
Akustische und perzeptive Analysen von Sprache unter Alkoholeinfluss

Barbara Susanne Baumeister

Inaugural-Dissertation

zur Erlangung des Doktorgrades der Philosophie

an der Ludwig–Maximilians–Universität

München

vorgelegt von

Barbara Susanne Baumeister

aus Duisburg

München 2014

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

Korreferent: Prof. Dr. Jonathan M. Harrington

Tag der mündlichen Prüfung: 29.01.2015

Abstract

In this thesis, several acoustic and perceptive studies on the influence of alcoholic intoxication on speech are presented. The speech samples for the experiments are taken from the Alcohol Language Corpus, a large database containing recordings of 162 (77 female and 85 male) speakers in both sober and intoxicated conditions with various blood alcohol concentrations. An acoustic analysis of recordings of 148 speakers in intoxicated and sober condition reveals a significant increase of the mean and also of the interquartile ranges of the fundamental frequency in the majority of speakers. A large inter-speaker variability was found, but no correlation with their individual blood alcohol concentration. An analysis of the microperturbation parameters jitter and shimmer in the vowels /a:/ and /i:/ showed no consistent results and a large variance among the data. Jitter is significantly lower in intoxicated condition but only in the vowel /a:/. Shimmer is found to be significantly higher in intoxicated condition but only for male speakers in the vowel /a:/ and in read speech of male speakers in the vowel /i:/. Additionally, a forced choice discrimination test was conducted with speech samples of 161 speakers. 71 naive listeners reached an average discrimination rate of 63,1% which is well above chance. No gender-dependent effects could be found. The ability of the listeners to pick the correct stimulus was higher for speakers with high blood alcohol concentrations, with no connection to the drinking habits of the speakers. To see, whether fundamental frequency also functions as a perceptual cue to reveal a person's intoxication solely by the speech signal, the results of the perception test were compared with the acoustic measurements of fundamental frequency of the speakers. Those who tend to use higher fundamental frequency in intoxicated condition were judged correctly more often. Similar results could be found for speakers who use larger pitch ranges. The listeners also showed a general preference for the stimulus with higher fundamental frequency and higher pitch range. To further test fundamental frequency as a perceptual cue, two perception tests with manipulated stimuli were conducted. In the first one, the fundamental frequency effects in the stimulus in intoxicated condition were compensated by adjusting mean fundamental frequency and the interquartile range of the intoxicated stimulus to that of the sober stimulus in each discrimination pair. In the second one, fundamental frequency effects were simulated in a sober stimulus and compared to another sober stimulus in one pair. The simulation of fundamental frequency effects was performed by up-shifting and stretching the contour by 5%, according to the findings in the acoustic study. In the test with compensated fundamental frequency effects, listeners did not perform worse than in the basis test. In the second test with simulated fundamental frequency effects, they show a slight tendency to pick the manipulated stimulus as the intoxicated one. Fundamental frequency seems not to function as a strong perceptual cue for listeners in revealing a person's intoxication, maybe changes of fundamental frequency occur as a side-effect of other effects that play the major role. These could be due to other acoustic features and/or maybe linguistic or paralinguistic information. Listeners may use fundamental frequency as kind of a fall-back feature if no other signs of intoxication are present as is the case in the test with simulated effects. Probably listeners do not rely on fundamental frequency as a perceptual cue, because it is also prone to changes caused by other speaker states.

Inhalt

1	Einleitung	10
2	Das Alcohol Language Corpus	15
3	Akustische Veränderungen in der Sprache alkoholisierter Sprecher	20
3.1	Literaturüberblick	20
3.2	Langzeitanalyse der Grundfrequenz	33
3.2.1	Methode	33
3.2.2	Hypothesen	35
3.2.3	Ergebnisse	35
3.2.3.1	F ₀ -Median	35
3.2.3.2	F ₀ -Interquartilsabstand	37
3.2.3.3	Zusammenhang mit der Blutalkoholkonzentration .	40
3.2.3.4	Zusammenhang zwischen F ₀ -Median und F ₀ -IQA .	44
3.2.4	Zwischendiskussion	46
3.3	Jitter und Shimmer	50
3.3.1	Methode	50
3.3.2	Hypothesen	57
3.3.3	Ergebnisse	57
3.3.3.1	Jitter	57
3.3.3.2	Shimmer	59
3.3.4	Zwischendiskussion	61
4	Perzeption der Sprache alkoholisierter Sprecher	64
4.1	Literaturüberblick	64

4.2	Diskriminationstest	79
4.2.1	Methode	79
4.2.1.1	Stimuli	79
4.2.1.2	Hörer	82
4.2.1.3	Durchführung	82
4.2.2	Hypothesen	83
4.2.3	Ergebnisse	84
4.2.3.1	Erkennungsleistung der Hörer und Erkennungsraten der Sprecher	84
4.2.3.2	Nicht-akustische Einflussfaktoren auf die Hörerent- scheidung	87
4.2.3.3	Einfluss der F_0 -Parameter auf die Hörerentscheidung	88
4.2.4	Zwischendiskussion	93
4.3	Diskriminationstests mit manipulierter Grundfrequenz	100
4.3.1	Methode	100
4.3.1.1	Stimuli	101
4.3.1.2	Kompensation der F_0 -Effekte – Test 1	101
4.3.1.3	Simulation der F_0 -Effekte – Test 2	106
4.3.2	Hypothesen	106
4.3.3	Ergebnisse	107
4.3.3.1	Kompensation der F_0 -Effekte – Test 1	107
4.3.3.2	Simulation der F_0 -Effekte – Test 2	110
4.3.4	Zwischendiskussion	112
5	Zusammenfassende Diskussion und Ausblick	115
	Literaturverzeichnis	121
A	Danksagung	125

Abbildungsverzeichnis

3.1	Verteilung der Blutalkoholkonzentrationen	33
	(a) Männliche Sprecher	33
	(b) Weibliche Sprecher	33
3.2	Verteilung des Alters der Sprecher	34
3.3	F ₀ -Median	36
	(a) Verteilung der F ₀ -Mediane	36
	(b) Sprecherindividuelle Veränderungen des F ₀ -Medians über alle Sprachstile	36
3.4	F ₀ -Interquartilsabstand	38
	(a) Verteilung der F ₀ -Interquartilsabstände	38
	(b) Sprecherindividuelle Veränderungen des F ₀ -Interquartilsabstands in <i>gelesener Sprache</i>	38
3.5	F ₀ -Interquartilsabstand	39
	(a) Sprecherindividuelle Veränderungen des F ₀ -Interquartilsabstands in <i>Spontansprache</i>	39
	(b) Sprecherindividuelle Veränderungen des F ₀ -Interquartilsabstands in <i>Kommandosprache</i>	39
3.6	Streudiagramme der Veränderung des F ₀ -Medians und der BAK	42
3.7	Streudiagramme der Veränderung des F ₀ -IQA und der BAK	43
3.8	Streudiagramme der Veränderung des F ₀ -Medians und der Verände- rung des F ₀ -IQA	45
3.9	Beispiel zur Berechnung des Jitter (<i>RAP</i>) anhand eines kurzen Signal- stücks	54
3.10	Beispiel zur Berechnung des Shimmer (<i>APQ3</i>) anhand eines kurzen Signalstücks	55
3.11	Jitter (<i>RAP</i>) getrennt nach Alkoholisierung und Sprachstil	58
	(a) Weibliche Sprecher – Vokal /a:/	58

(b)	Männliche Sprecher – Vokal /a:/	58
3.12	Jitter (<i>RAP</i>) getrennt nach Alkoholisierung und Sprachstil	59
(a)	Weibliche Sprecher – Vokal /i:/	59
(b)	Männliche Sprecher – Vokal /i:/	59
3.13	Shimmer (<i>APQ3</i>) getrennt nach Alkoholisierung und Sprachstil	60
(a)	Weibliche Sprecher – Vokal /a:/	60
(b)	Männliche Sprecher – Vokal /a:/	60
3.14	Shimmer (<i>APQ3</i>) getrennt nach Alkoholisierung und Sprachstil	61
(a)	Weibliche Sprecher – Vokal /i:/	61
(b)	Männliche Sprecher – Vokal /i:/	61
4.1	Ergebnisse des Diskriminationstests	85
(a)	Erkennungsleistung der <i>Hörer</i>	85
(b)	Erkennungsrate der <i>Sprecher</i>	85
4.2	Anzahl richtiger Antworten in den unterschiedlichen Sprachstilen	86
4.3	Einfluss der Blutalkoholkonzentration der Sprecher	87
(a)	Zusammenhang von Erkennungsrate und BAK der Sprecher	87
(b)	BAK der Sprecher in den Zuordnungen der Hörer	87
4.4	Veränderung des F_0 -Medians	89
(a)	Veränderung des F_0 -Medians der Sprecher in den Stimuluspaaren	89
(b)	Zusammenhang von Erkennungsrate und Veränderung des F_0 - Medians	89
4.5	Veränderung des F_0 -IQA	90
(a)	Veränderung des F_0 -IQA der Sprecher in den Stimuluspaaren	90
(b)	Zusammenhang von Erkennungsrate und Veränderung des F_0 -IQA	90
4.6	Veränderungen der F_0 -Parameter der Sprecher in den Zuordnungen der Hörer	91
(a)	F_0 -Median	91
(b)	F_0 -IQA	91
4.7	Präferenz für den Stimulus mit höherem F_0 -Median	92

(a)	Verteilung der Präferenz der 72 Hörer für den Stimulus mit höherem F_0 -Median	92
(b)	Zusammenhang von Erkennungsleistung und Präferenz der 72 Hörer für den Stimulus mit höherem F_0 -Median	92
4.8	Präferenz für den Stimulus mit höherem F_0 -IQA	93
(a)	Verteilung der Präferenz der 72 Hörer für den Stimulus mit höherem F_0 -IQA	93
(b)	Zusammenhang von Erkennungsleistung und Präferenz der 72 Hörer für den Stimulus mit höherem F_0 -IQA	93
4.9	Beseitigung von Oktavfehlern	102
(a)	F_0 -Kontur mit Oktavfehler vor Anwendung des Algorithmus . .	102
(b)	F_0 -Kontur mit Interpolation statt Oktavfehler nach Anwendung des Algorithmus	102
4.10	Probleme in der Oktavfehlerkorrektur	103
(a)	F_0 -Kontur mit Oktavfehler vor Anwendung des Algorithmus . .	103
(b)	Beispiel für einen Fehler in der F_0 -Kontur nach Anwendung des Algorithmus	103
4.11	F_0 -Konturen der Äußerung „Temperatur 23°C“ eines männlichen Sprechers	104
(a)	alkoholisiert	104
(b)	nüchtern	104
4.12	F_0 -Konturen des a -Stimulus unverändert und nach der Manipulation .	105
4.13	F_0 -Median-Unterschiede	108
(a)	Verteilung der F_0 -Median-Unterschiede eines Stimuluspaares .	108
(b)	Zusammenhang der F_0 -Median-Unterschiede und der Erkennungsrate der Sprecher	108
4.14	Vergleich der Erkennungsleistung der Hörer	109
(a)	Test mit Originalstimuli	109
(b)	Kompensation der F_0 -Effekte	109

4.15	Vergleich der Erkennungsrate der Sprecher	110
	(a) Test mit Originalstimuli	110
	(b) Kompensation der F_0 -Effekte	110
4.16	Vergleich der Erkennungsleistung der Hörer	111
	(a) Test mit Originalstimuli	111
	(b) Simulation der F_0 -Effekte	111
4.17	Vergleich der Erkennungsrate der Sprecher	112
	(a) Test mit Originalstimuli	112
	(b) Simulation der F_0 -Effekte	112

Tabellenverzeichnis

2.1	Einteilung der Sprecher in Gruppen nach Trinkgewohnheiten	19
3.1	Überblick über die bisherigen Forschungsergebnisse zu akustischen Veränderungen in Sprache unter Alkoholeinfluss	31
3.2	Ergebnisse der linearen Regressionen der Veränderung des F_0 -Medians und der BAK getrennt nach Sprechergeschlecht und Sprachstil	40
3.3	Ergebnisse der linearen Regressionen der Veränderung des F_0 -IQA und der BAK getrennt nach Sprechergeschlecht und Sprachstil	41
3.4	Ergebnisse der linearen Regressionen der Veränderung des F_0 -Medians und des F_0 -IQA getrennt nach Sprechergeschlecht und Sprachstil	44
3.5	Übersicht über die Hypothesentests	47
3.6	Ergebnisse der <i>MEM</i> -Analyse – Jitter (<i>RAP</i>)	59
3.7	Ergebnisse der <i>MEM</i> -Analyse und des <i>post-hoc</i> -Tests für den Vokal <i>/i:/</i> bei männlichen Sprechern – Shimmer (<i>APQ3</i>)	62
3.8	Übersicht über die Hypothesentests	63
4.1	Übersicht über die bisherigen Perzeptionsexperimente	75
4.2	Übersicht der Hörer	82
4.3	Ergebnisse nach Geschlecht	86
4.4	Nicht-akustische Einflussfaktoren auf die Antwort der Hörer im Dis- kriminierungstest	88
4.5	Ergebnisse des <i>post-hoc</i> -Tests für die unterschiedlichen Sprachstile	88
4.6	Einflussfaktoren (F_0) auf die Antwort der Hörer im Diskriminationstest	91
4.7	Ergebnisse der linearen Regression der Veränderung des F_0 -IQA und der Erkennungsrate	91
4.8	Diskriminationsraten von Teilanalysen zur Vergleichbarkeit mit vorherigen Studien	96
4.9	Übersicht über die Hypothesentests	99

4.10 Übersicht der Hörer in den Tests mit manipulierten Stimuli	101
4.11 Einflussfaktoren auf die Antwort der Hörer in Test 1	110
4.12 Einflussfaktoren auf die Antwort der Hörer in Test 2	112
4.13 Übersicht über die Hypothesentests	114

1

Einleitung

Die chemische Verbindung Ethanol (C_2H_6O), im allgemeinen Sprachgebrauch auch als Alkohol bezeichnet, ist als Bestandteil von Getränken ein beliebtes und weitverbreitetes Genussmittel. Häufig treffen wir daher im Alltag auf mehr oder weniger stark alkoholisierte Personen – es gibt wohl kaum einen Erwachsenen in Deutschland, der noch nie mit einem Betrunkenen gesprochen hat. In den meisten Fällen glauben wir von uns, beurteilen zu können, ob unser Gegenüber alkoholisiert ist oder nicht. Neben zum Beispiel motorischen Beeinträchtigungen und dem Alkoholgeruch spielt auch die Sprache als Merkmal der Alkoholisierung eine Rolle. Dabei ist nicht nur der Inhalt des Gesagten von Bedeutung, sondern auch die akustische Realisierung. Die verwaschene Aussprache von alkoholisierten Personen wird im Alltag oft als *Lallen* bezeichnet und so auch von Schauspielern gerne als solches imitiert. Was genau in der Produktion der Sprache von Alkohol beeinflusst wird ist allerdings nicht restlos erforscht.

Das lautsprachliche Verhalten gehört zu den komplexesten und störungsanfälligsten intellektuellen Leistungen eines Menschen (Künzel et al. 1992). Aber welche akuten Auswirkungen auf Sprache und Stimme hat der Konsum von Alkohol tatsächlich? Um der Antwort auf diese Frage näherzukommen, soll zunächst der Einfluss von Alkohol auf den menschlichen Körper und das zentrale Nervensystem (ZNS) beleuchtet werden.

Die Alkoholisierung durch das Trinken von alkoholhaltigen Getränken beginnt mit der Aufnahmephase, die bereits zu einem geringen Teil über die Mundschleimhaut einsetzt. Die Aufnahme des wasserlöslichen Alkohols erfolgt weiterhin über verschiedene Organe, hierbei wird die Verteilung durch den Wassergehalt der Schleimhäute und Organe bestimmt. Das Verhältnis der Aufnahme zwischen den hauptsächlich aufnehmenden Organen Magen und Darm beträgt 20% (Magen) zu 80% (Darm). Wie schnell und wieviel Alkohol aus der Trinkmenge aufgenommen wird hängt allerdings von vielen verschiedenen Faktoren ab, zum Beispiel dem Mageninhalt, der Art des Getränks und möglichen Wechselwirkungen mit anderen Drogen oder Medikamenten. Aufgrund der vielen Einflussfaktoren im Magen-Darm-Trakt lässt sich die Blutalkoholkonzentration (BAK) bei oral aufgenommenem Alkohol anhand der konsumierten Menge schwieriger voraussagen als z.B. bei intravenös aufgenommenem Alkohol (Chin & Pisoni 1997). Auf die Resorptionsphase folgt die Distributionsphase, die sich teilweise mit der Resorptionsphase überlappt und nach ca. ein bis zwei Stunden abgeschlossen ist. In der Distributionsphase verteilt sich der Alkohol gleichmäßig in der Körperflüssigkeit und gelangt über das arterielle Blut in das Gewebe und von dort aus über die Kapillaren in das venöse Blut (Künzel et al. 1992). In Leber, Lunge, Nieren und dem Gehirn ist die Distribution relativ schnell abgeschlossen, da diese stark durchblutet bzw. sehr wasserhaltig sind, im Gegensatz zu Muskeln in denen sich der Alkohol wesentlich langsamer verteilt und Knochen und Fett, die generell nur sehr wenig Alkohol aufnehmen. Die Elimination des Alkohols geschieht hauptsächlich durch enzymatischen Abbau über die Leber, nur ein geringer Teil wird unverändert über Nieren, Lunge oder Haut ausgeschieden (Chin & Pisoni 1997).

