalkoholkonzentration größer als 0.5‰ als Sprache unter Alkoholeinfluss. Das beste sprecherunabhängige Klassifikationsmodell von Bone et al. [2011] erreichte dabei eine Erkennungsrate von ca. 70% und verwendete zur Klassifizierung eine Kombination verschiedener aus dem Sprachsignal extrahierbarer akustischer Merkmale. Im Vergleich dazu erreichten Hörer in einem Perzeptionsexperiment auf Basis derselben Daten eine Erkennungsrate von 72% (Schiel [2011]). Diese Ergebnisse zusammen mit den gefundenen Ergebnissen der hier beschriebenen Experimente deuten darauf hin, dass ein allgemeines statistisches Modell nicht in ausreichendem Maße in der Lage ist, den Blutalkoholwert eines unbekannten Sprechers bzw. ob dieser alkoholisiert ist oder nicht, anhand des Sprachmaterials vorherzusagen. Dies zeigen auch die Ergebnisse der Pearson Korrelationen einzelner Parameter mit dem Blutalkoholwert. Der Grad der Veränderung der Parameter lässt im Allgemeinen keinen Rückschluss auf den Grad der Alkoholisierung der Sprecher zu (Korrelationen innerhalb eines Sprechers konnten nicht berechnet werden, da pro

Sprecher und Aufnahme nur eine Messung des Blutalkoholwerts vorgenommen wurde). Erfolgversprechend scheinen dagegen sprecherabhängige Modelle zu sein, die auf Basis der Sprache ohne Alkoholeinfluss einer Person aufgebaut und dann herangezogen werden können, um unbekanntes Sprachmaterial derselben Person zu klassifizieren.

7.6 Ausblick und Schlusswort

In dieser Arbeit wurden verschiedene umfangreiche Untersuchungen zur rhythmischen Struktur von Sprache unter Alkoholeinfluss auf Basis derselben statistisch aussagekräftigen Daten vorgestellt. Die Ergebnisse der einzelnen grundlegenden Studien liefern wichtige Erkenntnisse zum Verhalten automatisch aus dem Sprachsignal extrahierbarer Merkmale bei Sprache unter Einfluss von Alkohol im Vergleich zu Sprache ohne Einfluss von Alkohol. Forensische Untersuchungen oder Gutachten könnten von den vorgestellten Untersuchungen profitieren, indem zusätzlich zu den bereits bewährten Parametern wie der Grundfrequenz die vorgestellten Parameter zur rhythmischen Struktur im Rahmen einer Gesamtbewertung von Sprachmaterial hinzugezogen werden, um damit validere Aussagen zu ermöglichen. Da Veränderungen in der Grundfrequenz außer durch Alkoholisierung auch durch andere Faktoren wie Stress, Freude oder Traurigkeit bedingt sind, muss ebenfalls das Verhalten der Parameter zur rhythmischen Struktur der Sprache anhand unabhängiger Studien und entsprechendem Sprachmaterial hinsichtlich dieser Faktoren untersucht werden. Treten hierbei dieselben Effekte auf, ist es umso mehr erforderlich, die verschiedenen Parameter zu kombinieren, damit eine Bewertung nicht auf Basis weniger Merkmale unsicher ausfallen muss. Interessant wäre auch eine Analyse der Parameter zur rhythmischen Struktur der Sprache bei Müdigkeit, da die sedativen Effekte von Alkohol am ehesten vergleichbar mit den Effekten bei Müdigkeit sind. Würde das Verhalten der Parameter bei Müdigkeit ähnlich ausfallen wie bei Sprache unter Alkoholeinfluss, müssten vor allem automatische Systeme, die eine Bewertung des Sprachsignals

hinsichtlich einer eventuellen Alkoholisierung des Sprechers treffen sollen (z.B. im Hintergrund der Sprachbefehlseingabe eines Navigationssystems bei Kfz) dahingehend angepasst werden. Wird ein Fahrzeugführer nämlich bei Müdigkeit fälschlicherweise als alkoholisiert eingestuft, weil dieselben Effekte gefunden werden, würde eine potenzielle Bewertungsfunktion schnell in Ungnade fallen und sofern möglich höchstwahrscheinlich deaktiviert werden.

Die Analysen beruhen alle auf den Signalen des Headset Mikrofons. Es handelt sich dabei um Nahbesprechungsmikrofonsignale, die durch die Richtcharakteristik des Mikrofons und die unmittelbare Nähe zur Signalquelle (ca. 4 cm seitlich vom Mund des Sprechers) relativ unbelastet von Nebengeräuschen sind. Neben den Headset Signalen existieren ebenfalls Aufnahmesignale eines Grenzflächenmikrofons, das durch seine Richtcharakteristik die Raumakustik mit erfasst. Dadurch sind das Sprachsignal sowie jegliche Art von Nebengeräuschen, dazu gehören beispielsweise auch Sprachdaten einer weiteren anwesenden Person, enthalten. Nachdem in der Realität zumeist mit Nebengeräuschen gerechnet werden muss oder die Aufnahmequalität schlechter ist als bei einem Headset Mikrofon, sollten die Experimente auf Basis der Grenzflächenmikrofonsignale wiederholt werden, um das Verhalten der Parameter unter wirklichkeitsnahen Bedingungen zu untersuchen. Sollten sich vergleichbare Effekte ergeben, wäre dies ein Hinweis darauf, dass die untersuchten Parameter unempfindlich gegenüber Nebengeräuschen und damit gut geeignet für eine Bewertung von Sprachsignalen hinsichtlich Alkoholisierung sind.

Des Weiteren wurden bisher noch keine automatisch zu ermittelnden Merkmale des Sprachsignals gefunden, die in einer sprecherunabhängigen Klassifikationsaufgabe eine eindeutige und erfolgreiche Aussage hinsichtlich einer eventuell vorliegenden Alkoholisierung des Sprechers zulassen. Auch die Ergebnisse der präsentierten Experimente deuten darauf hin, dass sprecherab-

hängige Bewertungsschemata erfolgversprechender sind. Jedoch besteht dabei im Allgemeinen die Notwendigkeit, dass ein Modell speziell für einen Sprecher angefertigt wird und dann zur Bewertung herangezogen werden kann. Ein erweiterter Nutzerkreis ist damit weitestgehend ausgeschlossen, sofern nicht für jeden einzelnen Nutzer Modelle generiert werden können.