Alkohol wirkt im zentralen Nervensystem wie ein Sedativum und stört den Erregungsablauf. Ab einer BAK von ca. 1‰ wird die Nervenleitgeschwindigkeit reduziert und es treten Veränderungen im EEG (Elektroenzephalogramm¹) auf, Durchblutungsstörungen im Gehirn sind ab ca. 2‰ nachweisbar (Künzel et al. 1992). Aber auch schon bei geringeren Dosen sind Veränderungen in psychomotorischen Fähigkeiten

¹ grafische Darstellung der elektrischen Aktivität des Gehirns

erkennbar. Zunächst lässt sich allerdings bei geringer, aber steigender BAK eine Leistungssteigerung und stimulierende Wirkung erkennen – vor allem bei Menschen, die von Natur aus eher nervös oder ängstlich sind – ebenso kommt es zu Gefühlen von Euphorie und Entspannung. Bei stärkerer Alkoholisierung setzt dann aber der sedative Effekt ein. Ab einer sehr hohen BAK von ca. 3‰ können Benommenheit, Schläfrigkeit und komatöse Zustände eintreten und schließlich ab ca. 4‰ auch der Tod durch Lungenversagen (Chin & Pisoni 1997).

Ein besonderer Fall, in dem eine Verschlechterung der Psychomotorik fatale Konsequenzen haben kann, ist das Autofahren. In Deutschland gilt im Allgemeinen die 0,5‰-Grenze, die das Autofahren mit geringerer Alkoholisierung erlaubt. Verschiedene Studien, die in Chin & Pisoni (1997) zusammengefasst werden, konnten im Zusammenhang mit Alkoholisierung z.B. Beeinträchtigungen in der visuellen und auditiven Wahrnehmung, eine Verlängerung der Reaktionszeit (vor allem in stark alkoholisiertem Zustand) und Reduktion der Geschwindigkeit und Genauigkeit in der Ausführung von verschiedenen Handlungen nachweisen, außerdem Probleme in der motorischen Koordination. Eine Beeinträchtigung dieser Fähigkeiten erschwert das Autofahren und kann die alkoholisierte Person und eventuell andere Teilnehmer des Straßenverkehrs in erhebliche Gefahr bringen. Problematisch ist hier die biphasische Wirkung des Alkohols, die eine Person in der Anfangsphase in eine Art Hochstimmung versetzt, in der sie zur Selbstüberschätzung neigt und trotz Alkoholisierung glaubt, noch am Straßenverkehr teilnehmen zu können.

Die Alltagserfahrung und auch einige wissenschaftliche Studien, die bisher zum Einfluss von Alkohol auf die Stimme (vgl. Kapitel 3.1) durchgeführt wurden, zeigen, dass sowohl sprachliche Fähigkeiten als auch die Stimme vom Alkoholkonsum beeinflusst werden. Außerdem verbreitet sich zunehmend der Einsatz von sprachgesteuerten Geräten, die immer mehr auch in Fahrzeugen vorzufinden sind (z.B. Navigationssysteme). Eine Präventionsmöglichkeit des Fahrens unter Alkoholeinfluss könnte eine eingebaute automatische Erkennung der Alkoholisierung einer Person anhand der Sprache sein, zum Beispiel im Navigationssystem eines Fahrzeugs.

Von besonderem Interesse wäre sowohl eine solche automatische Erkennung der Alkoholisierung einer Person, als auch die Möglichkeiten der Erkennbarkeit in der menschlichen Wahrnehmung für das Gebiet der forensischen Phonetik. Nicht selten sind hier Einschätzungen über eine mögliche Alkoholisierung eines Sprechers², möglicherweise im Zusammenhang mit seiner Schuldfähigkeit, gefragt.

Die Motivation eine solche automatische Erkennung zu entwickeln, stützt sich darauf, dass die meisten Menschen glauben, anhand der Stimme einer Person erkennen zu können, ob diese Person alkoholisiert ist oder nicht. Wie oben bereits erwähnt, tritt eine solche Situation in der Realität allerdings selten auf – meist gibt es weitere sichtbare oder geruchliche Faktoren, die bei der Einschätzung (möglicherweise unbewusst) hilfreich sind. Ziel dieser Arbeit soll sein, zunächst den Einfluss von Alkohol auf die Stimme anhand der Tonhöhe und deren verwendeten Umfang zu prüfen – gut hörbare Merkmale, die auch akustisch messbar sind. Der nächste Schritt wird sein, zu testen, ob Hörer, die eine Person nicht kennen und auch nicht sehen können, anhand von zwei Sprachbeispielen eines Sprechers entscheiden können, in welchem der Beispiele der Sprecher alkoholisiert war. Weiterführende Forschungsfrage wird hier sein, ob die Tonhöhe oder der Tonumfang in diesem Fall eine Rolle in der Wahrnehmung der Alkoholisierung eines Sprechers spielt. Um dies zu testen wurden zwei weitere Perzeptionstests durchgeführt. Ziel des einen Tests ist es, die Beispiele eines alkoholisierten Sprechers in Tonhöhe und -umfang mittels Grundfrequenzmanipulation an die des Beispiels in nüchternem Zustand anzupassen, und so die Alkoholisierung, sollte sie sich auf die Grundfrequenz eines Sprechers auswirken, hörbar zu kompensieren. Erwartet wird eine Verschlechterung der Erkennbarkeit des unter Alkoholeinfluss produzierten Beispiels im Vergleich zu dem ursprünglichen Test. Im anderen Test soll das Gegenteil erreicht werden. Hier werden zwei Beispiele eines Sprechers, zu deren Aufnahmezeitpunkten er in beiden Fällen nüchtern war, miteinander verglichen. Allerdings wird in einem Beispiel, ebenfalls mittels Grundfrequenzmanipulation, ei-

² Aufgrund der besseren Lesbarkeit wird in dieser Arbeit das generische Maskulinum verwendet. Sofern nicht explizit anders gekennzeichnet, beziehen sich alle Personenbezeichnungen auf Männer und Frauen.

ne Alkoholisierung simuliert. Erwartet wird hier, dass die Hörer das Beispiel, in dem die Tonhöhe manipuliert wurde, häufiger für das Beispiel halten, in dem der Sprecher alkoholisiert war.

Die Arbeit ist folgendermaßen aufgebaut: zunächst wird in Kapitel 2 das Alcohol Language Corpus (ALC) vorgestellt, das die Datenbasis für die folgenden Experimente bildet. Kapitel 3.1 liefert einen Literaturüberblick über die bisherigen Forschungsergebnisse zu akustischen Veränderungen in Sprache unter Alkoholeinfluss. Schwerpunkt liegt hier auf der Grundfrequenz. Eine Untersuchung der Grundfrequenz der Sprecher des ALC folgt in Kapitel 3.2. Hier wird auf mögliche sprecherindividuelle Auswirkungen der Alkoholisierung eingegangen und auf Zusammenhänge mit der jeweiligen BAK geprüft. Eine weitere Untersuchung der Mikrovariation der Grundfrequenz und der Lautstärke, ebenfalls anhand von Sprechern des ALC, wird in Kapitel 3.3 vorgestellt. In Kapitel 4.1 folgt erneut ein Literaturüberblick, diesmal über bisherige Studien zur Wahrnehmbarkeit von Alkoholisierung anhand der Sprache eines Sprechers. In Kapitel 4.2 wird der Basistest zur Diskriminationsfähigkeit naiver Hörer mit Sprechern des ALC vorgestellt, in dem auch mögliche Einflüsse der Grundfrequenz auf die Hörerentscheidungen geprüft werden. Nachfolgend werden die beiden Tests mit manipulierter Grundfrequenz in Kapitel 4.3 beschrieben und die Einflüsse der Manipulation anhand eines Vergleichs mit den Ergebnissen des Basistests geprüft. Eine abschließende zusammenfassende Diskussion mit Ausblick auf zukünftige Forschung bildet Kapitel 5.

2

Das Alcohol Language Corpus

Das Alcohol Language Corpus (ALC) entstand in einer Kooperation des *Bavarian Archive of Speech Signals* (BAS) der Ludwig-Maximilians-Universität, dem Institut für Rechtsmedizin und der Firma BAS Services Schiel in München. In erster Linie sollte eine große, öffentlich zugängliche Sprachdatenbank mit Aufnahmen von Sprechern in nüchternem Zustand und unter Alkoholeinfluss entstehen.

Der „Bund gegen Alkohol und Drogen im Straßenverkehr“ (B.A.D.S) und das Institut für Rechtsmedizin der Ludwig-Maximilians-Universität veranstalten regelmäßig freiwillige Trinkversuche für Rechtsreferendare¹. Ziel dieser Trinkversuche ist es, den Teilnehmern ein Gefühl dafür zu vermitteln, wie man sich mit einer bestimmten Blutalkoholkonzentration (BAK) fühlt, hauptsächlich, um eine Fahruntüchtigkeit unter Alkoholeinfluss zu demonstrieren. Im Rahmen dieser Trinkversuche entstanden die Aufnahmen des ALC. In den Versuchen konnten die Teilnehmer selbst wählen, welche BAK sie am Ende erreichen wollten, möglich waren hier Werte von 0,3‰ bis 1,5‰. Die individuelle Trinkmenge an Bier und Wein, um die gewählte BAK zu erreichen, wurde für jeden Teilnehmer mithilfe der Formeln von Watson et al. (1980) und Widmark (1932) berechnet, die im Folgenden kurz erläutert werden.

¹ Die Trinkversuche fanden an unterschiedlichen Orten in Bayern statt, die Aufnahmen der nüchternen Sprecher am Institut für Sprachverarbeitung der Ludwig-Maximilians-Universität in München.

Anhand der *Widmark*-Formel (2.1) lässt sich die BAK anhand der aufgenommenen Menge an Alkohol in Abhängigkeit des Körpergewichts abschätzen:

$$c = \frac{V}{m * r} \quad (2.1)$$

wobei c die BAK darstellt (in ‰), V die Menge an konsumiertem Alkohol (in g), m das Körpergewicht der Person (in kg) und r den sogenannten Reduktionsfaktor, der abhängig von Geschlecht, Alter und Körpergewicht der Person ist. Um diesen zu berechnen, werden die *Watson*-Formeln 2.2 angewandt, die zunächst geschlechtsspezifisch das Gesamtkörperwasser (*GKW*) abschätzen:

$$GKV_{\text{männlich}} = 2,447 - 0,09516 * t + 0,1074 * h + 0,3362 * m \quad (2.2)$$

$$GKV_{\text{weiblich}} = 0,203 - 0,07 * t + 0,1069 * h + 0,2466 * m \quad (2.3)$$

wobei t für das Alter (in Jahren), h für die Körpergröße (in cm) und m auch hier für das Körpergewicht (in kg) steht. Um dann den Reduktionsfaktor auszurechnen, werden noch die Dichte des Blutes ($\rho_{\text{Blut}} = 1,055 \text{ g/cm}^3$) und der Anteil des Wassers im Blut ($f = 0,8$) berücksichtigt:

$$r = \frac{\rho_{\text{Blut}} * GKW}{f * m} \quad (2.4)$$

Löst man die *Widmark*-Formel nun nach V auf, ersetzt r durch Formel 2.4 und setzt die von der Versuchsperson gewünschte zu erreichende BAK (c) ein, erhält man die Menge an Alkohol (in g), die für das Erreichen der gewünschten BAK konsumiert werden muss (Schiel et al. 2012).

$$V = \frac{c * \rho_{\text{Blut}} * GKW}{f} \quad (2.5)$$

Anhand dieser Menge wurde dann berechnet, wieviel Bier oder Wein von dem Probanden innerhalb höchstens zwei Stunden konsumiert werden sollte.

Nach der Trinkphase folgte eine 20-minütige Pause, nach der sowohl ein Atemals auch ein Blutalkoholtest durchgeführt wurden. Die 20-minütige Pause diente dazu, dass sich der Alkohol im Mundraum verflüchtigt, erst danach sind die Messungen der Atemalkoholkonzentration (AAK) aussagekräftig. Unmittelbar nach dem Blutalkoholtest fanden die freiwilligen Sprachaufnahmen statt. Der Umfang der Sprachaufnahmen wurde so gewählt, dass sie in etwa 15 Minuten Zeit in Anspruch nahmen, um eventuelle Veränderungen der gemessenen BAK durch den Abbau von Alkohol über die Dauer der Aufnahme zu vermeiden. Die Aufnahmen der alkoholisierten Sprecher umfassen insgesamt 30 Einzelaufnahmen in drei Sprachstilen:

- gelesene Sprache:
 - 5 Nummern (Telefonnummern oder Kreditkartennummern)
 - 5 Adressen
 - 5 Zungenbrecher
- Spontansprache (Bildbeschreibungen oder Antworten auf Fragen des Versuchsleiters, maximale Länge: 60 Sekunden):
 - 3 Monologe
 - 2 Dialoge mit dem Versuchsleiter
- Kommandosprache (Kontrollkommandos um z.B. das Radio oder das Navigationsgerät im Auto zu steuern):
 - 5 gelesene Kommandos
 - 5 selbst formulierte Kommandos (aufgrund lesbarer Beschreibungen, wie z.B. „Sie möchten zum Hilton Nürnberg und dafür Ihr Navigationssystem benutzen. Geben Sie Ihrem Auto den entsprechenden Befehl.“)

Etwa zwei Wochen später fanden die Aufnahmen derselben Sprecher in nüchternem Zustand statt, mit doppelt so vielen Aufnahmen in jeder Kategorie (insgesamt also

60 Aufnahmen und ca. 30 Minuten Aufnahmezeit). Von 20 Sprechern (10 weiblichen und 10 männlichen) wurde das Material, das in alkoholisiertem Zustand aufgenommen wurde, als Kontrollgruppe ein weiteres Mal nüchtern aufgenommen. Die Sprecher wurden sowohl nüchtern als auch in alkoholisiertem Zustand in einem Auto aufgenommen, um dieselbe akustische Umgebung für beide Aufnahmen sicherzustellen. Der Motor des Aufnahmeautos blieb bei zwei Dritteln aller Aufnahmen ausgeschaltet, lediglich die Kommandosprache wurde mit laufendem Motor aufgenommen, um eine realistische akustische Umgebung für die Kontrollkommandos zu schaffen. Für die Aufnahmen wurden zwei Autos mit jeweils einem AKG Q400 Grenzflächen-Mikrofon mittig am Dachhimmel ausgestattet, zusätzlich bekamen die Sprecher ein Beyerdynamic Opus 54.16/3 *Headset*-Mikrofon. In den Autos führten zwei unterschiedliche Aufnahmeleiter die Aufnahmen durch. Der jeweilige Aufnahmeleiter überwachte die überwiegend selbständig geführte Aufnahme² und diente gleichzeitig als Dialogpartner, wenn dies in der Aufgabe vorgesehen war. Die Aufnahmen eines Sprechers (sowohl im nüchternen als auch im alkoholisierten Zustand) fanden immer im selben Auto und mit dem selben Versuchsleiter statt.

Zusätzlich zu Angaben zu Geschlecht, Alter, Atem- und Blutalkoholkonzentration, Gewicht, Größe, Dialektregion³, Beruf und Rauchgewohnheiten wurden auch die Trinkgewohnheiten abgefragt und die Sprecher je nach Häufigkeit und Intensität des Alkoholkonsums in drei Gruppen (*light*, *moderate*, und *heavy drinkers*) eingeteilt. Um die Sprecher in diese Gruppen einzuteilen, wurden sie nach der Häufigkeit ihres Alkoholkonsums pro Woche gefragt und das Schema in Tabelle 2.1 angewandt. *Oft* bedeutet in diesem Fall, dass täglich oder mehrmals pro Woche Alkohol konsumiert wird, *seldom*, dass nur einmal oder weniger als einmal pro Woche Alkohol konsumiert wird. Hinsichtlich der Menge des Alkohols bedeuten ein bis zwei Einheiten *wenig* und mehr als zwei Einheiten *viel*.