Ob die gefundenen Ergebnisse auf andere Sprachen übertragen werden können, bleibt ungeklärt. Eine Beantwortung dieser Frage bedarf gesonderten Untersuchungen in den jeweiligen Sprachen. Zu erwarten sind jedoch auf Grund der verschiedenen rhythmischen Strukturen der Sprachen (bzw. die durch die Rhythmusparameter propagierten Rhythmusklassen, siehe Kapitel 3) unterschiedliche Ergebnisse für verschiedene Sprachen bzw. ähnliche Ergebnisse innerhalb der Gruppen von Sprachen mit vergleichbaren rhythmischen Strukturen.

Die präsentierten Untersuchungen wurden mit größter Sorgfalt durchgeführt. Dennoch können auf Grund der Vielzahl von Verarbeitungsschritten einzelne Fehlerquellen nicht ganz ausgeschlossen werden. Eine Wiederholung der Experimente auf Basis desselben Sprachmaterials in unabhängigen Studien zur Verifizierung der Ergebnisse wäre deshalb wünschenswert. An der Quintessenz wird sich jedoch nicht viel ändern. Die Auswirkungen von Alkohol auf das Verhalten einzelner automatisch extrahierbarer Merkmale des Sprachsignals sind zu einem großen Teil sprecherabhängig. Genauso sind die Auswirkungen von Alkohol auf den menschlichen Körper individuell unterschiedlich, einzig die Blutalkoholkonzentration liefert einen objektiven Hinweis auf den Zustand des Konsumenten. Eine Pauschalaussage zum Verhalten von Menschen unter Alkoholeinfluss bzw. zur Wirkungsweise von Alkohol sollte genau wie bei den Parametern zur rhythmischen Struktur der Sprache vermieden werden, auch wenn sich aus statistischer Sicht grundsätzliche Effekte erkennen lassen.

Literaturverzeichnis

D. Abercrombie. *Elements of general phonetics.* Edinburgh University Press, Edinburgh, 1967.

G. A. Aldermann, H. Hollien, C. A. Martin, and G. DeJong. Shifts in fundamental frequency and articulation resulting from intoxication. *Journal of the Acoustical Society of America*, 97:3363–3364, 1995.

W. S. Allen. *Accent and Rhythm.* Cambridge University Press: Cambridge, 1973.

A. Arvaniti. Rhythm, timing and the timing of rhythm. *Phonetica*, 66:46–63, 2009.

R. H. Baayen. *Analysing Linguistic Data: A Practical Introduction to Statistics Using R.* Cambridge University Press: Cambridge, 2008.

M. G. Barlow and M. Wagner. Prosody as a basis for determining speaker characteristics. In *Proceedings of the 2nd Australian International Conference on Speech Science and Technology*, pages 80–85, Sydney, Australia, 1988.

W. J. Barry, B. Andreeva, M. Russo, S. Dimitrova, and T. Kostadinova. Do rhythm measures tell us anything about language type? In *Proceedings of the 15th ICPhS*, pages 2693–2696, Barcelona, Spain, 2003.

D. M. Behne and S. M. Rivera. Effects of alcohol on speech: Acoustic analyses of spondees. *Research on Speech Perception Progress Report*, 16:263–291, 1990.

D. M. Behne, S. M. Rivera, and D. B. Pisoni. Effects of alcohol on speech: Durations of isolated words, sentences and passages. *Research on Speech Perception*, 17:285–301, 1991.

B. Bloch. Studies in colloquial Japanese IV: Phonemics. *Language*, 26(1):86–125, 1950.

P. Boersma. Praat, a system for doing phonetics by computer. *Glot International*, 5:9(10):341–345, 2001.

D. L. Bolinger. Pitch accent and sentence rhythm. In I. Abe and T. Kanekiyo, editors, *Forms of English: Accent, Morpheme, Order*, pages 139–180, Cambridge, 1965. Harvard University Press.

D. Bone, M. P. Black, M. Li, A. Metallinou, S. Lee, and S. Narayanan. Intoxicated speech detection by fusion of speaker normalized hierarchical features and GMM supervectors. In *Proceedings of Interspeech 2011*, pages 3217–3220, Florence, Italy, 2011. International Speech Communication Association.

A. Braun. Speaking while intoxicated: Phonetic and forensic aspects. In *Proceedings of the XIIth International Congress of Phonetic Sciences*, pages 146–149, Aix-en-Provence, France, 1991.

S. Cassidy and J. Harrington. Multi-level annotation in the EMU speech database management system. *Speech Communication*, 33(1-2):61–77, 2001.

C. Cheng and M. Gubian. Predicting Taiwan Mandarin tone shapes from their duration. In *Proceedings of Interspeech 2011*, pages 1073–1076, Florence, Italy, 2011. International Speech Communication Association.

S. B. Chin and D. B. Pisoni. *Alcohol and Speech*. Academic Press, CA, USA, 1997.

R. A. J. Clark and K. E. Dusterhoff. Objective methods for evaluating synthetic intonation. In *Proceedings of Eurospeech 1999*, volume 4, pages 1623–1626, Budapest, Hungary, 1999.

O. M. Cooney, K. G. McGuigan, and P. J. P. Murphy. Acoustic analysis of the effects of alcohol on the human voice. *Journal of the Acoustical Society of America*, 103(5):2895, 1998.

K. E. Cummings, S. B. Chin, and D. B. Pisoni. Acoustic and glottal excitation analyses of sober vs. intoxicated speech: A first report. *Research on Spoken Language Processing Progress Report*, 20:359–386, 1995.

R. M. Dauer. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11(1):51–62, 1983.

A. W. De Groot. Phonetics in its relation to aesthetics. In B. Malmberg, editor, *Manual of Phonetics*, pages 533–549, Amsterdam, 1968.

T. Dekens, M. Demol, W. Verhelst, and P. Verhoeve. A comparative study of speech rate estimation techniques. In *Proceedings of Interspeech 2007*, pages 510–513, Antwerp, Belgium, 2007. International Speech Communication Association.

V. Dellwo. Rhythm and speech rate: A variation coefficient for deltaC. In P. Karnowski and I. Szigeti, editors, *Language and Language-processing. Proceedings of the 38th linguistic Colloquium*, pages 231–241. Peter Lang, Frankfurt am Main, 2006.

V. Dellwo, I. Steiner, B. Aschenberner, J. Dankovičová, and P. Wagner. BonnTempo-corpus and BonnTempo-tools: A database for the study of speech rhythm and rate. In *Proceedings of Interspeech 2004*, pages 777–780, Jeju Island, Korea, 2004. International Speech Communication Association.

C. Draxler. Webtranscribe - an extensible web-based speech annotation framework. In *Proceedings of TSD 2005*, pages 61–68, 2005.