² Auf einem Laptop war die aktuelle Aufgabe sichtbar, die Versuchsperson konnte eigenständig zur nächsten Aufgabe schalten.

³ Bundesland, in dem der Sprecher die Grundschule besucht hat

	selten	oft
wenig	<i>light</i>	<i>moderate</i>
viel	<i>moderate</i>	<i>heavy</i>

Tab. 2.1: Einteilung der Sprecher in Gruppen nach Trinkgewohnheiten

Weiterhin wurde der Sprecher nach seinem generellen und momentanen emotionalen Befinden gefragt.

Alle Aufnahmen wurden mit dem webbasierten Programm *WebTranscribe* (Draxler 2005) manuell transkribiert und außerdem in allen Aufnahmen mit Hilfe des automatischen Segmentiersystems *MAUS* (*Munich AUtomatic Segmentation system*) (Schiel 1999) Lautgrenzen für jede Aufnahme bestimmt.

3

Akustische Veränderungen in der Sprache alkoholisierter Sprecher

In diesem Kapitel werden Untersuchungen von akustischen Parametern in der Sprache alkoholisierter Sprecher des in Kapitel 2 beschriebenen ALC präsentiert. Zunächst folgt ein Literaturüberblick in Kapitel 3.1 mit einer Zusammenfassung der Ergebnisse bisheriger Studien in Tabelle 3.1. Danach wird in Kapitel 3.2 eine Langzeitanalyse der mittleren Grundfrequenz und der Variationsbreite der Grundfrequenz vorgestellt. Abschließend wird in Kapitel 3.3 eine Analyse der Mikroperturbationsparameter *Jitter* und *Shimmer* beschrieben.

3.1 Literaturüberblick

Bereits 1968 führten Trojan & Kryspin-Exner Sprachaufnahmen mit drei männlichen Sprechern unter Alkoholeinfluss durch, um eventuelle Veränderungen in Sprache und Stimme im Vergleich zu Aufnahmen in nüchternem Zustand derselben Sprecher aufzuzeigen. Neben einigen sprecherspezifischen sprachlichen Besonderheiten beobachteten die Autoren auch einen Anstieg der mittleren Grundfrequenz.

Das Hauptaugenmerk der Studie von Sobell, Sobell & Coleman (1982) liegt in der Untersuchung von Unregelmäßigkeiten in der Sprache bei alkoholisierten Spre-

chern. Zusätzlich analysierten sie aber auch akustische Parameter wie die Grundfrequenz (F_0), Amplitude (Lautheit) und die Sprechgeschwindigkeit. Sie untersuchten 16 männliche Sprecher, die ein kurzes Textstück in nüchternem, mäßig alkoholisiertem (0,16‰ – 0,51‰) und stark alkoholisiertem (0,45‰ – 1,17‰) Zustand vorlasen. Bezüglich der F_0 konnte allerdings kein signifikanter Effekt festgestellt werden.

Eine größere Forschungsgruppe um David Pisoni veröffentlichte eine Reihe von Studien zum Thema Alkohol und Sprache. Basis der Studien ist eine Datenbank von zunächst vier männlichen Sprechern in alkoholisiertem und nüchternem Zustand, die später unter gleichen Aufnahmebedingungen erweitert wurde. Die Datenbank umfasst Sprache von einzelnen Wörtern bis hin zu kurzen Sätzen und etwas längeren Textstücken, zum Teil gelesen und zum Teil nachgesprochen. Aufnahmen in alkoholisiertem Zustand wurden erst vorgenommen, wenn der Sprecher eine AAK von mindestens 1,0‰ aufwies.

In Pisoni, Hathaway & Yuchtman (1985) dienten 34 kurze Sätze von vier Sprechern mit einer AAK von 1,0‰ – 1,7‰ in intoxikiertem Zustand als Datenmaterial. Gegenstand des akustischen Teils der Studie war unter anderem eine Analyse der Grundfrequenz, die bei allen vier Sprechern eine deutliche größere Variabilität (gemessen an der Standardabweichung) in den Aufnahmen in alkoholisiertem Zustand im Gegensatz zu denen in nüchternem Zustand aufweist. Bezüglich der mittleren F_0 lassen sich anhand dieser vier Sprecher allerdings keine großen Unterschiede zwischen den beiden Zuständen erkennen.

Behne & Rivera (1990) konzentrierten sich in ihrer Studie auf die Analyse von – im Deutschen nicht existent – zweisilbigen Wörtern mit Akzenten auf beiden Silben. Sie verwendeten hierfür Aufnahmen von sechs männlichen Sprechern, die höchste erreichte AAK betrug auch hier 1,7‰. Neben einigen anderen akustischen Parametern untersuchten Behne & Rivera auch die Grundfrequenz. Es konnten signifikant höhere Werte in alkoholisiertem Zustand gemessen werden, ebenso zeigt sich im Mittel mehr Variation der F_0 , wenn auch die Ergebnisse sprecherabhängig sind.

Eine weitere Analyse von Material aus derselben Datenbank wird in Cummings, Chin & Pisoni (1995) beschrieben. Vokale aus acht einzeln gesprochenen Wörtern von vier Sprechern wurden für die Analyse einiger akustischer Parameter verwendet. Schwerpunkt der Studie liegt auf der Analyse der glottalen Anregung. Der am stärksten alkoholisierte Sprecher weist eine AAK von 1,7‰ auf. Auch in diesen Daten variiert die Grundfrequenz in den Aufnahmen der Sprecher in alkoholisiertem Zustand stärker als in nüchternem Zustand und auch die mittlere F_0 ist höher anzusetzen. Zusätzlich führten Cummings et al. auch Jitter- und Shimmernessungen durch. Viele der Ergebnisse für einzelne Jitter- und Shimmerparameter zeigten zwar keine Konsistenz innerhalb dieser vier Sprecher aber signifikante Unterschiede zwischen der Aufnahme eines Sprechers in nüchternem und der in alkoholisiertem Zustand.

Chin, Large & Pisoni (1997) arbeiteten ebenfalls mit der gleichen (erweiterten) Datenbank wie Pisoni et al. (1985). Basis für diese Studie sind Aufnahmen von kurzen gelesenen Sätzen, artikuliert von neun männlichen Sprechern zwischen 21 und 26 Jahren. Die höchste gemessene AAK betrug bei diesen Sprechern 1,9‰. Neben einigen anderen akustischen Parametern untersuchten Chin et al. auch die mittlere F_0 , Minimum und Maximum der F_0 und die Standardabweichung. In der mittleren F_0 ist nur ein leichter Anstieg in alkoholisiertem Zustand zu verzeichnen, alle anderen F_0 -Parameter unterscheiden sich signifikant zwischen beiden Aufnahmen. Das F_0 -Minimum liegt im Mittel in alkoholisiertem Zustand tiefer als in nüchternem Zustand, das F_0 -Maximum dagegen höher – der Umfang der verwendeten Grundfrequenz ist also im Durchschnitt größer in alkoholisiertem Zustand. Gleichzeitig steigt auch die Variabilität der F_0 , hier gemessen an der Standardabweichung, in alkoholisiertem Zustand an.

Klingholz, Penning & Liebhardt (1988) untersuchten in ihrer Studie 16 männliche, deutsche Sprecher zwischen 25 und 35 Jahren, wobei 11 Sprecher in alkoholisiertem und nüchternem Zustand aufgenommen wurden und 5 Sprecher, die – zweimal nüchtern aufgenommen – als Kontrollgruppe fungierten. Der gemessene Blutalkoholwert der alkoholisierten Gruppe variierte zwischen 0,67‰ und 1,59‰. Drei dieser alkoholisierten Sprecher wurden nach einer weiteren Trinkphase mit einer höheren BAK

erneut aufgenommen. Gegenstand der Aufnahme war ein bekanntes deutsches Märchen von 520 Wörtern Länge. Klingholz et al. untersuchten die Grundfrequenz, indem sie die Häufigkeitsverteilung der Frequenzen für kurze Signalstücke analysierten. Die Verteilungen der Sprecher in nüchternem Zustand, verglichen mit denen der Sprecher in alkoholisiertem Zustand, zeigt, dass die am Häufigsten genutzte Frequenz bei alkoholisierten Sprechern höher liegt, und außerdem die Verteilung breiter ist. Die Autoren schließen daraus, dass die Grundfrequenz bei Sprechern in alkoholisiertem Zustand sowohl im Mittel höher liegt als auch stärker variiert.

Eine interdisziplinäre Studie von Künzel, Braun & Eysholdt (1992) beschreibt sehr ausführlich phonetisch-linguistische, phoniatisch-logopädische und neurologische Einflüsse von Alkohol auf Sprache und Stimme. Am Trinkversuch, bei dem die Aufnahmen für die Analysen entstanden, nahmen 33 männliche Polizeibeamte teil. Die Aufnahmen beinhalten sowohl gelesene Sprache, als auch Semispontansprache. Die Autoren teilen ihre Sprecher nach ihrer maximalen gemessenen Atemalkoholkonzentration in Gruppen von $< 0,4\text{‰}$ bis $> 2,0\text{‰}$, mit einer Breite von $0,4\text{‰}$ ein. Neben vielen weiteren Untersuchungen analysierten Künzel et al. auch mögliche Veränderungen in der Grundfrequenz von alkoholisierten Sprechern. Bei einem Großteil der hier untersuchten Sprecher zeigt sich eine höhere Grundfrequenz in alkoholisiertem Zustand, bei einigen wenigen allerdings auch eine niedrigere. Die Unterschiede der über die beiden Gruppen gebildeten Mittelwerte ist jedoch nicht signifikant, es sei denn, man unterteilt die Gruppe der alkoholisierten Sprecher erneut in stark und schwach alkoholisiert. Hier zeigt sich ein signifikanter Unterschied zwischen der F_0 stark alkoholisierten Sprecher ($> 0,8\text{‰}$) und der F_0 gemessen in den jeweiligen nüchternen Aufnahmen. Die Erhöhung der Grundfrequenz in alkoholisiertem Zustand fällt aber verhältnismäßig um einiges stärker aus als eine Verringerung. Außerdem wird deutlich, dass die Erhöhung der F_0 auch mit der Stärke der Alkoholintoxikation steigt, ebenso die Variabilität des Stimmtons – hier gemessen anhand der Standardabweichung der F_0 . Die Variabilität des Stimmtons ist insgesamt in intoxikiertem Zustand signifikant höher als in nüchternem Zustand, bei der Einteilung in zwei Gruppen ergibt sich aber

nur für die Gruppe der stark alkoholisierten Sprecher ein signifikant höherer Wert. Jitterwerte berechneten die Autoren mit Hilfe des RAP (*Relative Average Perturbation*), also einer Berechnung der durchschnittlichen Schwankung der Mittelwerte aufeinanderfolgender Perioden. Auf verschiedene Jittermaße und deren Berechnung wird noch genauer in Kapitel 3.3 eingegangen. Als Grundlage für die Jitterberechnung diente der Mittelwert zweier Langvokale /i:/ aus dem Wort ‚blies‘, das in der gelesenen Sprache zweimal vorkam. Die Jitterwerte der alkoholisierten Sprecher sind deutlich (und signifikant) höher als die Jitterwerte der Aufnahmen in nüchternem Zustand. Signifikante Werte zeigen sich sowohl bei Betrachtung des Gesamtmittelwertes, als auch in der Betrachtung der Mittelwerte der Gruppen der leicht bzw. stark alkoholisierten Sprecher. Es lässt sich allerdings kein direkter Zusammenhang zum Grad der Alkoholisierung feststellen.

Eine Studie von Johnson et al. (1993)¹ beschäftigt sich hauptsächlich mit Elektropalatographie (EPG)- und Elektroglottographie (EGG)-Daten von zwei weiblichen und zwei männlichen Sprechern. Zusätzlich untersuchten sie aber auch andere akustische Parameter wie F_0 , Jitter und Shimmer anhand eines anhaltend gesprochenen Vokals /a:/. Die Probanden wurden alle nüchtern und in vier verschiedenen Alkoholisierungsgraden aufgenommen, drei in der Phase steigender Intoxikation und eine in der Abbauphase nach der maximalen gemessenen BAK. Insgesamt reichten die BAK von 0,35‰ bis 1,3‰. Zwar konnten Johnson et al. einen Anstieg in der Variabilität der F_0 bestätigen, nicht aber einen Anstieg der mittleren F_0 mit zunehmender Alkoholisierung. Jitter- und Shimmermessungen führen zu relativ ähnlichen Ergebnissen: mit steigender Alkoholisierung steigen im Allgemeinen auch Jitter- und Shimmerwerte. Allerdings steigen diese Werte bis zur letzten Aufnahme, in der die BAK bereits wieder sinkt. Die Autoren hypothetisieren, dass, sollte der Anstieg der Jitter- und Shimmerwerte auf eventuelle morphologische Veränderungen der Stimmlippen zurückzuführen sein, solche Veränderungen einen verspäteten Effekt nach sich ziehen könnten.

¹ zitiert nach Chin & Pisoni (1997)

An einer Studie von Watanabe et al. (1994) nahmen 48 Freiwillige japanische Muttersprachler teil (11 Frauen und 37 Männer), die sowohl nüchtern, als auch unter Alkoholeinfluss die Vokale /a/, /e/, /i/, /o/ und /u/ produzierten. Die jeweiligen BAK oder AAK der Sprecher während der Produktion der Vokale lassen sich der Studie leider nicht entnehmen. Unter anderem untersuchten Watanabe et al. auch die Grundfrequenz in diesen Vokalen und konnten sowohl bei den Frauen als auch bei den Männern eine niedrigere mittlere F_0 in alkoholisiertem Zustand feststellen. Außerdem ist auch hier die Variationsbreite der F_0 signifikant höher als in nüchternem Zustand.

In ihrer Masterarbeit untersuchte Cooney (1998) Sprachaufnahmen von 12 Sprechern (8 weiblichen und 4 männlichen) und kommt aufgrund von unterschiedlichem Sprachmaterial zu unterschiedlichen Ergebnissen. Zum einen analysierte sie den anhaltend gesprochenen Vokal /a:/ und einige isoliert gesprochene Wörter und fand außer einer nicht signifikanten Erhöhung der Variabilität mit steigendem Alkoholkonsum der Sprecher keine Zusammenhänge zwischen Alkoholisierung und der mittleren F_0 . In einer Analyse einiger Vokale, herausgeschnitten aus einigen vorgelesenen Sätzen, lässt sich aber ein signifikant steigender Zusammenhang zwischen Alkoholisierung und mittlerer F_0 erkennen. Leider sind die Messwerte der AAK laut Aussage der Autorin unbrauchbar und die Stärke der Alkoholisierung wird hier anhand von ml(konsumierter Alkohol)/kg(Körpergewicht) ausgedrückt und spiegelt somit leider den Alkoholisierungsgrad nicht so exakt wider wie eine Atem- oder Blutalkoholanalyse.

Ein weiteres größeres Projekt um Harry Hollien steuerte mehrere Studien zu dem Thema bei, die alle auf einer (später erweiterten) Datenbank basieren. Hauptziel dieses Projekts war es, eine größere Anzahl an bisher analysierten Sprechern aufzunehmen und gleichzeitig die Alkoholisierung zu kontrollieren. Zu Anfang (Hollien & Martin 1996) bestand die Datenbank aus 16 Sprechern (5 Frauen und 11 Männer zwischen 21 und 32 Jahren), wobei 3 davon als Kontrollgruppe fungieren und Placebos bekommen sollten. Dies stellte sich im weiteren Verlauf des Projekts allerdings als schwierig heraus, da die Probanden trotz Alkoholgeruch im Glas mit einem unalkoholischen Getränk

und weiteren Maßnahmen nicht glaubten, dass sie laut Versuchsleiter stark alkoholisiert sein sollten. Die Idee der Kontrollgruppe wurde also vorerst fallengelassen. Da Menschen mit unterschiedlichen Konstitutionen und Trinkgewohnheiten mutmaßlich auch unterschiedlich auf eine bestimmte Alkoholmenge bzw. eine bestimmte AAK reagieren, sollten die Versuchspersonen zu mehreren Zeitpunkten in verschiedenen Trunkenheitsgraden Sprachaufnahmen machen. Die festgelegten Fenster für die AAK lagen hier bei 0,4‰ – 0,5‰ (leichte Alkoholisierung), 0,8‰ – 0,9‰ (Grenze vor dem Gesetz) und 1,2‰ – 1,3‰ (starke Alkoholisierung). Zusätzlich wurde eine weitere Aufnahme in der Abbauphase des Alkohols gemacht, also wenn die AAK nach Erreichen des höchsten Wertes wieder sinkt und wieder bei 0,9‰ – 0,8‰ angekommen ist. In jedem Trunkenheitszustand wurde das gleiche Sprachmaterial aufgenommen, das aus gelesenen Sätzen, Textstücken, Antworten auf Fragen und diadochokinetischen Übungen² bestand.

Aufgrund der Einschätzungen von Hörern in einem Perzeptionstest, der in Hollien & Martin (1996) beschrieben wird (auf die Ergebnisse dieses Tests wird noch gesondert in Kapitel 4.1 eingegangen), untersuchten die Autoren bei einigen der Sprecher aus der Datenbank auch die Grundfrequenz. Hierfür unterteilten die Autoren sechs Sprecher in zwei Gruppen. Eine Gruppe bestand aus Sprechern, die unter Alkoholeinfluss von den Hörern auch tatsächlich als stärker alkoholisiert wahrgenommen wurden, die andere aus Sprechern, bei denen die Hörer kaum oder gar keinen Unterschied zwischen den beiden Zuständen in ihrer Beurteilung machten. In dem Perzeptionstest wurde von jedem Sprecher nur eine nüchterne und eine stark alkoholisierte Aufnahme beurteilt. Auch in der Grundfrequenzanalyse betrachteten Hollien & Martin für jeden der analysierten Sprecher nur die Aufnahmen in nüchternem und die in stark alkoholisiertem (AAK 1,2‰ – 1,3‰) Zustand und keinen Verlauf über die verschiedenen Alkoholisierungsgrade. Anhand der wenigen Sprecher lässt sich aber keine allgemeine Aussage über den Einfluss von Alkohol auf die Grundfrequenz treffen – einige Sprecher spre-

² schnelle Silbenwiederholungen von echten oder Nonsenswörtern, meist mit wechselnden Artikulationsstellen – wird oft in der Sprachtherapie eingesetzt

chen alkoholisiert mit höherer F_0 , andere mit niedrigerer und bei wieder anderen bleibt sie gleich. Auch zwischen den beiden Gruppen ist kein signifikanter Unterschied festzustellen. Die Autoren merken an, dass hier eine größere Population vonnöten sei.

In Aldermann, Hollien, Martin & DeJong (1995) wird ebenfalls eine Einteilung in hörbar betrunkene Sprecher und nicht hörbar betrunkene Sprecher vorgenommen, und die F_0 der nüchternen Aufnahmen mit den Aufnahmen der stark alkoholisierten Sprecher verglichen. Wieviele Sprecher Teil der Analyse waren, wird hier leider nicht deutlich. In der Gruppe der hörbar alkoholisierten Sprecher ist eine niedrigere F_0 in alkoholisiertem Zustand messbar, allerdings ist der Unterschied zur Aufnahme in nüchternem Zustand nicht signifikant. In der Gruppe der nicht hörbar alkoholisierten Sprecher ist kein klarer Trend ablesbar. Eine weitere Gruppe von Sprechern wurde nach ihren Trinkgewohnheiten in *light*, *medium* und *heavy* unterteilt, also leichte, moderate und starke Trinker, doch auch hier war bezüglich der mittleren F_0 in keiner der Gruppen ein signifikanter Unterschied feststellbar.

Eine Studie (unter anderem) zur Grundfrequenz mit einer größeren Population von 19 männlichen Sprechern ist in Hollien et al. (1999) beschrieben. Hier wurde sowohl Spontansprache, als auch gelesene Sprache in allen vier Alkoholisierungsgraden in der Phase steigender Alkoholisierung untersucht. In beiden Sprachstilen zeigt sich ein signifikant steigender Trend, ausgehend von der nüchternen Aufnahme bis hin zu der stark alkoholisierten Aufnahme. Die Variabilität der F_0 steigt zwar auch mit zunehmender Alkoholisierung an, allerdings nicht so stark wie es von den Autoren aufgrund der vorhergehenden Ergebnisse angenommen wurde.

Auch in Hollien et al. (2001a) ist ein signifikanter Unterschied zwischen der F_0 in den Aufnahmen in nüchternem Zustand und denen unter Alkoholeinfluss zu verzeichnen. Hier untersuchten Hollien et al. die Sprachaufnahmen von 16 Frauen und 19 Männern in allen drei Alkoholisierungsgraden in der steigenden Phase – auch hier unterteilt in gelesene und Spontansprache. In der gelesenen Sprache ist der Unterschied in der mittleren F_0 bei den Männern zwischen den beiden Aufnahmen doppelt so stark ausgeprägt wie bei den Frauen. Betrachtet man den Trend über die verschiedenen Al-

koholisierungsgrade, zeigt sich bei den Männern ein signifikanter Unterschied zwischen allen Stufen, außer zwischen der leicht und der mittelstark alkoholisierten Stufe (ca. 0,4‰ bzw. 0,8‰), bei den Frauen hingegen besteht nur zwischen diesen beiden Aufnahmen ein signifikanter Unterschied. Die Ergebnisse der Spontansprache fallen beinahe identisch aus, lediglich das Ausmaß der F_0 -Veränderung in alkoholisiertem Zustand ist nicht so hoch. Eine Aufteilung in die drei Trinkgewohnheitsgruppen zeigt in allen drei Gruppen sowohl für Frauen als auch für Männer signifikante Unterschiede zwischen der Aufnahme in nüchternem und der in stark alkoholisiertem Zustand. Es fällt allerdings auf, dass dieser Unterschied bei beiden Geschlechtern in der Gruppe der starken Trinker wesentlich geringer ausfällt als bei den Gruppen mit weniger ausgeprägtem Alkoholkonsum. Bezüglich der Variation der Grundfrequenz (auch hier gemessen an der Standardabweichung) lässt sich zwar generell ein leichter Anstieg in der Aufnahme in stark alkoholisiertem Zustand beobachten, allerdings ist dieser Unterschied nicht signifikant. Als weiteren interessanten Aspekt stellen Hollien et al. heraus, dass trotz der signifikanten Ergebnisse ca. ein Fünftel der Sprecher einen gegenläufigen Trend zeigen und in alkoholisiertem Zustand mit niedrigerer Grundfrequenz sprechen.

Um der Fragestellung nachzugehen, inwiefern ein Sprecher alkoholisiert noch dazu in der Lage ist, Einfluss auf die Wahrnehmbarkeit seiner Intoxikation zu nehmen – also in welchem Maß ein alkoholisierter Sprecher fähig ist, seine Trunkenheit zu verdecken und umgekehrt, ob Alkoholisierung in der Sprache auch vorgetäuscht von naiven Hörern als echte Alkoholisierung wahrgenommen wird, führten Hollien, DeJong & Martin (1998) Experimente mit Schauspielern durch. Hierfür baten sie 12 Schauspieler (5 Frauen und 7 Männer zwischen 21 und 37 Jahren) im nüchternen Zustand drei verschiedene Grade an Alkoholisierung zu imitieren, die eine AAK von 0,4‰ – 0,5‰, 0,8‰ – 0,9‰ und 1,2‰ – 1,3‰ widerspiegeln sollen. Außerdem wurden von den Schauspielern Sprachaufnahmen mit eben diesen AAK gemacht, bei denen sie gebeten wurden ganz normal zu sprechen, und zusätzlich eine Aufnahme, bei der sie mit der höchsten AAK nüchterne Sprache simulieren sollten. Auf die durchgeführten Perzeptionstests und deren Ergebnisse wird in Kapitel 4.1 noch eingegangen. Bei der

Untersuchung der akustischen Eigenschaften der Aufnahmen (Hollien et al. 2001b) wird zunächst das Ergebnis aus Hollien et al. (2001a) bestätigt – auch die Aufnahmen der stark alkoholisierten Schauspieler (ohne Verdeckungsintention) weisen einen ähnlichen Effekt bezüglich der Grundfrequenz auf. Auch hier ist die mittlere F_0 in der nüchternen Aufnahme niedriger. In den Aufnahmen mit simulierter Trunkenheit zeigt sich bei den Männern ebenfalls eine höhere mittlere F_0 , bei den Frauen allerdings eine niedrigere. Für eine Analyse der Aufnahmen, in denen Nüchternheit vorgetäuscht wurde, stützen die Autoren sich auf Ergebnisse aus Perzeptionstests aus Hollien et al. (1998) (vgl. Kapitel 4.1) und unterteilen die Sprecher nach ihrem Erfolg, der sehr sprecherabhängig war. Solche Sprecher, die ihre Alkoholisierung gut maskieren konnten, sprachen hier auch mit niedrigerer F_0 – im Mittel sogar niedriger als in tatsächlich nüchternem Zustand. Bei denjenigen Sprechern, die keinen Erfolg in der Vortäuschung der Nüchternheit hatten, lässt sich auch kein gemeinsamer Trend bezüglich der F_0 erkennen.

In einer relativ aktuellen Studie von Tisljár-Szabó et al. (2013) konnte wiederum kein signifikanter Unterschied in der Grundfrequenz bei alkoholisierten Sprechern im Vergleich zu Aufnahmen derselben Sprecher in nüchternem Zustand festgestellt werden, aber auch hier wird ein leichter Anstieg der F_0 in der unter Alkoholeinfluss produzierten Aufnahme beobachtet. Die Basis der Analysen bildet eine hauptsächlich aus Zungenbrechern bestehende Datenbank mit Aufnahmen von 15 ungarischen Muttersprachlern (7 weibliche und 8 männliche) in nüchternem und alkoholisiertem Zustand. Leider wird nicht deutlich, wie hoch die individuellen AAK oder BAK der einzelnen Sprecher zum Zeitpunkt der Aufnahmen waren, sondern lediglich beschrieben, dass jeder Proband dreimal 50 ml eines 40%igen alkoholischen Getränks konsumierte und dazu angehalten war, vier Stunden vor Beginn des Versuchs nichts zu essen. Aus den Daten von Tisljár-Szabó et al. (2013) geht allerdings hervor, dass der (wenn auch nicht signifikante) Unterschied zwischen den beiden Aufnahmen für Männer ausgeprägter ist als für Frauen, und auch sie stellen fest, dass die Ergebnisse sehr sprecherspezifisch sind und sich einige Sprecher genau gegenläufig verhalten.

Die Ergebnisse aus den bisherigen Studien, vor allem den Einfluss von Alkohol auf die mittlere F_0 betreffend, führen also bisher noch zu keiner Einigkeit darüber, ob die Unterschiede zwischen Aufnahmen eines Sprechers in alkoholisiertem Zustand und Aufnahmen in nüchternem Zustand von regelmäßiger Natur sind, oder ob es sich um Einzelphänomene handelt. Zwar lässt sich in den bisherigen Studien zum Teil ein signifikanter Anstieg in der mittleren F_0 in alkoholisiertem Zustand feststellen (z.B. Klingholz et al. 1988; Behne & Rivera 1990; Hollien et al. 2001a) in einigen Studien wird aber auch eine niedrigere F_0 bei alkoholisierten Sprechern gemessen (z.B. Watanabe et al. 1994). Dass so unterschiedliche Ergebnisse aufgezeigt werden, liegt vermutlich nicht nur an individuellen Reaktionen der Sprecher auf Alkohol, sondern auch an den vielen unterschiedlich aufgebauten Studien mit zum Teil sehr wenigen (meist männlichen) Probanden. Auch problematisch ist die schlechte Vergleichbarkeit aufgrund sehr unterschiedlich stark alkoholisierter Sprecher und die eventuelle Ungenauigkeit der häufig gemessenen AAK im Gegensatz zur tatsächlichen BAK, die nur in den seltensten Fällen erhoben wurde. Trotz vieler unterschiedlicher Ansätze lässt sich dennoch zusammenfassend sagen, dass bezüglich der mittleren F_0 in den meisten Studien ein Anstieg in alkoholisiertem Zustand gemessen wurde, es besteht jedoch größere Einigkeit über einen Anstieg der Variabilität der F_0 . Jitter- und Shimmermessungen wurden bisher sehr selten durchgeführt. Dabei wurde aber immer von einem deutlichen Anstieg der Parameter berichtet (Cummings et al. 1995; Künzel et al. 1992; Johnson et al. 1993).

Tab. 3.1: Überblick über die bisherigen Forschungsergebnisse zu akustischen Veränderungen in Sprache unter Alkoholeinfluss

Studie	Sprecher	Alkoholisierung in %	mittlere F ₀	F ₀ -Variabilität	Jitter/Shimmer
Trojan & Kryspin-Exner (1968)	3m	•	↑	•	•
Sobell et al. (1982)	16m	AAK 0,16 – 1,17	-	•	•
Klingholz et al. (1988)	11m	BAK 0,67 – 1,59	↑	↑	•
Pisoni et al. (1985)	4m	AAK 1,0 – 1,7	-	↑	•
Behne & Rivera (1990)	6m	AAK 1,0 – 1,7	↑	•	•
Cummings et al. (1995)	4m	AAK 1,0 – 1,7	↑	↑	↑/↓
Chin et al. (1997)	9m	AAK 1,0 – 1,9	↑	↑	•
Künzel et al. (1992)	33m	AAK < 0,4 – > 2,0	↓↑	↑	RAP: ↑
Johnson et al. (1993) ³	2w/2m	BAK 0,35 – 1,3	-	↑	↑
Watanabe et al. (1994)	11w/37m	BAK •	↓	↑	•

w = weiblich, m = männlich, ↑/↓ = signifikant höher/niedriger in alkoholisiertem Zustand, ↑/↓ = Trend erkennbar, nicht signifikant, - = kein Ergebnis, • = nicht untersucht

³ zitiert nach Chin & Pisoni (1997)

Tab. 3.1: Überblick über die bisherigen Forschungsergebnisse zu akustischen Veränderungen in Sprache unter Alkoholeinfluss: Fortsetzung

Studie	Sprecher	Alkoholisierung in %	mittlere F ₀	F ₀ -Variabilität	Jitter/Shimmer
Cooney (1998)	8w/4m	•	-	↑	•
Hollien & Martin (1996)	2w/4m	AAK 1,2 – 1,3	↓↑	•	•
Hollien et al. (1999)	19m	AAK 0,4 – 1,3	↑↑	↑	•
Aldermann et al. (1995)	•	AAK 1,2 – 1,3	↓	•	•
Hollien et al. (2001a)	16w/19m	AAK 0,4 – 1,3	↑↑	↑	•
Hollien et al. (2001b)	5w/7m	AAK 1,2 – 1,3	↑↑	•	•
Tiszlár-Szabó et al. (2013)	7w/8m	•	↑	•	•

w = weiblich, m = männlich, ↑/↓ = signifikant höher/niedriger in alkoholisiertem Zustand, ↑/↓ = Trend erkennbar, nicht signifikant, - = kein Ergebnis, • = nicht untersucht

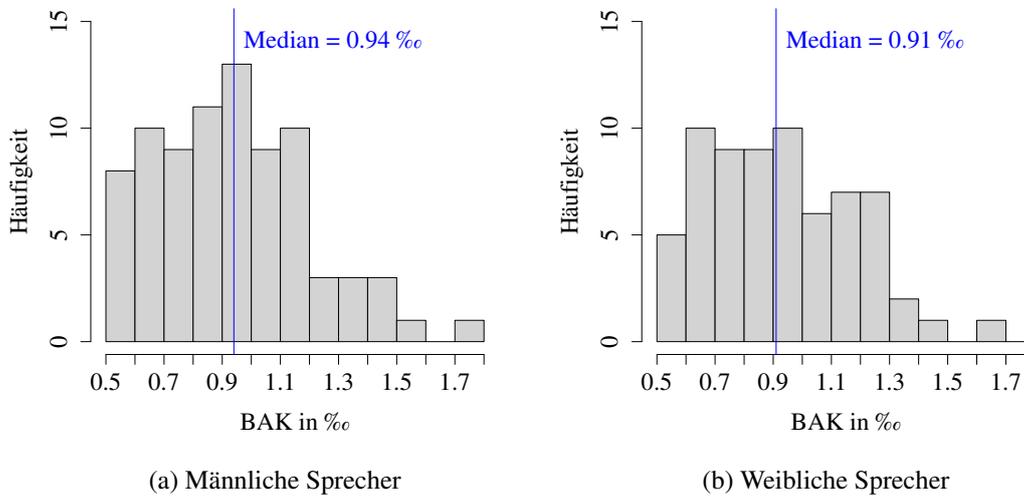


Abb. 3.1: Verteilung der Blutalkoholkonzentrationen

3.2 Langzeitanalyse der Grundfrequenz

Im folgenden Abschnitt wird eine Langzeitanalyse des F_0 -Medians und des F_0 -Interquartilsabstands (F_0 -IQA) vorgestellt. Zunächst folgt in Kapitel 3.2.1 eine Beschreibung der Extraktion der Parameter, in Kapitel 3.2.2 werden Hypothesen aufgestellt und in Kapitel 3.2.3 die Ergebnisse präsentiert. Hier wird auf sprecherindividuelle Unterschiede eingegangen und auf mögliche Zusammenhänge zwischen dem Ausmaß der Veränderung der F_0 -Parameter zwischen den Aufnahmen eines Sprechers in alkoholisiertem und nüchternem Zustand und der BAK eines Sprechers getestet. Abschließend werden die Ergebnisse in Kapitel 3.2.4 diskutiert.