L. Folk and F. Schiel. The lombard effect in spontaneous dialog speech. In *Proceedings of Interspeech 2011*, pages 2701–2704, Florence, Italy, 2011. International Speech Communication Association.

C. Francu and C. G. Nevill-Manning. Distance metrics and indexing strategies for a digital library of popular music. In *Proceedings of IEEE International Conference on Multimedia and Expo*, volume 2, pages 889–892, 2000.

E. Grabe and E. L. Low. Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology*, 7:515–546, 2002.

M. Gubian, F. Torreira, H. Strik, and L. Boves. Functional data analysis as a tool for analyzing speech dynamics - A case study on the French word c'était. In *Proceedings of Interspeech 2009*, pages 2199–2202, Brighton, UK, 2009. International Speech Communication Association.

M. Gubian, F. Cangemi, and L. Boves. Automatic and data driven pitch contour manipulation with functional data analysis. In *Proceedings of the Fifth International Conference on Speech Prosody 2010*, 2010.

M. Gubian, L. Boves, and F. Cangemi. Joint analysis of f0 and speech rate with functional data analysis. In *Proceedings of ICASSP 2011*, pages 4972–4975, 2011.

M. S. Han. The feature of duration in Japanese. *Onsei no Kenkyuu [Study of Sounds]*, 10:65–80, 1962.

S. Handel. *Listening - an introduction to the perception of auditory events*. MIT Press, Cambridge, 1986.

J. H. L. Hansen and S. Patil. Speech under stress: Analysis, modeling and recognition. In C. Müller, editor, *Speaker Classification I, LNAI*, volume 4343, pages 108–137. Springer, New York, 2007.

J. Harrington. *Phonetic Analysis of Speech Corpora*. Wiley-Blackwell, 2010.

D. J. Hermes. Measuring the perceptual similarity of pitch contours. *Journal of Speech, Language & Hearing Research*, 41(1):73–82, 1998a.

D. J. Hermes. Auditory and visual similarity of pitch contours. *Journal of Speech, Language & Hearing Research*, 41(1):63–72, 1998b.

H. Hollien, K. Liljegren, C. A. Martin, and G. DeJong. Prediction of intoxication levels by speech analysis. In A. Braun, editor, *Advances in Phonetics*, volume 106, pages 40–50, Stuttgart, 1999. Steiner Verlag.

H. Hollien, G. DeJong, C. A. Martin, R. Schwartz, and K. Liljegren. Effects of ethanol intoxication on speech suprasegmentals. *Journal of the Acoustical Society of America*, 110(6):3198–3206, 2001.

K. Johnson, D. B. Pisoni, and R. H. Bernacki. Do voice recordings reveal whether a person is intoxicated? A case study. *Phonetica*, 47:215–237, 1990.

A. Kipp, M.-B. Wesenick, and F. Schiel. Pronunciation modeling applied to automatic segmentation of spontaneous speech. In *Proceedings of Eurospeech 1997*, pages 1023–1026, Rhodes, Greece, 1997.

S. Kitaazawa, H. Ichikawa, S. Kobayashi, and Y. Nishinuma. Extraction and representation rhythmic components of spontaneous speech. In *Proceedings of Eurospeech 1997*, pages 641–644, 1997.

E. Klabbers and J. P. H. van Santen. Clustering of foot-based pitch contours in expressive speech. In *ISCA Speech Synthesis Workshop 5*, pages 73–78, 2004.

F. Klingholz, R. Penning, and E. Liebhardt. Recognition of low-level alcohol intoxication from speech signal. *Journal of the Acoustical Society of America*, 84(3):929–935, 1988.

K. J. Kohler. *Einführung in die Phonetik des Deutschen*. Erich Schmidt, Berlin, 1977.

H. J. Künzel and A. Braun. The effect of alcohol on speech prosody. In *Proceedings of ICPhS 2003*, pages 2645–2648, 2003.

H. J. Künzel, A. Braun, and U. Eysholdt. *Einfluss von Alkohol auf Sprache und Stimme*. Kriminalistik-Verlag, Heidelberg, 1992.

P. Ladefoged. *A Course in Phonetics*. Harcourt Brace Jovanovich, New York, 1975.

V. L. Latsch and S. L. Netto. Pitch-synchronous time alignment of speech signals for prosody transplantation. In *International Symposium on Circuits and Systems (ISCAS)*, pages 2405–2408. IEEE, 2011.

I. Lehiste. Isochrony reconsidered. *Journal of Phonetics*, 5(3):253–263, 1977.

I. Lehiste. Phonetic investigation of metrical structure in orally produced poetry. *Journal of Phonetics*, 18:123–133, 1990.

M. Levit, R. Huber, A. Batliner, and E. Noeth. Use of prosodic speech characteristics for automated detection of alcohol intoxication. In M. Bacchiani, J. Hirschberg, D. Litman, and M. Ostendorf, editors, *Proceedings of the Workshop on Prosody in Speech Recognition and Understanding*, pages 103–106, Red Bank, NJ, 2001.

A. Loukina, G. Kochanski, C. Shih, E. Keane, and I. Watson. Rhythm measures with language-independent segmentation. In *Proceedings of Interspeech 2009*, pages 1531–1534, Brighton, UK, 2009. International Speech Communication Association.

L. Lu, H. You, and H.-J. Zhang. A new approach to query by humming in music retrieval. In *Proceedings of IEEE International Conference on Multimedia and Expo*, pages 595–598, 2001.

C. S. Martin and M. Yuchtman. Using speech as an index of alcohol-intoxication. *Research on Speech Perception*, 12:413–426, 1986.

F. Martinez, D. Tapias, and J. Alvarez. Towards speech rate independence in large vocabulary continuous speech recognition. In *Proceedings of ICASSP 1998*, pages 725–728, 1998.

C. Mathon and S. de Abreu. Emotion from speakers to listeners. Perception and prosodic characterization of affective speech. In C. Müller, editor, *Speaker Classification II, LNAI*, volume 4441, pages 70–82. Springer, New York, 2007.

G. Moehler. *Theoriebasierte Modellierung der deutschen Intonation für die Sprachsynthese*. PhD thesis, Universität Stuttgart, Stuttgart, Germany, 1998.

N. Morgan and E. Fosler-Lussier. Combining multiple estimators of speaking rate. In *Proceedings of ICASSP 1998*, volume 2, pages 729–732, 1998.

N. Morgan, E. Fosler, and N. Mirghafori. Speech recognition using on-line estimation of speaking rate. In *Proceedings of Eurospeech 1997*, pages 2079–2082, 1997.