3.2.1 Methode

Für die Analyse der Grundfrequenz wurden aus dem ALC nur das Sprachmaterial derjenigen Sprecher verwendet, deren BAK über 0,5‰ liegt, um eine deutlichere Unterscheidung zwischen nüchtern und alkoholisiert artikulierter Sprache zu erreichen. Da 0,5‰ in Deutschland die legale Grenze für Autofahrer darstellt, erscheint diese Unterteilung angemessen. In der folgenden Analyse bilden also 148 Sprecher (67 weibliche und 81 männliche) die Grundlage. Eine Verteilung der BAK der einzelnen Sprecher ist

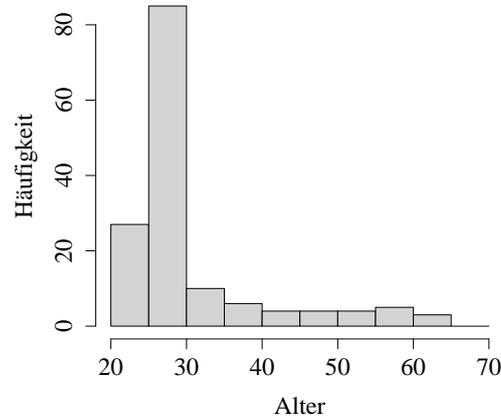


Abb. 3.2: Verteilung des Alters der Sprecher

in den Abb. 3.1a und 3.1b zu sehen. Da das Alter des Großteils der Sprecher zwischen 20 – 30 Jahren liegt, werden in der folgenden Analyse keine Alterseffekte untersucht. Die Altersverteilung der Sprecher zeigt Abb. 3.2.

Zunächst wurde für jedes Sprachsignal mit dem Schäfer-Vincent-Algorithmus (Schäfer-Vincent 1983) alle 5 ms ein F_0 -Wert berechnet. Bei weiblichen Sprechern wurde nur der Frequenzbereich von 100 – 500 Hz betrachtet, bei männlichen Sprechern der Bereich von 50 – 250 Hz. Alle vom Algorithmus als stimmlos klassifizierten Bereiche wurden in der folgenden Analyse ignoriert. Die Werte der stimmhaften Bereiche aus den Dateien eines Sprechers wurden dann zusammengefasst und nach Sprachstil und Alkoholisierung getrennt. Für jede dieser so entstandenen Teilmengen wurde als Variabilitätsmaß der IQA der F_0 -Verteilung sowie der Median als Mittelwert berechnet. Diese beiden Maße lassen sich einerseits mit den in bisherigen Studien verwendeten Maßen vergleichen und sind gleichzeitig unempfindlich gegenüber Störgeräuschen und Ausreißern, was in einem Fall wie diesem wichtig ist, da die Sprachaufnahmen keine Studioqualität haben. Wetterbedingte Störgeräusche oder der zum Teil laufende Motor des Aufnahmefahrzeugs machen die Aufnahmen für die Anwendungsmöglichkeit in einem Navigationssystem o.Ä. realistisch, aber auch anfälliger für Fehler in der automatischen F_0 -Erkennung. Die Mittelung über alle Äußerungen eines Sprechers

pro Sprachstil und Alkoholisierungszustand hat den Vorteil, dass für die Analyse von jedem Sprecher gleich viele Werte zur Verfügung stehen, unabhängig davon, wieviel Sprachmaterial insgesamt von einem Sprecher vorhanden ist – dies variiert natürlicherweise in den spontansprachlichen Äußerungen stark.

3.2.2 Hypothesen

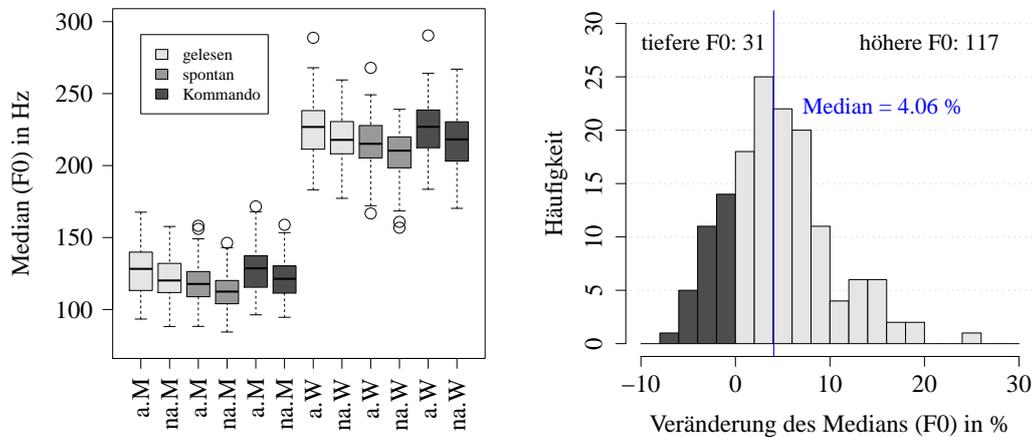
Anhand der gesichteten und in 3.1 vorgestellten Literatur zur Grundfrequenz bei alkoholisierten Sprechern lassen sich folgende Hypothesen aufstellen:

1. In alkoholisiertem Zustand liegt die mittlere Grundfrequenz der Sprecher höher als in nüchternem Zustand.
2. Der Umfang der genutzten Grundfrequenz ist in alkoholisiertem Zustand größer als in nüchternem Zustand.
3. Die Erweiterung des Tonumfangs in alkoholisiertem Zustand zeigt einen deutlicheren Effekt als die Erhöhung der mittleren Grundfrequenz.
4. Je höher die Blutalkoholkonzentration eines Sprechers, desto ausgeprägter sind die Unterschiede aus den Hypothesen 1 und 2.
5. Eine Erhöhung der mittleren Grundfrequenz geht auch mit einer Erweiterung des Tonumfangs einher.

3.2.3 Ergebnisse

3.2.3.1 F_0 -Median

Abb. 3.3a zeigt die Verteilung der F_0 -Mediane, getrennt nach Sprachstil und Alkoholisierung. In allen drei Sprachstilen ist deutlich zu erkennen, dass der Median dieser Verteilungen für alkoholisierte Sprecher höher liegt als für nüchterne Sprecher. Ein Unterschied ist außerdem zwischen den verschiedenen Sprachstilen zu erkennen –



(a) Verteilung der F_0 -Mediane – a = alkoholisiert, na = nicht alkoholisiert, W = weibliche Sprecher, M = männliche Sprecher
 (b) Sprecherindividuelle Veränderungen des F_0 -Medians über alle Sprachstile

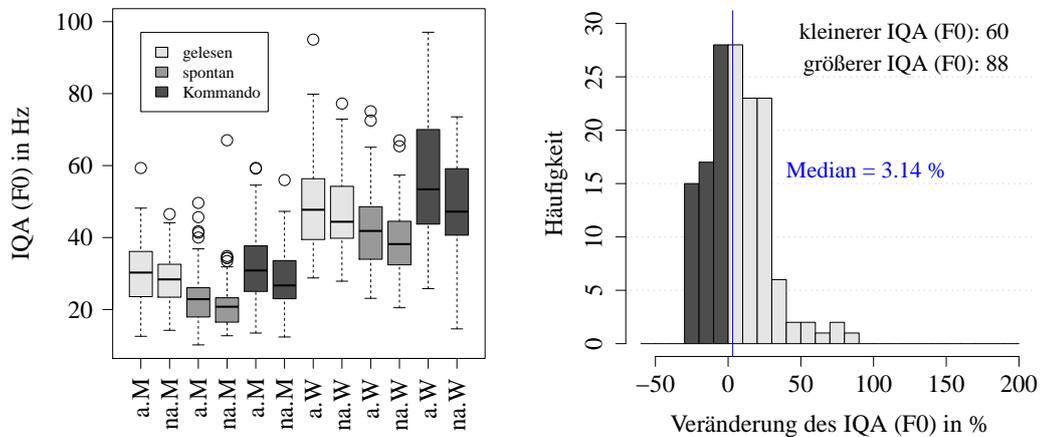
Abb. 3.3: F_0 -Median

Kommandosprache weist die höchsten F_0 -Werte auf, gefolgt von gelesener und Spontansprache. Der relative Unterschied zwischen den Medianen der Sprecher in alkoholisiertem und in nüchternem Zustand ist sowohl bei männlichen als auch bei weiblichen Sprechern ebenfalls in der Kommandosprache am deutlichsten, gefolgt von gelesener und Spontansprache. Insgesamt ist der Unterschied bei den Männern etwas stärker ausgeprägt. Eine Varianzanalyse mit Messwiederholungen und Sprecher als *random factor* zeigt einen höchst signifikanten Unterschied zwischen alkoholisiertem und nüchternem Zustand der Sprecher ($F(1, 146) = 99, 2, p < 0, 001$). Signifikante Effekte des Sprechergeschlechts ($F(1, 146) = 1298, 3, p < 0, 001$) und des Sprachstils ($F(1, 145) = 180, 5, p < 0, 001$) zeigen allerdings keine Interaktion mit der Alkoholisierung, sondern spiegeln lediglich zum einen natürliche Grundfrequenzunterschiede zwischen Männern und Frauen wider und zum anderen Unterschiede in der Tonlage in den unterschiedlichen Sprachstilen. Insgesamt lässt sich also feststellen, dass alkoholisierte deutsche Sprecher des ALC im Mittel mit einer höheren mittleren Grundfrequenz sprechen, als wenn sie nüchtern sind. Dabei spielt weder das Sprechergeschlecht, noch der Sprachstil eine Rolle. Um allerdings sprecherindividuelle Unterschiede festzustellen

len und die jeweilige Frequenzlage der einzelnen Sprecher zu berücksichtigen, wurde pro Sprecher ein relativer Unterschied der F_0 zwischen den Aufnahmen in alkoholisiertem und nüchternem Zustand berechnet. Da in der oben beschriebenen Varianzanalyse keine geschlechts- oder sprachstilspezifischen Unterschiede feststellbar waren, wurde hier nicht mehr nach Geschlecht und Sprachstil getrennt. In der Betrachtung der prozentualen Unterschiede spielen aufgrund der Normierung weder der natürliche Unterschied in der Tonhöhe zwischen Männern und Frauen, noch Unterschiede in der Tonlage zwischen verschiedenen Sprachstilen eine Rolle. Eine Verteilung dieser sprecherindividuellen F_0 -Unterschiede zeigt Abb. 3.3b. 117 von 148 Sprechern – also ca. 79,1% sprechen in alkoholisiertem Zustand mit höherer F_0 . Bei den übrigen 20,9% der Sprecher konnte allerdings eine tiefere F_0 in alkoholisiertem Zustand gemessen werden. Im Mittel erhöhen die Sprecher ihre F_0 in alkoholisiertem Zustand um ca. 4,1%. Auffällig ist aber, dass die Spannbreite der Sprecher, die alkoholisiert mit höherer F_0 sprechen (in Abb. 3.3b durch helle Balken verdeutlicht), wesentlich größer ist, als die der Sprecher, die die F_0 in alkoholisiertem Zustand absenken.

3.2.3.2 F_0 -Interquartilsabstand

In Abb. 3.4a ist die Verteilung der IQA der F_0 , getrennt nach Geschlecht, Sprachstil und Alkoholisierung zu sehen. Auch hier fällt ein deutlicher Unterschied zwischen alkoholisiertem und nüchternem Zustand auf. Bei der Kommandosprache sind die größten IQA zu beobachten, gefolgt von gelesener und Spontansprache. Der relative Unterschied ist sowohl bei männlichen als auch bei weiblichen Sprechern zwischen alkoholisiertem und nüchternem Zustand in der Kommandosprache am größten und in der gelesenen Sprache am kleinsten. Zusätzlich lässt sich eine insgesamt größere Streuung der IQA bei männlichen Sprechern erkennen. Auch hier zeigt eine Varianzanalyse mit Messwiederholungen und Sprecher als *random factor* ein höchst signifikantes Ergebnis für den Faktor Alkoholisierung ($F(1, 146) = 35,21, p < 0,001$). Auch das Sprechergeschlecht ($F(1, 146) = 205,4, p < 0,001$) und der Sprachstil ($F(1, 145) = 119,68, p < 0,001$) zeigen hier einen signifikanten Effekt und im

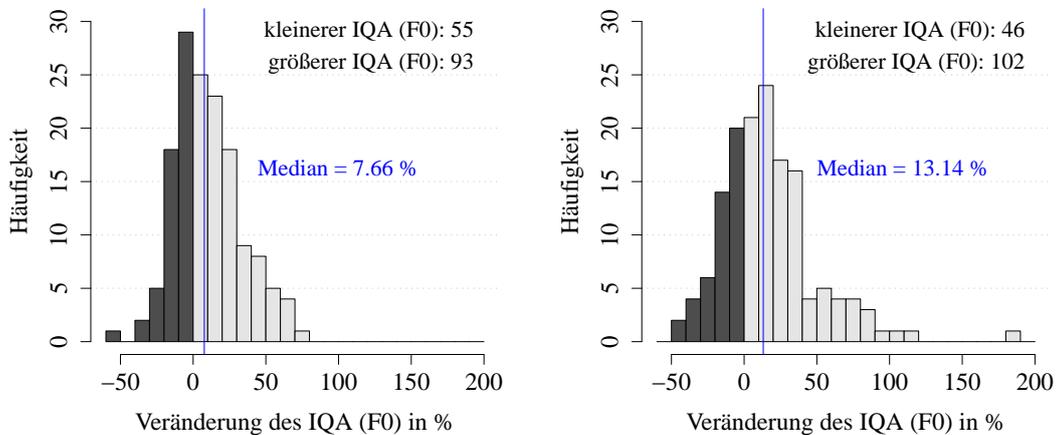


(a) Verteilung der F_0 -Interquartilsabstände – a = *alkoholisiert*, na = *nicht alkoholisiert*, W = *weibliche Sprecher*, M = *männliche Sprecher* (b) Sprecherindividuelle Veränderungen des F_0 -Interquartilsabstands in **gelesener Sprache**

Abb. 3.4: F_0 -Interquartilsabstand

Gegensatz zu der Analyse des F_0 -Median tritt hier auch eine signifikante Interaktion für Sprachstil und Alkoholisierung auf ($F(1, 145) = 7,42, p < 0,001$) und eine schwach signifikante Interaktion zwischen Sprechergeschlecht und Alkoholisierung ($F(1, 146) = 5,03, p < 0,01$). Ein *post-hoc*-Test mit Bonferroni-Korrektur zeigt allerdings, dass der Effekt der Alkoholisierung für beide Geschlechter nicht unterschiedlich stark ausgeprägt ist ($p < 0,001$). Bezüglich des Sprachstils lässt sich allerdings anhand des *post-hoc*-Tests ein stärkerer signifikanter Unterschied zwischen der Aufnahme in alkoholisiertem und der in nüchternem Zustand bei der Kommandosprache und der Spontansprache ($p < 0,001$) im Gegensatz zur gelesenen Sprache ($p < 0,05$) erkennen. Zusammenfassend lässt sich also sagen, dass die analysierten Sprecher im Mittel in alkoholisiertem Zustand eine größere Spanne der F_0 nutzen als in nüchternem Zustand. Dies ist jedoch vom Sprachstil abhängig ist und in der Kommandosprache am stärksten zu beobachten.