E. Moulines and F. Charpentier. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, 9: 453–467, 1990.

S. Narayanan and D. Wang. Speech rate estimation via temporal correlation and selected sub-band correlation. In *Proceedings of ICASSP 2005*, pages 413–416, 2005.

O. Niebuhr. F0-based rhythm effects on the perception of local syllable prominence. *Phonetica*, 66:95–112, 2009.

T. Pfau and G. Ruske. Estimating the speaking rate by vowel detection. In *Proceedings of ICASSP 1998*, pages 945–948, 1998.

K. L. Pike. *The intonation of American English*. University of Michigan Press, Ann Arbor, Michigan, 2 edition, 1946.

J. Ramsay, G. Hooker, and S. Graves. *Functional Data Analysis with R and MATLAB*. Springer, New York, 2009.

F. Ramus, M. Nespor, and J. Mehler. Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3):265–292, 1999.

U. Reubold, J. Harrington, and F. Kleber. Vocal aging effects on f0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication*, 52: 638–651, 2010.

H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(1):43–49, 1978.

K. Schaefer-Vincent. Pitch period detection and chaining: Method and evaluation. *Phonetica*, 40:177–202, 1983.

F. Schiel. Automatic phonetic transcription of non-prompted speech. In *Proceedings of ICPhS 1999*, pages 607–610, 1999.

F. Schiel. Perception of alcoholic intoxication in speech. In *Proceedings of Interspeech 2011*, pages 3281–3284, Florence, Italy, 2011. International Speech Communication Association.

F. Schiel, C. Heinrich, and S. Barfüßer. Alcohol Language Corpus: the first public corpus of alcoholized German speech. *Language Resources and Evaluation*, 46 (3), 2012.

B. Schuller, S. Steidl, A. Batliner, F. Schiel, and J. Krajevski. The Interspeech 2011 speaker state challenge. In *Proceedings of Interspeech 2011*, pages 3201–3204, Florence, Italy, 2011. International Speech Communication Association.

I. Shmulevich. A note on the pitch contour similarity index. *Journal of New Music Research*, 33(1):17–18, 2004.

L. C. Sobell, M. B. Sobell, and R. F. Coleman. Alcohol-induced dysfluency in nonalcoholics. *Folia Phoniatrica*, 34:316–323, 1982.

S. Tilsen and K. Johnson. Low-frequency fourier analysis of speech rhythm. *Journal of the Acoustical Society of America*, 124(2):34–39, 2008.

F. Trojan and K. Kryspin-Exner. The decay of articulation under the influence of alcohol and paraldehyde. *Folia Phoniatrica*, 20:217–238, 1968.

G. Turco and M. Gubian. L1 prosodic transfer and priming effects: A quantitative study on semi-spontaneous dialogues. In *Proceedings of the Sixth International Conference on Speech Prosody 2012*, pages 386–389, 2012.

G. Turco, M. Gubian, and J. Schertz. A quantitative investigation of the prosody of verum focus in Italian. In *Proceedings of Interspeech 2011*, pages 961–964, Florence, Italy, 2011. International Speech Communication Association.

S. Ultes, A. Schmitt, and W. Minker. Attention, sobriety checkpoint! Can humans determine by means of voice, if someone is drunk... and can automatic classifiers compete? In *Proceedings of Interspeech 2011*, pages 3221–3224, Florence, Italy, 2011. International Speech Communication Association.

P. Wagner and V. Dellwo. Introducing YARD (yet another rhythm determination) and re-introducing isochrony to rhythm research. In *Proceedings of Speech Prosody 2004*, pages 227–230, Nara, Japan, 2004.

D. Wang and S. Narayanan. Robust speech rate estimation for spontaneous speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 15: 2190–2201, 2007.

H. Watanabe, T. Shin, H. Matsuo, F. Okuno, T. Tsuji, M. Matsuoka, J. Fakaura, and H. Matsunaga. Studies on vocal fold injection and changes in pitch associated with alcohol intake. *Journal of Voice*, 8(4):340–346, 1994.

P. E. Watson, W. I. D., and R. D. Batt. Total body water volumes for adult males and females estimated from simple anthropometric measurements. *The American journal of clinical nutrition*, 33(1):27–39, 1980.

L. White and S. L. Mattys. Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35:501–522, 2007.

E. M. P. Widmark. *Die theoretischen Grundlagen und die praktische Verwendbarkeit der gerichtlich-medizinischen Alkoholbestimmung*. Urban und Schwarzenberg, Berlin, Wien, 1932.

L. Wiget, L. White, B. Schuppler, I. Grenon, O. Rauch, and S. L. Mattys. How stable are acoustic metrics of contrastive speech rhythm. *Journal of the Acoustical Society of America*, 127(3):1559–1569, March 2010.

Z. Xie and P. Niyogi. Robust acoustic-based syllable detection. In *Proceedings of Interspeech 2006*, pages 1571–1574, Pittsburgh, Pennsylvania, USA, 2006. International Speech Communication Association.

S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, C. Busso, and Z. Deng. An acoustic study of emotions expressed in speech. In *Proceedings of Interspeech 2004*, pages 2193–2196, Jeju Island, Korea, 2004. International Speech Communication Association.

M. Zellers, M. Gubian, and B. Post. Redescribing intonational categories with functional data analysis. In *Proceedings of Interspeech 2010*, pages 1141–1144, Chiba, Japan, 2010. International Speech Communication Association.

Y. Zhang and J. R. Glass. Speech rhythm guided syllable nuclei detection. In *Proceedings of ICASSP 2009*, pages 3797–3800, 2009.

Y. Zhu and M. Kankanhalli. Similarity matching of continuous melody contours for humming querying of melody databases. In *Proceedings of IEEE Workshop on Multimedia Signal Processing*, pages 249–252, 2002.

Anhang A

Rhythmusparameter

A.1 Alternative Lauteinteilung

Die alternative Einteilung vokalischer und konsonantischer Laute (nach IPA) in die Kategorien C und V ist aus Tabelle A.1 ersichtlich.

V Elemente	aː,ɑː,eː,iː,oː,uː,ɛː,yː,øː,a,ɛ,i,o,u,e,ɔ,ɪ,ʊ,ə,ʏ,œ,ɐ aʊ,aɪ,ɔʏ,ɛɪ,ɛ̃ː,ɛ̃,ãː,ã,õː,m,n,ŋ,l,j
C Elemente	p,b,t,d,k,g,ʔ,f,v,ð,s,z,ʃ,ʒ,ç,x,h,r,ʊ

Tabelle A.1: *Alternative Einteilung vokalischer und konsonantischer Laute in die Kategorien C und V nach IPA.*

Auf Grund der Zuweisung des Approximanten /j/, des Laterals /l/ und der Nasale /m,n,ŋ/ zu den Vokalen bei der alternativen Lauteinteilung ergeben sich gegebenenfalls insgesamt mehr und längere vokalische Elemente V. Damit könnten sich die erwarteten Werte vor allem für die Rhythmusparameter, die auf den vokalischen Elementen beruhen, verändern. Die Variabilität der Dauern vokalischer Elemente steigt höchstwahrscheinlich, und damit liegen $\%V$, den Δ-Werten, den $Varco\Delta$-Werten, den PVI-Werten, den $YARD$-Werten und den Silbenraten-Werten veränderte Ausgangswerte bei der Berechnung zugrunde. Die alternative Lauteinteilung führte jedoch nicht zu qualitativ anderen Ergebnissen als die in Kapitel 3.1 verwendete und wurde deshalb dort nicht näher beschrieben. Dennoch sind die Ergebnisse im folgenden Abschnitt A.2 aufgeführt.

A.2 Alternative Lauteinteilung - Ergebnisse

Parameter	% ⇑ na zu a	% ⇓ na zu a	p-level	Sprechstil(e)
$\%V$	58.67	41.33	$p < 0.001$	r
ΔV	84.67	15.33	$p < 0.001$	r,s
ΔC	69.33	30.67	$p < 0.001$	r
ΔS	74	26	$p < 0.001$	r
ΔS_P	66	34	$p < 0.001$	r,s
ΔSN	68.67	31.33	$p < 0.001$	r
ΔSN_P	67.33	32.67	$p < 0.001$	r,s
$Varco\Delta V$	70.67	29.33	$p < 0.001$	r
$Varco\Delta C$	66	34	$p < 0.001$	r
$Varco\Delta S$	68	32	$p < 0.001$	r
$Varco\Delta S_P$	60.67	39.33	$p < 0.001$	r
$rPVI_V$	84	16	$p < 0.001$	r,s
$rPVI_C$	64.67	35.33	n.s.	-
$rPVI_S$	81.33	18.67	$p < 0.001$	r,s
$rPVI_{S_P}$	76	24	$p < 0.001$	r,s
$rPVI_{SN}$	67.33	32.67	$p < 0.001$	r,s
$rPVI_{SN_P}$	73.33	26.67	$p < 0.001$	r,s
$nPVI_V$	72	28	$p < 0.001$	r,s
$nPVI_C$	59.33	40.67	n.s.	-
$nPVI_S$	72	28	$p < 0.001$	r,s
$nPVI_{S_P}$	70.67	29.33	$p < 0.001$	r
$nPVI_{SN}$	59.33	40.67	n.s.	-
$nPVI_{SN_P}$	64	36	$p < 0.01$	r,s,c
$YARD$	44.67	55.33	$p < 0.01$	c
$YARD_P$	53.33	46.67	$p < 0.01$	r
SR	20	80	$p < 0.001$	r,s
SR_P	20	80	$p < 0.001$	r,s
$mean(P_{short})$	52	48	n.s.	-
$sd(P_{short})$	54.67	45.33	n.s.	-
$mean(P_{long})$	53.33	46.67	n.s.	-
$sd(P_{long})$	46	54	n.s.	-
SP	56	44	$p < 0.001$	r
LP	58	42	n.s.	-

Tabelle A.2: *Auswertungsergebnisse zu den Rhythmusparametern bei alternativer Lauteinteilung mit Prozentwerten der Sprecher, bei welchen sich der jeweilige Parameter bei a-Sprache gegenüber na-Sprache erhöht (% ⇑ na zu a) bzw. verringert (% ⇓ na zu a). Sprechstile r (gelesen), s (spontan) und c (Kommando). Ein P im Index des Parameters bedeutet Berechnung inklusive Pausen.*

A.3 Alternative Lauteinteilung - Ergebnisse Kontrollgruppe

Parameter	% ⇑ na zu a	% ⇓ na zu a	p-level	Sprechstil(e)
$\%V$	65	35	$p < 0.001$	r
ΔV	90	10	$p < 0.01$	r
ΔC	70	30	$p < 0.1$	r
ΔS	85	15	$p < 0.05$	r
ΔS_P	70	30	$p < 0.001$	r
ΔSN	85	15	$p < 0.05$	r
ΔSN_P	75	25	$p < 0.001$	r
$Varco\Delta V$	70	30	n.s.	-
$Varco\Delta C$	70	30	n.s.	-
$Varco\Delta S$	65	35	$p < 0.1$	r
$Varco\Delta S_P$	70	30	$p < 0.001$	r
$rPVI_V$	95	5	$p < 0.001$	r
$rPVI_C$	70	30	n.s.	-
$rPVI_S$	95	5	$p < 0.01$	r,s
$rPVI_{S_P}$	90	10	$p < 0.001$	r
$rPVI_{SN}$	85	15	$p < 0.01$	r,s
$rPVI_{SN_P}$	90	10	$p < 0.001$	r
$nPVI_V$	80	20	$p < 0.01$	r,s
$nPVI_C$	80	20	n.s.	-
$nPVI_S$	90	10	$p < 0.01$	r,s
$nPVI_{S_P}$	80	20	$p < 0.001$	r
$nPVI_{SN}$	80	20	n.s.	-
$nPVI_{SN_P}$	80	20	$p < 0.01$	r
$YARD$	50	50	n.s.	-
$YARD_P$	65	55	$p < 0.05$	c
SR	10	90	$p < 0.01$	r
SR_P	5	95	$p < 0.001$	r
$mean(P)$	70	30	$p < 0.05$	r
$sd(P)$	55	45	$p < 0.05$	r
SP	45	55	$p < 0.05$	r,s

Tabelle A.3: *Auswertungsergebnisse zu den Rhythmusparametern bei alternativer Lauteinteilung Kontrollgruppe alkoholisiert-nüchtern mit Prozentwerten der Sprecher, bei welchen sich der jeweilige Parameter bei a-Sprache gegenüber in na-Sprache erhöht (% ⇑ na zu a) bzw. verringert (% ⇓ na zu a). Sprechstile r (gelesen), s (spontan) und c (Kommando). Ein P im Index des Parameters bedeutet Berechnung inklusive Pausen.*