Auch hier zeigt die große Streuung der Daten allerdings, dass die Ergebnisse stark sprecherabhängig sind. Um das Verhalten der einzelnen Sprecher genauer zu analysieren, wurde auch hier für jeden Sprecher ein individueller relativer Unterschied in der



(a) Sprecherindividuelle Veränderungen des F_0 -Interquartilsabstands in *Spontansprache* (b) Sprecherindividuelle Veränderungen des F_0 -Interquartilsabstands in *Kommandosprache*

Abb. 3.5: F_0 -Interquartilsabstand

Variationsbreite seiner Grundfrequenz zwischen der Aufnahme in alkoholisiertem und der in nüchternem Zustand berechnet. Da die Varianzanalyse bezüglich des verwendeten Grundfrequenzumfangs keine Unterschiede im Verhalten von männlichen und weiblichen Sprechern unter Alkoholeinfluss zeigte, wird in der folgenden sprecherindividuellen Analyse keine Gruppierung nach Sprechergeschlecht vorgenommen. Da die unterschiedlichen Sprachstile allerdings unterschiedlich stark ausgeprägte Unterschiede zwischen den Aufnahmen in alkoholisiertem und nüchternem Zustand zeigten, wird hier im Gegensatz zur Analyse des sprecherindividuellen F_0 -Medians nach Sprachstilen getrennt. Abb. 3.4b zeigt die Verteilung für gelesene Sprache, Abb. 3.5a die für spontansprachliche Äußerungen und Abb. 3.5b die Verteilung in der Kommandosprache. Insgesamt ist der Anteil an Sprechern, deren F_0 -Umfang in alkoholisiertem Zustand größer ist, nicht so hoch wie der Anteil der Sprecher, die insgesamt ihre mittlere F_0 in alkoholisiertem Zustand erhöhen. In gelesener Sprache sind das nur 59,5% der Sprecher, in der Spontansprache 62,8% und in der Kommandosprache 68,9%. Im Mittel erhöhen die Sprecher ihren genutzten Stimmumfang in gelesener Sprache um 3,14%, in der Spontansprache um 7,66% und in der Kommandosprache sogar um 13,14%.

Sprachstil	Weibliche Sprecher	Männliche Sprecher
gelesen	$F(1, 75) = 2, 52, \quad p > 0, 1 \text{ n.s.}$ $r = 0, 18$	$F(1, 83) = 0, 05, \quad p > 0, 1 \text{ n.s.}$ $r = 0, 03$
spontan	$F(1, 75) = 3, 35, \quad p < 0, 1 .$ $r = 0, 21$	$F(1, 83) = 0, 34, \quad p > 0, 1 \text{ n.s.}$ $r = 0, 06$
Kommando	$F(1, 75) = 0, 21, \quad p > 0, 1 \text{ n.s.}$ $r = 0, 05$	$F(1, 83) = 0, 19, \quad p > 0, 1 \text{ n.s.}$ $r = -0, 05$

Tab. 3.2: Ergebnisse der linearen Regressionen der Veränderung des F_0 -Medians und der BAK getrennt nach Sprechergeschlecht und Sprachstil – *n.s.* = *nicht signifikant*, *.* = *signifikanter Trend*, *** = *signifikant*, **** = *hoch signifikant*, ***** = *höchst signifikant*

3.2.3.3 Zusammenhang mit der Blutalkoholkonzentration

Die oben beschriebenen und beobachteten sprecherindividuellen Unterschiede legen die Vermutung nahe, dass die Unterschiede aufgrund der unterschiedlichen BAK der Sprecher auftreten. Um dies zu testen, wurden lineare Regressionen für die BAK in Abhängigkeit verschiedener Parameter berechnet. Hierfür wurden die in Kapitel 3.2.3.1 und 3.2.3.2 beschriebenen relativen Unterschiede der Parameter F_0 -Median und F_0 -IQA zwischen den alkoholisierten und nüchternen Aufnahmen eines Sprechers verwendet. Für die lineare Regression wurden alle 162 Sprecher des ALC, also auch diejenigen Sprecher mit einer BAK von 0,5‰ oder weniger miteinbezogen. Aufgeteilt wurden die Daten nach Sprechergeschlecht und Sprachstil. Abb. 3.6 zeigt die Streudiagramme der Veränderung des F_0 -Median und der BAK jedes Sprechers, getrennt nach Sprachstilen für weibliche (oben) und männliche Sprecher (unten). Eingezeichnet sind die jeweiligen Regressionsgeraden und der berechnete Regressionskoeffizient r . Die Ergebnisse der Regressionsmodelle sind in Tabelle 3.2 zusammengefasst. Es lässt sich beobachten, dass die für weibliche Sprecher berechneten Regressionskoeffizienten generell etwas höher liegen, als die für männliche Sprecher berechneten – allerdings liegt auch hier der höchste Korrelationskoeffizient bei $r = 0, 21$ in der Spontansprache und lässt somit darauf schließen, dass der Zusammenhang zwischen der F_0 -Median-Veränderung und der jeweiligen BAK eines Sprechers nicht besonders stark ist.

Sprachstile	Weibliche Sprecher	Männliche Sprecher
gelesen	$F(1, 75) = 0, 26, \quad p > 0, 1 \text{ n.s.}$ $r = 0, 06$	$F(1, 83) = 0, 56, \quad p > 0, 1 \text{ n.s.}$ $r = 0, 08$
spontan	$F(1, 75) = 0, 03, \quad p > 0, 1 \text{ n.s.}$ $r = -0, 02$	$F(1, 83) = 0, 0, \quad p > 0, 1 \text{ n.s.}$ $r = -0, 01$
Kommando	$F(1, 75) = 1, 37, \quad p > 0, 1 \text{ n.s.}$ $r = 0, 13$	$F(1, 83) = 0, 35, \quad p > 0, 1 \text{ n.s.}$ $r = 0, 06$

Tab. 3.3: Ergebnisse der linearen Regressionen der Veränderung des F_0 -IQA und der BAK getrennt nach Sprecher-geschlecht und Sprachstil – *n.s.* = *nicht signifikant*, *.* = *signifikanter Trend*, *** = *signifikant*, **** = *hoch signifikant*, ***** = *höchst signifikant*

Bezüglich des Zusammenhangs der F_0 -IQA-Veränderung eines Sprechers und seiner BAK (vgl. Abb. 3.7 und Tabelle 3.3) lassen sich ähnliche Aussagen treffen. Auch hier ist kein starker Zusammenhang erkennbar – der höchste Korrelationskoeffizient liegt hier bei $r = 0, 13$ in der Kommandosprache weiblicher Sprecher. Das Ausmaß der Veränderung des F_0 -IQA zwischen den Aufnahmen in alkoholisiertem und nüchternem Zustand eines Sprechers lässt sich also nicht auf seine BAK zurückführen.

Auf eine mögliche nicht-lineare Korrelation lässt keines der Streudiagramme schließen.

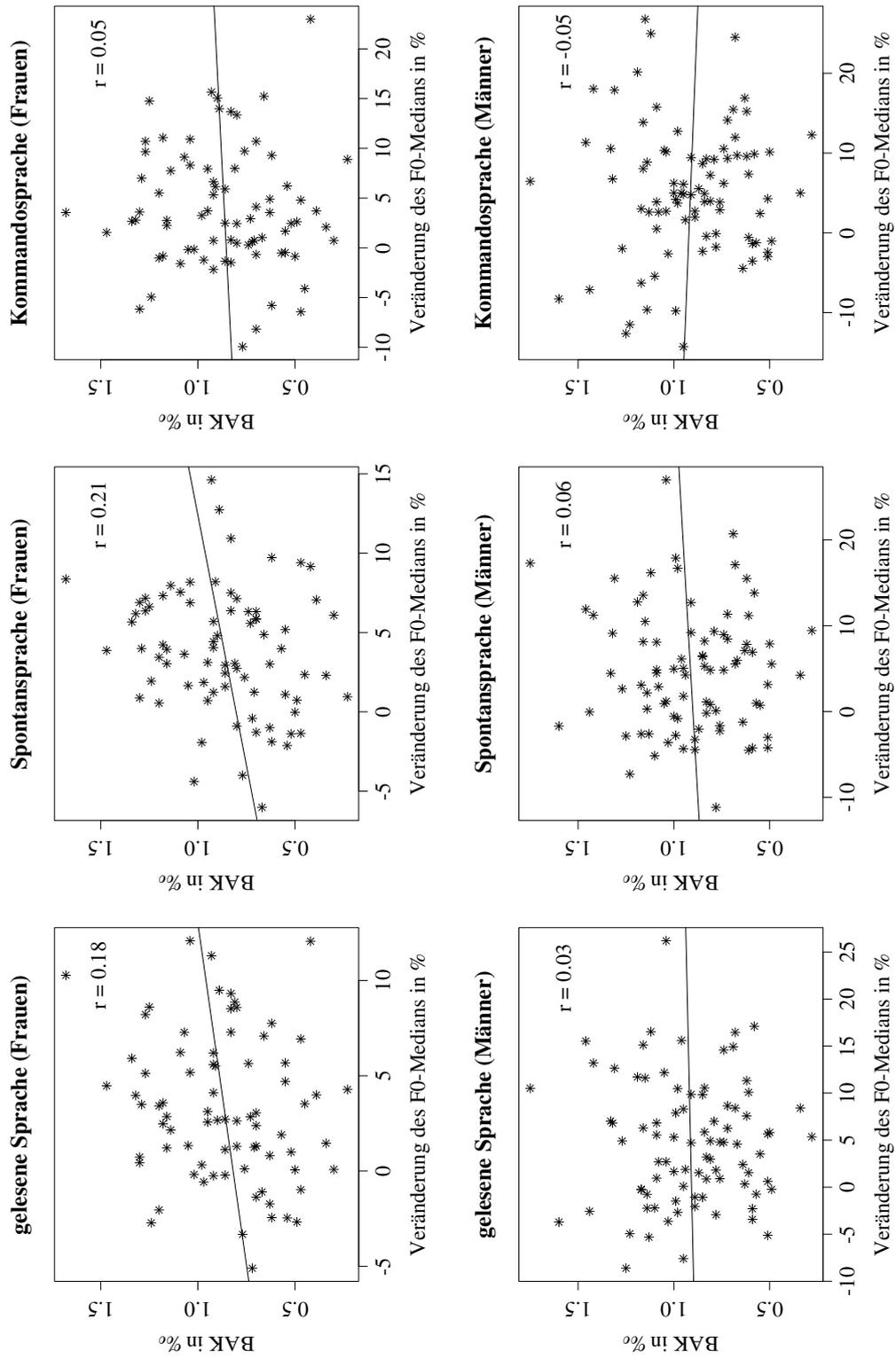


Abb. 3.6: Streudiagramme der Veränderung des F₀-Medians und der BAK

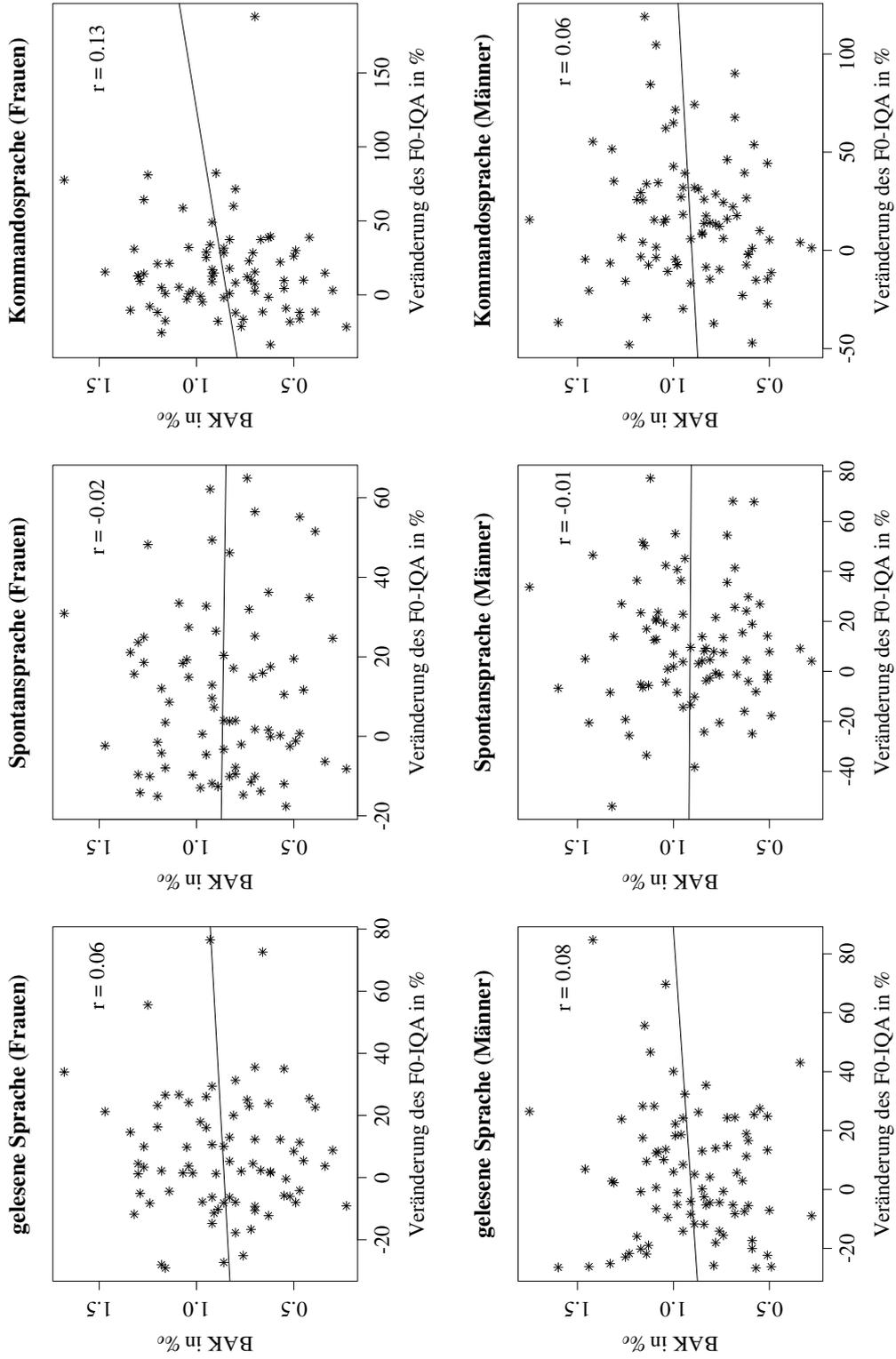


Abb. 3.7: Streudiagramme der Veränderung der F₀-IQA und der BAK

Sprachstile	Weibliche Sprecher	Männliche Sprecher
gelesen	$F(1, 65) = 20,36, p < 0,001$ *** $r = 0,49$	$F(1, 79) = 24,21, p < 0,001$ *** $r = 0,48$
spontan	$F(1, 65) = 17,73, p < 0,001$ *** $r = 0,46$	$F(1, 79) = 42,14, p < 0,001$ *** $r = 0,59$
Kommando	$F(1, 65) = 17,48, p < 0,001$ *** $r = 0,46$	$F(1, 79) = 57,86, p < 0,001$ *** $r = 0,65$

Tab. 3.4: Ergebnisse der linearen Regressionen der Veränderung des F_0 -Medians und des F_0 -IQA getrennt nach Sprechergeschlecht und Sprachstil – *n.s.* = *nicht signifikant*, *.* = *signifikanter Trend*, *** = *signifikant*, **** = *hoch signifikant*, ***** = *höchst signifikant*

3.2.3.4 Zusammenhang zwischen F_0 -Median und F_0 -IQA

Um zu testen, ob eine Erhöhung der mittleren Grundfrequenz in alkoholisiertem Zustand auch mit einer Vergrößerung des Tonumfangs einhergeht, wurden lineare Regressionen für die prozentualen Veränderungen des F_0 -IQA und des F_0 -Medians berechnet. Abb. 3.8 zeigt die Streudiagramme getrennt nach Sprachstil und Sprechergeschlecht, Tabelle 3.4 stellt die Ergebnisse dar. Alle berechneten Modelle zeigen einen signifikanten Zusammenhang zwischen der Veränderung des F_0 -IQA und der Veränderung des F_0 -Medians. Der stärkste Zusammenhang ist mit $r = 0,65$ in der Kommandosprache der männlichen Sprecher zu beobachten. Insgesamt fallen die berechneten Regressionskoeffizienten für männliche Sprecher etwas höher aus als für weibliche Sprecher. Eine Erhöhung des F_0 -Medians hat also in den meisten Fällen auch eine Vergrößerung des Tonumfangs zufolge.

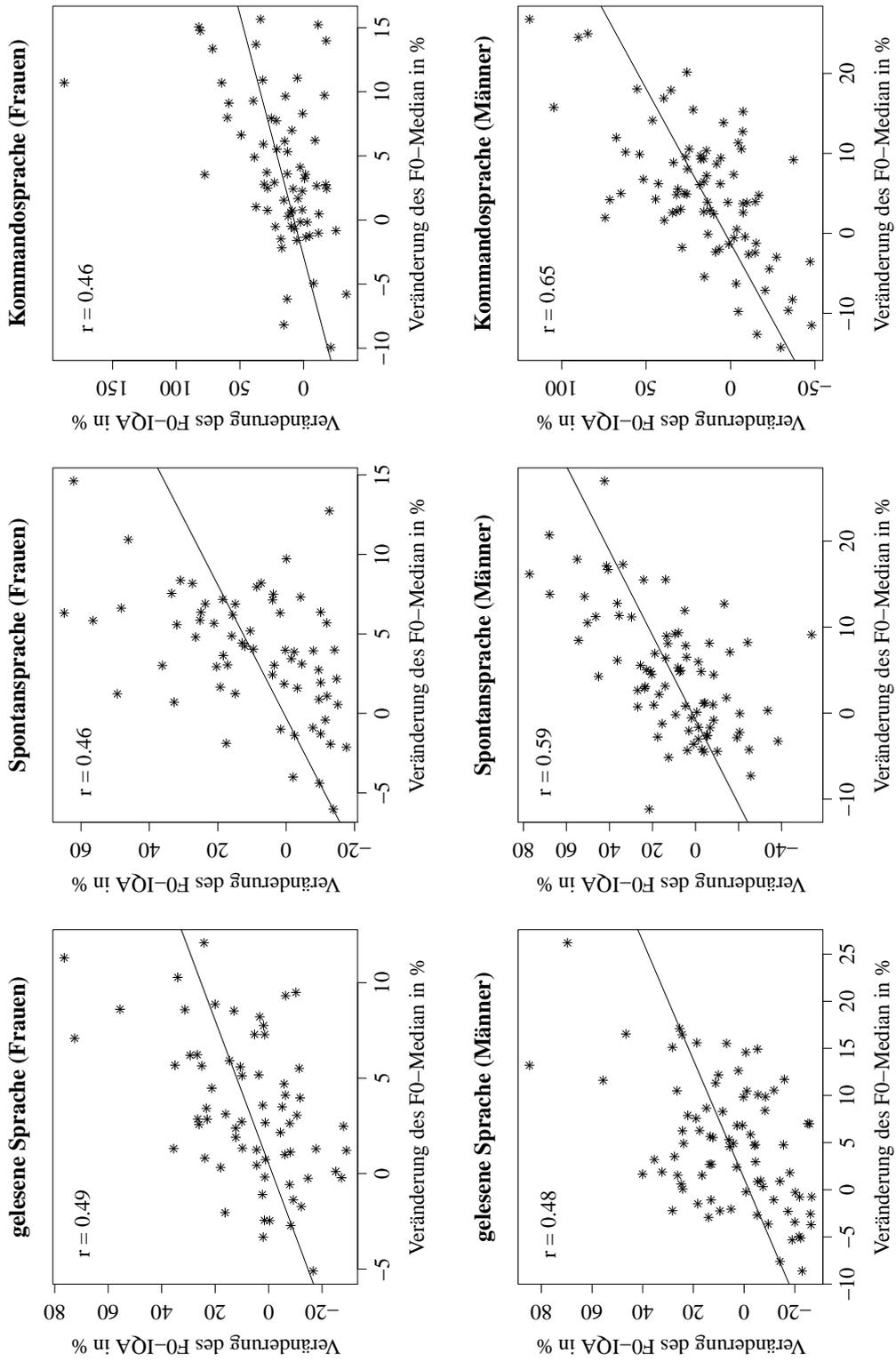


Abb. 3.8: Streudiagramme der Veränderung des F₀-Medians und der Veränderung des F₀-IQA

3.2.4 Zwischendiskussion

Die Langzeitanalyse der Parameter F_0 -Median und F_0 -IQA hat gezeigt, dass Sprecher des ALC mit einer BAK von mehr als 0,5‰ im Mittel sowohl mit einer höheren mittleren Grundfrequenz sprechen als in nüchternem Zustand als auch einen größeren Tonumfang nutzen. Die Hypothesen 1 und 2 lassen sich also bestätigen und entsprechen dem Ergebnis der Mehrheit der bisherigen Studien (vgl. Kapitel 3.1). Dieses Ergebnis ist statistisch höchst signifikant und aufgrund der großen Menge an untersuchten Sprechern repräsentativ. Über einen Anstieg der mittleren Grundfrequenz in alkoholisiertem Zustand herrschte bisher allerdings in der Literatur weniger Einigkeit als über den Anstieg der F_0 -Variabilität. Hier lässt sich jedoch bezüglich der Variabilität ein weniger deutliches Ergebnis als bezüglich der mittleren Grundfrequenz erkennen. Bei einer genaueren Betrachtung der einzelnen Sprecher zeigte sich, dass ca. 79% aller 148 untersuchten Sprecher mit einer BAK von über 0,5‰ alkoholisiert mit höherer Grundfrequenz sprechen als nüchtern, ein höherer F_0 -IQA ist in der sprecherindividuellen Analyse allerdings abhängig vom Sprachstil und fällt prozentual mit 59,5% in der gesprochenen Sprache und 68,9% in der Kommandosprache insgesamt geringer aus. Hypothese 3 kann demnach nicht bestätigt werden.

Hollien et al. (2001a) konnten beispielsweise ein ganz ähnliches Ergebnis aufzeigen. Auch hier lieferte die Analyse der Variationsbreite (hier gemessen an der Standardabweichung) ein weniger deutliches Ergebnis als die Analyse der mittleren Grundfrequenz. Zwar untersuchten sie mit 16 weiblichen und 19 männlichen Sprechern wesentlich weniger Sprecher, jedoch beobachteten sie auch eine signifikant höhere Grundfrequenz in alkoholisiertem Zustand, bei ca. einem Fünftel der Sprecher allerdings einen gegenläufigen Trend. Auch in dieser Arbeit konnte bei 21% der Sprecher eine niedrigere mittlere F_0 in alkoholisiertem Zustand festgestellt werden, was etwa einem Fünftel entspricht. Diese zwei Gruppen lassen sich allerdings, wie in der Verteilung in Abb. 3.3b erkennbar ist, nicht scharf voneinander abgrenzen. Auch zeigt eine lineare Regression keinen Zusammenhang des Ausmaßes der F_0 -Median-Veränderung mit der gemessenen BAK des Sprechers. Künzel et al. (1992) konnten einen solchen

Hypothese	Angenommen?
1. In alkoholisiertem Zustand liegt die mittlere Grundfrequenz der Sprecher höher als in nüchternem Zustand.	✓
2. Der Umfang der genutzten Grundfrequenz ist in alkoholisiertem Zustand größer als in nüchternem Zustand.	✓
3. Die Erweiterung des Tonumfangs in alkoholisiertem Zustand zeigt einen deutlicheren Effekt als die Erhöhung der mittleren Grundfrequenz.	✗
4. Je höher die Blutalkoholkonzentration eines Sprechers, desto ausgeprägter sind die Unterschiede aus den Hypothesen 1 und 2.	✗
5. Eine Erhöhung der mittleren Grundfrequenz geht auch mit einer Erweiterung des Tonumfangs einher.	✓

Tab. 3.5: Übersicht über die Hypothesentests

Zusammenhang beschreiben, allerdings innerhalb verschiedener BAK *eines* Sprechers, wenn diese den Wert von 0,8‰ übersteigen. Hypothese 4 muss also auch verworfen werden.

Hypothese 5 dagegen lässt sich bestätigen – lineare Regressionen der sprecherindividuellen Unterschiede der beiden Parameter F_0 -Median und F_0 -IQA haben gezeigt, dass diese beiden Parameter einen signifikanten Zusammenhang aufweisen. Sprecher, die in alkoholisiertem Zustand eine höhere mittlere Grundfrequenz nutzen als in nüchternem Zustand, erweitern alkoholisiert auch ihren genutzten Tonumfang. Eine übersichtliche Darstellung der Hypothesentests zeigt Tabelle 3.5.

Eine mögliche Erklärung für die Erhöhung der Grundfrequenz in alkoholisiertem Zustand könnte eine gleichzeitige Erhöhung der Lautstärke sein. Diese lässt sich allerdings schwer vergleichbar messen. Die Aufnahmen wurden sowohl mit einem *Headset*-Mikrofon, als auch mit einem im Auto fest installierten Grenzflächen-Mikrofon durchgeführt. Doch trotz sorgfältiger Anbringung des *Headset*-Mikrofons kann hier nicht gewährleistet werden, dass der Abstand von Mikrofon zu Mund bei jedem Sprecher derselbe war. Somit sind bezüglich der Lautstärke weder die Aufnahmen verschie-

dener Sprecher, noch die beiden Aufnahmen eines Sprechers miteinander vergleichbar. Gramming (1991) beobachtete bei gesunden Sprechern eine Anhebung der Lautstärke von 0,3 – 0,5 Halbtönen pro dB. Bei einem Sprecher mit einer mittleren Grundfrequenz von 120 Hz entspräche das bei einer Erhöhung der Lautstärke um beispielsweise 5 dB eine gleichzeitige Erhöhung der mittleren Grundfrequenz um 1,5 – 2,5 Halbtöne, also um ca. 9 – 16% auf ca. 131 – 139 Hz. Der in dieser Arbeit gefundene Effekt könnte also auf eine Erhöhung der Lautstärke in alkoholisiertem Zustand zurückzuführen sein, allerdings erklärt dieser nicht die gleichzeitige Erweiterung des Tonumfangs. Im Rahmen einer Dissertation zur rhythmischen Struktur von Sprache unter Alkoholeinfluss, der als Datenbasis ebenfalls Sprecher des ALC zugrunde lagen, wurden unter anderem Kurzzeitenergiefunktionen (*root mean square* (RMS)) berechnet, die vereinfacht als Abfolge von leisen und lauten Abschnitten betrachtet werden können (Heinrich 2014). Diese RMS-Kurve wurde vor der Extraktion des Parameters normalisiert, indem der durchschnittliche RMS-Wert jeder Kurve von der ursprünglichen Kurve subtrahiert wurde. Ein anhand dieser Kurve berechneter Parameter beschreibt den Interquartilsabstand der relativen Distanz von Minimum oder Maximum zum Mittelwert, beschreibt also das Ausmaß der Änderung der Energiedynamik eines Signals. Ein probeweise getesteter Zusammenhang zwischen diesem Parameter und dem Median der Grundfrequenz eines Sprechers in den Aufnahmen, die unter Alkoholeinfluss produziert wurden, zeigt allerdings keine Relevanz.

Ein weiterer Aspekt, der in Betracht gezogen werden sollte, sind mögliche kulturelle Unterschiede. Im Großteil der in Kapitel 3.1 vorgestellten Studien wurde die Sprache amerikanischer Muttersprachler untersucht. Möglicherweise aufgrund der unterschiedlichen Experimentbedingungen herrscht hier dennoch keine Einigkeit über einen Anstieg oder eine Absenkung der mittleren Grundfrequenz. In der Studie mit japanischen Muttersprachlern (Watanabe et al. 1994) ließ sich als eine der wenigen Fälle eine Tendenz zur Absenkung der mittleren Tonhöhe in alkoholisiertem Zustand nachweisen. Bei ungarischen Muttersprachlern (Tisljár-Szabó et al. 2013) konnte wiederum ein Trend zur höheren mittleren Grundfrequenz gezeigt werden. In den Studien

mit deutschen Muttersprachlern (Klingholz et al. (1988) und Künzel et al. (1992)) konnte in einer Studie ein Anstieg der mittleren Grundfrequenz (Klingholz et al. 1988) verzeichnet werden und in der zweiten signifikante Änderungen in beide Richtungen in Abhängigkeit der BAK eines Sprechers gezeigt werden (Künzel et al. 1992). Die Ergebnisse der in dieser Arbeit vorgestellten Untersuchung sind also mit den bisherigen Ergebnissen für deutsche Muttersprachler vereinbar.

3.3 Jitter und Shimmer

In diesem Kapitel wird eine Analyse der Parameter Jitter (*RAP*) und Shimmer (*APQ3*) der Vokale /a:/ und /i:/ der Sprecher des ALC vorgestellt. Zunächst werden in Kapitel 3.3.1 die beiden untersuchten Parameter erklärt und deren Extraktion aus den Sprachsignalen der Sprecher im ALC beschrieben. Aus der Literatur abgeleitete Hypothesen werden in Kapitel 3.3.2 aufgestellt. Es folgt eine Präsentation der Ergebnisse in Kapitel 3.3.3 und eine abschließende Diskussion in Kapitel 3.3.4.

3.3.1 Methode

Als Jitter bezeichnet man im Allgemeinen die Mikroperturbation der Grundfrequenz, also meist kleine zeitliche Schwankungen der Periodendauer von aufeinanderfolgenden glottalen Schwingungen im stimmhaften Signal. Im Sprachsignal einer menschlichen Stimme sind nie alle Schwingungen gleich lang und genau diese Unregelmäßigkeiten sorgen dafür, dass eine Stimme menschlich und nicht wie die typische „Roboterstimme“ klingt. Jittermessungen kommen normalerweise in der Phoniatrie zum Einsatz und dienen als akustisches Maß zur Einschätzung der Stimmqualität, also der Stabilität des phonatorischen Systems (Baken 1997:S. 166). Schwankungen über das normale Maß hinaus lassen also pathologische Veränderungen des Kehlkopfes und der Stimme vermuten (Titze 1991; Orlikoff 1989).

Normalerweise werden Jitter- und Shimmeranalysen zur Erkennung von Stimmstörungen anhand (meist unter Laborbedingungen aufgenommener) anhaltend gesprochener Vokale in für den Patienten komfortabler Lautstärke und Tonhöhe durchgeführt. Der Grund hierfür liegt darin, dass sich intonatorische Schwankungen in der Grundfrequenz auf den Jitterwert auswirken, und da es in der Stimmdiagnostik von Vorteil ist, mit möglichst vergleichbaren Werten zu arbeiten, sollen diese Einflüsse so gut es geht minimiert werden. Anhaltend gesprochene Vokale liegen im ALC nicht vor, die Messungen müssen also in Vokalen aus zusammenhängender Sprache vorgenommen werden.

Zhang & Jiang (2008) verglichen Jittermessungen in anhaltend gesprochenen Vokalen mit Jittermessungen in aus zusammenhängender Sprache herausgeschnittenen Vokalen. Ziel der Untersuchung war es, herauszufinden, ob Stimmstörungen auch in zusammenhängender Sprache vergleichbar gut erkannt werden. Zhang & Jiang beschreiben eine wesentlich größere Variabilität in den Jittermessungen in zusammenhängender Sprache, die sie zum Beispiel auf den Dialekt eines Sprechers, seinen emotionalen Zustand oder seine Sprechgeschwindigkeit zurückführen.

Ebenfalls einflussnehmend auf Jitter und Shimmer können F_0 , Lautstärke, Sprechergeschlecht und die Auswahl des untersuchten Vokals sein. Einen Literaturüberblick liefert hier Brockmann-Bauser (2012), die allerdings aufgrund der zum Teil sich widersprechenden und schwer vergleichbaren Studien zu dem Schluss kommt, dass sich zu keinem der Einflussfaktoren bisher ein verallgemeinerbarer Effekt herausgestellt hat. Weiterhin hypothetisiert sie, dass die Faktoren interagieren und nur zusammengefasst betrachtet werden können.

Ein weiterer häufig diskutierter Einflussfaktor auf den Jitter ist das Alter eines Sprechers. Sussman & Sapienza (1994) konnten zwar keinen signifikanten Unterschied im durchschnittlichen absoluten Jitter zwischen Kindern und Erwachsenen feststellen, jedoch ist bei älteren, im Vergleich zu jungen Frauen und Männern ein signifikanter Unterschied nachweisbar, vor allem im Vokal /a:/ (Linville 1988). Auch Wilcox & Horii (1980) konnten schon einen höheren Jitter bei älteren Sprechern aufzeigen. An Xue & Deliyski (2001) untersuchten unter anderem verschiedene Jitterparameter bei älteren Sprechern und erhielten wesentlich höhere Messwerte, als das Analyseprogramm MDVP (*Multidimensional Voice Program*, Kay Elemetrics (1993)) als Schwellwerte für eine gesunde (junge) Stimme angibt. Diese Schwellwerte sind aufgrund der vielen unterschiedlichen Aufnahmebedingungen aber mit Vorsicht zu betrachten. Linville (2004) merkt allerdings an, dass die gemessenen Jitterunterschiede in verschiedenen Altersgruppen nicht allein auf das steigende Lebensalter zurückzuführen sind, sondern auch von der generellen körperlichen Konstitution abhängen – vor allem bei Frauen. Nach Linville verschwinden diese Altersunterschiede sogar, wenn auf Gesundheitszu-

stand und Fitness kontrolliert wird. Eine Untersuchung hierzu ist in Ramig & Ringel (1983) nachzulesen.