Parameter	% ⇑ na zu cna	% ⇓ na zu cna	p-level	Sprechstil(e)
$\%V$	60	40	$p < 0.05$	r
ΔV	55	45	n.s.	-
ΔC	50	50	n.s.	-
ΔS	65	35	n.s.	-
ΔS_P	60	40	n.s.	-
ΔSN	55	45	n.s.	-
ΔSN_P	55	45	n.s.	-
$Varco\Delta V$	55	45	n.s.	-
$Varco\Delta C$	60	40	n.s.	-
$Varco\Delta S$	55	45	n.s.	-
$Varco\Delta S_P$	55	45	n.s.	-
$rPVI_V$	65	35	$p < 0.1$	r
$rPVI_C$	55	45	n.s.	-
$rPVI_S$	60	40	n.s.	-
$rPVI_{S_P}$	60	40	n.s.	-
$rPVI_{SN}$	55	45	n.s.	-
$rPVI_{SN_P}$	70	30	n.s.	-
$nPVI_V$	50	50	n.s.	-
$nPVI_C$	50	50	n.s.	-
$nPVI_S$	65	35	n.s.	-
$nPVI_{S_P}$	65	35	n.s.	-
$nPVI_{SN}$	65	35	n.s.	-
$nPVI_{SN_P}$	70	30	n.s.	-
$YARD$	45	55	$p < 0.1$	r
$YARD_P$	60	40	n.s.	-
SR	25	75	$p < 0.05$	r
SR_P	35	65	$p < 0.1$	r
$mean(P)$	50	50	n.s.	-
$sd(P)$	40	60	n.s.	-
SP	60	40	n.s.	-

Tabelle A.4: *Auswertungsergebnisse zu den Rhythmusparametern bei alternativer Lauteinteilung Kontrollgruppe Kontrollgruppenaufnahmen-nüchtern mit Prozentwerten der Sprecher, bei welchen sich der jeweilige Parameter bei cna-Sprache gegenüber na-Sprache erhöht (% ⇑ na zu cna) bzw. verringert (% ⇓ na zu cna). Sprechstile r (gelesen), s (spontan) und c (Kommando). Ein P im Index des Parameters bedeutet Berechnung inklusive Pausen.*

Anhang B

ALC Metadaten Sprecher

SCD	SEX	AGE	ACC	WEI	HEI	PRO	SMO	DRH
006	F	26	BY	50	165	assessor	N	light
007	M	25	BY	68	172	assessor	N	light
008	M	29	BY	79	175	jurist	N	light
009	M	26	BY	75	180	assessor	Y	moderate
010	M	29	BY	70	183	jurist	N	light
011	F	27	BY	53	174	assessor	N	light
012	F	52	HE	54	163	senior prosecutor	N	moderate
013	M	59	NI	83	185	prosecutor	N	light
015	F	25	BW	57	165	student	N	moderate
016	M	25	BY	68	176	student	N	light
017	M	32	BY	73	188	research assistant	N	moderate
018	F	27	BW	65	177	assessor	Y	moderate
019	F	26	BY	58	168	assessor	N	light
020	M	32	BE	88	178	jurist	N	light
022	F	33	BY	69	165	employee	N	light
023	F	23	BY	56	173	paralegal	N	moderate
024	F	23	BY	52	165	jurist	N	light
025	F	24	BY	52	169	jurist	N	light
026	M	25	BY	72	180	assessor	N	moderate
027	M	28	BY	85	180	assessor	N	light
028	M	27	BY	71	183	assessor	N	moderate
029	F	26	BY	62	178	assessor	N	light
030	F	26	BW	53	167	assessor	N	moderate
031	F	27	BY	-	163	jurist	Y	light
032	M	29	BY	85	179	assessor	Y	moderate
034	M	26	BY	80	177	assessor	Y	light

SCD	SEX	AGE	ACC	WEI	HEI	PRO	SMO	DRH
035	F	26	NW	54	168	assessor	Y	moderate
036	F	27	BY	64	174	assessor	N	moderate
037	M	26	BY	75	179	assessor	N	light
038	F	28	BY	62	175	assessor	N	light
039	F	39	BY	55	160	judical clerk	N	moderate
040	F	27	BY	57	168	assessor	N	light
041	M	28	BY	65	179	assessor	N	moderate
042	F	36	BW	63	158	advocate	N	moderate
043	M	31	RP	103	194	apprentice	Y	moderate
046	F	24	BY	62	168	assessor	N	light
047	M	28	BY	79	182	assessor	Y	light
048	M	47	BY	100	183	policeman	N	light
049	F	25	BY	54	161	assessor	N	light
050	F	26	BY	57	166	assessor	N	light
051	M	30	BY	82	192	jurist	N	moderate
053	F	26	BY	60	170	assessor	Y	moderate
054	M	29	BY	80	174	assessor	Y	moderate
055	F	24	BY	56	165	assessor	N	moderate
057	M	27	BY	72	186	assessor	N	light
058	F	24	BY	63	170	student	N	heavy
059	F	27	BY	56	168	assessor	Y	moderate
060	F	27	BY	56	168	assessor	Y	moderate
061	F	57	BE	70	156	clerk	N	light
062	M	62	BY	72	175	judge	N	moderate
063	M	53	BY	88	180	judge	N	moderate
064	F	29	BY	90	172	assessor	Y	light
065	F	25	BY	66	175	assessor	N	moderate
069	M	31	BY	65	172	assessor	Y	moderate
070	M	28	BY	62	175	assessor	Y	moderate
071	F	26	BY	70	170	assessor	N	moderate
072	M	27	BY	70	172	assessor	N	moderate
073	M	25	BY	80	178	assessor	Y	heavy
074	F	27	BY	55	163	assessor	N	light
075	M	25	BY	68	170	assessor	Y	heavy
077	M	27	BY	84	180	assessor	Y	moderate
078	F	26	BY	62	164	assessor	N	moderate
079	F	27	BY	63	167	student	Y	light
080	M	28	HE	77	172	assessor	N	light
081	M	27	BY	75	195	assessor	N	moderate
082	F	25	BY	70	173	assessor	N	moderate
083	M	27	BY	70	186	assessor	Y	moderate

SCD	SEX	AGE	ACC	WEI	HEI	PRO	SMO	DRH
086	M	33	BY	90	188	assessor	N	moderate
087	M	52	BY	90	186	detective	N	moderate
088	M	27	BY	63	174	assessor	N	light
089	M	26	BY	80	183	assessor	N	moderate
090	F	26	BY	53	161	assessor	N	light
091	F	26	BY	80	175	assessor	N	moderate
092	F	28	BY	62	168	housewife	Y	moderate
094	M	27	BY	85	196	jurist	N	moderate
095	M	28	BY	80	198	assessor	N	moderate
096	M	50	BY	70	179	judge	N	moderate
097	M	62	BY	88	182	judge	N	heavy
098	M	59	BY	83	184	judge	N	light
099	F	54	NW	63	172	industrial clerk	N	light
100	F	37	BY	55	162	employee	N	light
101	M	50	NI	71	181	employee	N	moderate
102	M	35	BY	76	176	engineer	N	light
103	M	48	BY	79	185	lecturer	N	moderate
500	F	25	BW	55	167	jurist	N	moderate
501	M	28	BY	65	178	assessor	N	moderate
502	F	26	BY	55	173	assessor	N	light
504	F	27	BY	75	173	assessor	N	light
505	F	30	BY	57	168	prosecutor	N	moderate
506	F	25	NW	57	173	jurist	N	light
507	F	29	BY	62	172	jurist	N	light
508	M	38	BY	77	180	jurist	N	moderate
509	M	30	BY	70	179	prosecutor	N	light
511	M	31	BY	68	180	jurist	N	light
513	F	28	BY	57	169	assessor	N	light
514	M	26	HE	70	177	jurist	N	light
515	M	28	BE	80	180	jurist	N	moderate
517	F	21	XX	50	170	paralegal	N	light
518	M	30	BY	77	189	assessor	N	moderate
520	F	27	BY	55	170	jurist	Y	light
521	F	25	SN	78	163	assessor	N	light
522	F	25	BW	72	172	assessor	N	light
523	M	26	BY	90	196	jurist	Y	moderate
524	M	27	HE	78	171	jurist	N	moderate
525	F	28	IT	54	163	jurist	N	moderate
526	M	29	BW	95	192	assessor	Y	moderate
527	M	26	NI	85	195	assessor	N	heavy
528	M	25	BY	80	184	jurist	N	light

SCD	SEX	AGE	ACC	WEI	HEI	PRO	SMO	DRH
529	F	25	BY	59	167	jurist	Y	moderate
530	F	27	BY	75	182	assessor	N	light
531	M	26	NW	105	193	jurist	N	moderate
532	M	26	SH	92	186	jurist	N	light
533	M	29	BY	85	180	assessor	Y	light
534	M	38	SA	81	172	constable	Y	moderate
536	M	30	BY	70	177	jurist	N	moderate
537	M	26	SR	72	186	jurist	N	light
541	F	26	BW	65	170	jurist	Y	moderate
542	F	28	HE	53	165	jurist	N	moderate
543	F	27	BY	58	165	jurist	N	light
544	M	27	BY	82	183	jurist	N	moderate
545	M	50	BY	84	180	policeman	N	moderate
546	F	25	BY	52	161	jurist	N	light
547	F	26	BY	64	175	jurist	N	light
548	F	25	RP	54	165	assessor	Y	light
549	F	26	BY	58	166	jurist	N	light
550	M	28	NS	80	186	jurist	Y	moderate
551	M	36	BY	90	180	jurist	N	light
552	F	25	BW	64	174	jurist	N	moderate
554	F	26	BY	48	160	jurist	N	light
555	M	28	BY	83	180	assessor	N	light
556	M	24	BY	80	190	medical scientist	N	heavy
557	F	28	NI	65	178	assessor	N	light
558	F	26	NW	60	172	jurist	N	moderate
560	F	32	BY	55	168	jurist	Y	light
561	F	56	BY	63	164	judge	N	moderate
562	M	52	BY	80	180	judge	N	moderate
563	M	41	BY	96	178	prosecutor	N	moderate
565	M	26	BY	78	175	jurist	N	moderate
568	M	26	NS	70	173	jurist	Y	moderate
569	F	25	BY	58	165	jurist	N	light
570	M	28	BW	65	173	jurist	N	moderate
571	M	26	BY	70	176	jurist	N	light
572	M	31	BY	98	188	jurist	N	moderate
573	M	27	BY	86	186	jurist	Y	heavy
574	F	25	BY	53	160	jurist	N	light
575	M	28	BY	78	188	jurist	N	moderate
576	F	25	BY	60	162	assessor	Y	moderate
577	F	25	BY	65	170	jurist	N	moderate
578	F	25	BW	74	172	assessor	N	light

SCD	SEX	AGE	ACC	WEI	HEI	PRO	SMO	DRH
581	M	50	BY	92	189	detective	Y	moderate
582	M	44	BY	72	181	detective	N	moderate
583	F	29	BY	75	178	jurist	N	moderate
584	M	26	BY	87	181	jurist	Y	heavy
585	F	25	BY	52	163	jurist	N	light
586	M	27	BY	73	180	jurist	N	light
587	M	29	BY	80	191	chemist	N	moderate
589	F	27	NW	85	168	jurist	N	light
590	M	28	BY	85	178	assessor	N	light
591	M	44	BY	68	172	jurist	N	light
594	F	59	BY	70	178	secretary	N	light
595	F	64	BY	73	163	housewife	N	moderate
596	M	44	BY	72	170	social pedagogue	N	heavy

Tabelle B.1: *Metadaten aller 162 Sprecher des ALC mit Kennzeichnung der Kontrollgruppensprecher in grau. SCD = Sprechercode, SEX = Geschlecht, AGE = Alter, ACC = Deutscher Akzent, WEI = Gewicht, HEI = Größe, PRO = Beruf, SMO = Raucher, DRH = Trinkverhalten; Detailbeschreibung in Schiel et al. [2012]*

Anhang C

ALC gleichlautende Aufnahmeelemente

Die folgenden 19 gelesenen Aufnahmeelemente sind sowohl in nüchternem Zustand der Sprecher als auch unter Alkoholeinfluss und ggf. bei der Kontrollgruppe in gleichem Wortlaut vorhanden.

1. Bitte lesen Sie die Telefonnummer: +491763582901
2. Bitte so schnell wie möglich lesen: Messwechsel, Wachsmaske, Wachsmaske, Messwechsel
3. Bitte Telefonnummer lesen: 0862359286
4. Bitte so schnell wie möglich lesen: Bemoost wächst nächst dem Strom ein Stamm, feststämmig stolz strebt sein Geäst stromwärts, und weist nach Ost und West.
5. Bitte lesen Sie die Adresse: Sportplatzweg 27, Marktgraitz
6. Bitte lesen Sie die Kreditkartennummer: 1390 7516 0281 9357
7. Bitte lesen Sie das Autokennzeichen: STA-PB 2759

8. Bitte lesen: Backer, Muselmann, Menschen, Massen, Morder, Mohren, Mutter, Manumentenmacher

9. Bitte lesen Sie die Adresse: 18546 Sassnitz

10. Bitte lesen: 06271 57390

11. Bitte so schnell wie möglich lesen: Die Köchin mit dem Tupfenkopftuch kocht Karpfen in dem Kupferkochtopf.

12. Bitte Adresse lesen: Schwester-Hermenegildis-Straße

13. Bitte Adresse lesen: Madapaka-Betegindis-Straße 77 B

14. Bitte so schnell wie möglich lesen: Ketzer Krächzer petzten jetzt kläglich, letztlich plötzlich leicht skeptisch.

15. Bitte Steuerbefehl lesen: Temperatur 23 °C

16. Bitte Steuerbefehl lesen: nächster Titel

17. Bitte Steuerbefehl lesen: Frequenz 92,2 MHz

18. Bitte Steuerbefehl lesen: Autobahnen meiden

19. Bitte buchstabieren Sie: M A R K T G R A I T Z

Anhang D

Segmentlisten

Listing D.1: Eintrag alcsegsall

```
           utt    start       end                                utts spn item
5634 0772060024 716.667 4686.667 BLOCK20:SES2060:0772060024_h_00 077  024
     o_item       o_utt alc sex age acc       drh      aak      bak ges ces wea
5634     017 0771078017  na   M  27  BY moderate 0.00000 0.00000  f1   r1 SUN
                 irreg anncom specom type content f0_uttlist
5634 0|0|0|1|0|0|0|0|0    null    null    R       A          *
```

Eine detaillierte Beschreibung der Metadaten findet sich in Schiel et al. [2012].

Listing D.2: Ausschnitt phonsegs

```
segment  list from database:  ALC
query was:  phonetic!=xxx
       labels     start       end                                utts
504007    &p:     0.000   716.667 BLOCK20:SES2060:0772060024_h_00
504008     S    716.667   856.667 BLOCK20:SES2060:0772060024_h_00
504009     v    856.667   926.667 BLOCK20:SES2060:0772060024_h_00
504010     E    926.667  1016.667 BLOCK20:SES2060:0772060024_h_00
504011     s   1016.667  1126.667 BLOCK20:SES2060:0772060024_h_00
504012     t   1126.667  1196.667 BLOCK20:SES2060:0772060024_h_00
504013     6   1196.667  1476.667 BLOCK20:SES2060:0772060024_h_00
504014    &p:  1476.667  2386.667 BLOCK20:SES2060:0772060024_h_00
504015     h   2386.667  2416.667 BLOCK20:SES2060:0772060024_h_00
504016     E   2416.667  2486.667 BLOCK20:SES2060:0772060024_h_00
504017     6   2486.667  2636.667 BLOCK20:SES2060:0772060024_h_00
504018     m   2636.667  3046.667 BLOCK20:SES2060:0772060024_h_00
504019     e   3046.667  3206.667 BLOCK20:SES2060:0772060024_h_00
504020     n   3206.667  3256.667 BLOCK20:SES2060:0772060024_h_00
504021     @   3256.667  3296.667 BLOCK20:SES2060:0772060024_h_00
504022     g   3296.667  3506.667 BLOCK20:SES2060:0772060024_h_00
504023     I   3506.667  3566.667 BLOCK20:SES2060:0772060024_h_00
504024     l   3566.667  3756.667 BLOCK20:SES2060:0772060024_h_00
504025     I   3756.667  3896.667 BLOCK20:SES2060:0772060024_h_00
504026     s   3896.667  4016.667 BLOCK20:SES2060:0772060024_h_00
504027     S   4016.667  4086.667 BLOCK20:SES2060:0772060024_h_00
504028     t   4086.667  4116.667 BLOCK20:SES2060:0772060024_h_00
504029     r   4116.667  4206.667 BLOCK20:SES2060:0772060024_h_00
504030    a:   4206.667  4376.667 BLOCK20:SES2060:0772060024_h_00
504031     s   4376.667  4516.667 BLOCK20:SES2060:0772060024_h_00
504032     @   4516.667  4686.667 BLOCK20:SES2060:0772060024_h_00
504033    &p:  4686.667  6736.803 BLOCK20:SES2060:0772060024_h_00
```

Listing D.3: Ausschnitt cvsegs

```
segment list from database: ALC
query was: phonetic!=xxx
       labels      start       end                                utts
366212    &p: 10437.211  716.667 BLOCK20:SES2060:0772060024_h_00
366213      C   716.667  926.667 BLOCK20:SES2060:0772060024_h_00
366214      V   926.667 1016.667 BLOCK20:SES2060:0772060024_h_00
366215      C  1016.667 1196.667 BLOCK20:SES2060:0772060024_h_00
366216      V  1196.667 1476.667 BLOCK20:SES2060:0772060024_h_00
366217    &p:  1476.667 2386.667 BLOCK20:SES2060:0772060024_h_00
366218      C  2386.667 2416.667 BLOCK20:SES2060:0772060024_h_00
366219      V  2416.667 2636.667 BLOCK20:SES2060:0772060024_h_00
366220      C  2636.667 3046.667 BLOCK20:SES2060:0772060024_h_00
366221      V  3046.667 3206.667 BLOCK20:SES2060:0772060024_h_00
366222      C  3206.667 3256.667 BLOCK20:SES2060:0772060024_h_00
366223      V  3256.667 3296.667 BLOCK20:SES2060:0772060024_h_00
366224      C  3296.667 3506.667 BLOCK20:SES2060:0772060024_h_00
366225      V  3506.667 3566.667 BLOCK20:SES2060:0772060024_h_00
366226      C  3566.667 3756.667 BLOCK20:SES2060:0772060024_h_00
366227      V  3756.667 3896.667 BLOCK20:SES2060:0772060024_h_00
366228      C  3896.667 4206.667 BLOCK20:SES2060:0772060024_h_00
366229      V  4206.667 4376.667 BLOCK20:SES2060:0772060024_h_00
366230      C  4376.667 4516.667 BLOCK20:SES2060:0772060024_h_00
366231      V  4516.667 4686.667 BLOCK20:SES2060:0772060024_h_00
```

Anhang E

Danksagung

Ich möchte mich recht herzlich bei Herrn Florian Schiel bedanken, der immer für Fragen und Anregungen zur Verfügung stand, weiterhin für seine stets hilfreiche Fachkenntnis und vor allem sein Interesse am behandelten Thema. Mein Dank gilt ebenfalls der Deutschen Forschungsgemeinschaft, die mir durch ihre finanzielle Unterstützung im Rahmen des DFG Projektes SCHI 1117/1-1 diese Arbeit ermöglicht hat. Vielen Dank!