In *Praat* (Boersma & Weenink 2014) stehen für die Jitter- und Shimmeranalyse verschiedene Methoden zur Berechnung unterschiedlicher Parameter zu Verfügung, zwei davon sind die Berechnungen der *Relative Amplitude Perturbation (RAP)* als Jittermaß und des *three-point Amplitude Perturbation Quotient (APQ3)* als Shimmermaß, die weiter unten kurz beschrieben werden. Im Gegensatz zur einfachen Berechnung des lokalen Jitter, der das Verhältnis der durchschnittlichen absoluten Differenz zwischen zwei aufeinanderfolgenden Perioden zu der durchschnittlichen Periodendauer des analysierten Signalstücks beschreibt, wird bei der Berechnung des Jitter (*RAP*) über drei Perioden geglättet. Der Vorteil an diesem Maß ist, dass der Effekt von langsamen F_0 -Veränderungen auf die Jittermessung minimiert wird, und so nur die schnellen Veränderungen als Perturbation isoliert werden. Da hier die Analyse nicht anhand von anhaltend gesprochenen, sondern aus zusammenhängender Sprache herausgeschnittenen Vokalen durchgeführt wird, empfiehlt sich dieses Maß, um prosodische Einflüsse zu minimieren. Zusätzlich sind die Ergebnisse dadurch vergleichbar mit denen von Künzel et al. (1992), die ebenfalls Vokale aus zusammenhängender Sprache analysierten und ebenfalls den Jitter (*RAP*) als Maß verwendeten.

In *Praat* lassen sich einige Voreinstellungen zur Jitterberechnung vornehmen. Zum einen lässt sich der *period floor* einstellen, die kleinstmögliche Periodendauer – kürzere Perioden als die angegebene werden vom Algorithmus ignoriert. Hier liegt der Standardwert bei 0,1 ms. Der Standardwert der größten möglichen Periodendauer (*period ceiling*) liegt bei 20 ms und lässt sich ebenfalls verändern. Außerdem ist noch die größtmögliche Differenz zwischen zwei aufeinanderfolgenden Periodendauern, hier *maximum period factor* genannt, einstellbar. Für alle aufeinanderfolgenden Perioden wird das Verhältnis zwischen ihren Dauern berechnet, liegt dies über dem eingestellten Wert, geht die Differenz der Periodendauern nicht in die Jitterberechnung mit ein. Der voreingestellte Wert liegt hier bei 1,3. In der Analyse dieser Arbeit wurde hier mit den Standardeinstellungen von *Praat* gearbeitet.

Jitter (RAP)

Der Jitter (*RAP*) – also die relative durchschnittliche Perturbation – bezeichnet das Verhältnis der durchschnittlichen absoluten Differenz zwischen der Dauer einer Periode und dem Durchschnitt der Dauer dieser und der beiden umliegenden Perioden und der durchschnittlichen Periodendauer des gesamten analysierten Signalstücks. Es wird also zunächst die Durchschnittsdauer von drei Perioden ermittelt und die Differenz dieser Dauer zu der Dauer der in der Mitte liegenden Periode (vgl. Abb. 3.9). Dies wird für alle Perioden im Analysebereich, zu denen eine vorhergehende und eine nachfolgende Periode existiert, berechnet und schließlich aus den Ergebnissen dieser Berechnungen wieder ein Durchschnitt ermittelt. Der Jitter (*RAP*)-Wert entspricht dann dem Verhältnis dieses Durchschnitts zu der durchschnittlichen Periodendauer:

$$averagePerturbation = \frac{\sum_{i=2}^{N-1} |T_i - (T_{i-1} + T_i + T_{i+1})/3|}{N - 2} \quad (3.1)$$

$$meanPeriod = \frac{\sum_{i=1}^N T_i}{N} \quad (3.2)$$

$$Jitter(RAP) = \frac{averagePerturbation}{meanPeriod} \quad (3.3)$$

wobei T_i für die i -te Periodendauer und N für die Anzahl der analysierten Perioden steht. Wenn eine der beiden Periodendauern T_i oder T_{i-1} allerdings nicht in dem durch die vorher eingestellten Grenzen zulässigen Bereich liegt, oder das Verhältnis von T_i zu einer der umliegenden Perioden das festgelegte Maximum übersteigt, wird die Berechnung auf der Basis von Periode T_i nicht durchgeführt und N um 1 verringert. Sollte N dabei kleiner als 2 werden, ist die gesamte Rechnung nicht möglich und es wird kein Wert ausgegeben. Für die Berechnung der *meanPeriod* gelten ebenfalls diese voreingestellten Beschränkungen. Werte für Jitter (*RAP*) liegen normalerweise zwischen 0% und 2%.

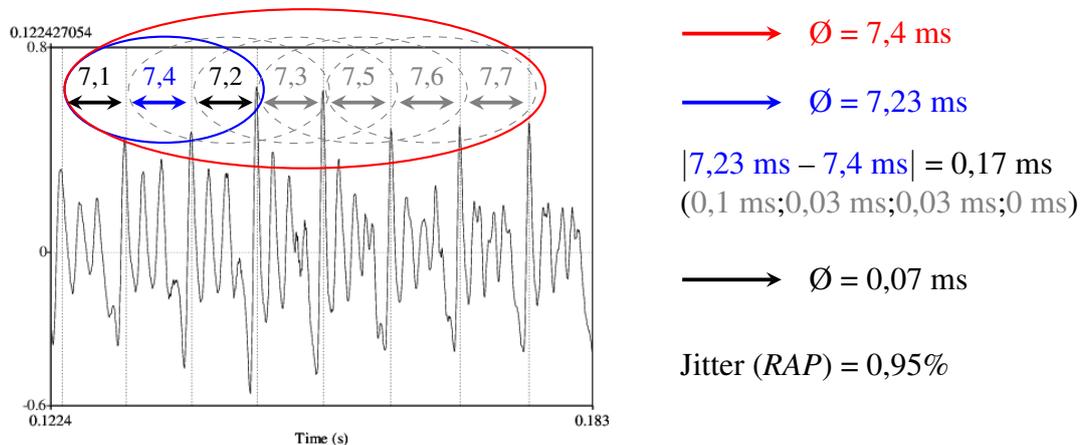


Abb. 3.9: Beispiel zur Berechnung des Jitter (RAP) anhand eines kurzen Signalstücks

Shimmer (APQ3)

Der Shimmer (APQ3) ist ebenso wie der Jitter (RAP) ein relatives Maß und wird ebenfalls auf der Basis von Berechnungen über drei aufeinanderfolgende Perioden berechnet. Wie in Abb. 3.10 zu sehen ist, handelt es sich hier aber nicht wie bei der Berechnung des lokalen Jitters um Dauern, sondern einzelne Punkte, an denen die Amplitude gemessen wird.

$$\text{averagePerturbation} = \frac{\sum_{i=2}^{N-1} |A_i - (A_{i-1} + A_i + A_{i+1})/3|}{N - 2} \quad (3.4)$$

$$\text{meanAmplitude} = \frac{\sum_{i=1}^N A_i}{N} \quad (3.5)$$

$$\text{Shimmer}(APQ3) = \frac{\text{averagePerturbation}}{\text{meanAmplitude}} \quad (3.6)$$

wobei A_i hier für das Amplitudenmaximum der i -ten Periode steht und N für die Gesamtperiodenanzahl.

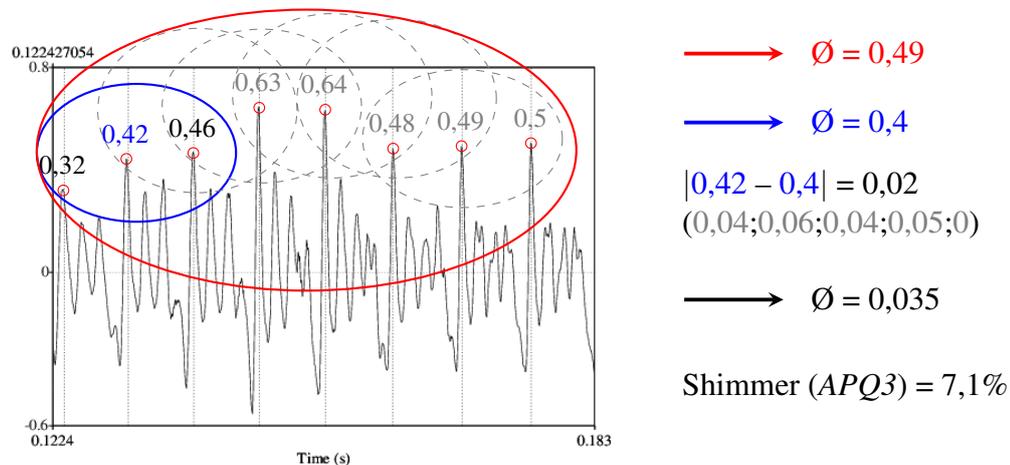


Abb. 3.10: Beispiel zur Berechnung des Shimmer (APQ3) anhand eines kurzen Signalstücks

Zusätzlich zu den Einstellungen bei der Jitterberechnung lassen sich bereits bei der Grundfrequenzberechnung schon einige Voreinstellungen machen. Knarrstimme kommt in der gesprochenen Sprache, je nach Sprecher, relativ häufig vor. Jitterwerte für diese Bereiche fallen sehr hoch aus und sollen auf die folgende Analyse möglichst wenig Einfluss nehmen. Der Grundfrequenzalgorithmus mit den Standardeinstellungen erkennt in der Knarrstimme meist eine sehr tiefe Grundfrequenz. Praat verwendet zur Bestimmung der Grundfrequenz einen Autokorrelationsalgorithmus, der in Boersma (1993) beschrieben wird. Mit dem Verfahren der Autokorrelation wird quasi eine ‚Selbstähnlichkeit‘ zu verschiedenen zeitlichen Verschiebungen des Signals bestimmt. Bei einem periodischen Signal ist dieser Wert natürlicherweise bei einer Verschiebung um ein Vielfaches der Periodendauer sehr hoch und bildet ein Maximum in der Autokorrelationskurve. Je nach Verschiebungsfenster und Ähnlichkeit der einzelnen Perioden eines Signals kann es dabei zu Fehlern kommen. Bei Knarrstimme zum Beispiel ist die Länge der einzelnen aufeinanderfolgenden Perioden sehr unterschiedlich lang und es kann vorkommen, dass z.B. zwei Perioden als eine erkannt werden, und so eine sehr viel tiefere Grundfrequenz angenommen wird. Der Parameter *octave cost* legt fest, ob bei möglichen Kandidaten für die F_0 eher ein hoher Wert favorisiert wird oder nicht. Je größer der Wert dieses Parameters eingestellt ist, desto niedriger ist also die Wahrscheinlichkeit, dass ein tiefer Wert als Grundfrequenz in Frage kommt. Der vor-

eingestellte Wert in *Praat* liegt hier bei 0,01. In der folgenden Analyse wurde der Wert auf 0,2 gesetzt, somit sind sehr niedrige Kandidaten für die Grundfrequenz nicht mehr wahrscheinlich. In dem Fall der Knarrstimme kommen auch keine höheren Kandidaten in Frage – der Teil, in dem Knarrstimme produziert wurde, wird also als stimmlos klassifiziert und geht so nicht in die Jitteranalyse mit ein. Der eingestellte Wert von 0,2 basiert auf Beobachtungen und führt in einer Vielzahl der Beispiele, in denen Knarrstimme vorkommt, zu einer Stimmlosklassifikation des Abschnittes, ohne die restliche F_0 -Kontur maßgeblich zu beeinflussen. Bei allen übrigen Parametern wurden die Standardeinstellungen von *Praat* übernommen.

Die Analyse der Jitter- und Shimmerparameter im ALC erfolgte automatisch mit Hilfe eines *Praat*-Skripts. In allen Äußerungen des ALC wurden mit dem automatischen Segmentiersystem *MAUS* bearbeitet. *MAUS* erstellt eine phonetische Segmentierung mit zeitlichen Information der einzelnen Lautgrenzen. Auf Basis dieser Zeitinformationen wurde in jeder Äußerung (sofern vorhanden) in den ersten Realisierungen der Vokale /a:/ und /i:/ die beiden Parameter Jitter (*RAP*) und Shimmer (*APQ3*) extrahiert. Mit diesen beiden Vokalen sind sowohl ein hoher Vokal (/i:/) als auch ein tiefer Vokal (/a:/) Gegenstand der Analyse. Insgesamt konnten in 6427 von 14580 Dateien Jitter- und Shimmerwerte für den Vokal /a:/ berechnet werden – von jedem Sprecher in durchschnittlich 40 verschiedenen Realisierungen. 2154 in den Dateien der in nüchternem Zustand produzierten Aufnahmen und 4273 in denen unter Alkoholeinfluss produzierten. Die meisten gefundenen /a:-Vokale stammen aus der gelesenen Sprache (ca. 40%), die wenigsten aus der Kommandosprache (ca. 25%), dazwischen liegt die Spontansprache (35%). Für den Vokal /i:/ konnten sogar in 8560 Dateien Jitter- und Shimmerwerte berechnet werden – in durchschnittlich 53 Vokalen für jeden Sprecher. Die Verteilung über die Sprachstile liegt hier bei ca. 50% in der gelesenen Sprache, ca. 27% in der Spontansprache und ca. 23 % in der Kommandosprache.

3.3.2 Hypothesen

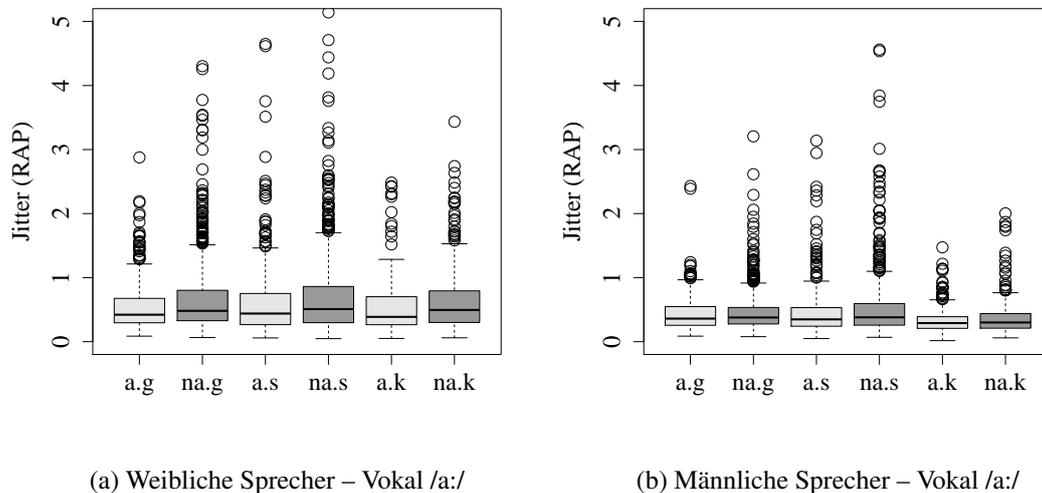
Aufgrund der gesichteten Literatur lassen sich folgende Hypothesen aufstellen:

1. Der Jitterwert eines Sprechers ist höher in den Aufnahmen unter Alkoholeinfluss als in den Aufnahmen ohne Alkoholeinfluss.
2. Der Shimmerwert eines Sprechers ist höher in den Aufnahmen unter Alkoholeinfluss als in den Aufnahmen ohne Alkoholeinfluss.

3.3.3 Ergebnisse

3.3.3.1 Jitter

Die Abb. 3.11a und 3.11b zeigen die Ergebnisse der Jitteranalyse für den Vokal /a:/ – links für weibliche Sprecher und rechts für männliche. Da die Vokale von den Sprechern nicht isoliert gesprochen, sondern aus dem Sprachfluss herausgeschnitten wurden, streuen die Daten sehr stark und es sind sehr viele Ausreißer vorhanden. Zur besseren Übersicht ist in den Abb. 3.11a und 3.11b nur der Bereich von 0 – 5 dargestellt. Insgesamt fällt auf, dass der Jitterwert in den Aufnahmen in nüchternem Zustand für den Vokal /a:/ in allen Sprachstilen und für beide Geschlechter höher liegt als in den Aufnahmen in alkoholisiertem Zustand. Die Daten für weibliche Sprecher streuen zwar etwas stärker als für männliche Sprecher, jedoch ist bei den weiblichen Sprechern auch der Unterschied zwischen den beiden Aufnahmen deutlicher zu erkennen. Im Vokal /i:/ (dargestellt in den Abb. 3.12a und 3.12b) streuen die Jitterwerte für beide Sprecher-geschlechter stärker als im Vokal /a:/, aber auch hier ist in gelesener und Spontansprache ein im Mittel höherer Jitterwert in den Aufnahmen in nüchternem Zustand erkennbar. In der Kommandosprache kehrt sich dieser Effekt sowohl bei weiblichen als auch bei männlichen Sprechern allerdings um – hier wurden höhere Jitterwerte in den Aufnahmen, in denen der Sprecher alkoholisiert war, gemessen. Bei männlichen Sprechern ist dieser Unterschied deutlicher zu erkennen als bei weiblichen Sprechern.

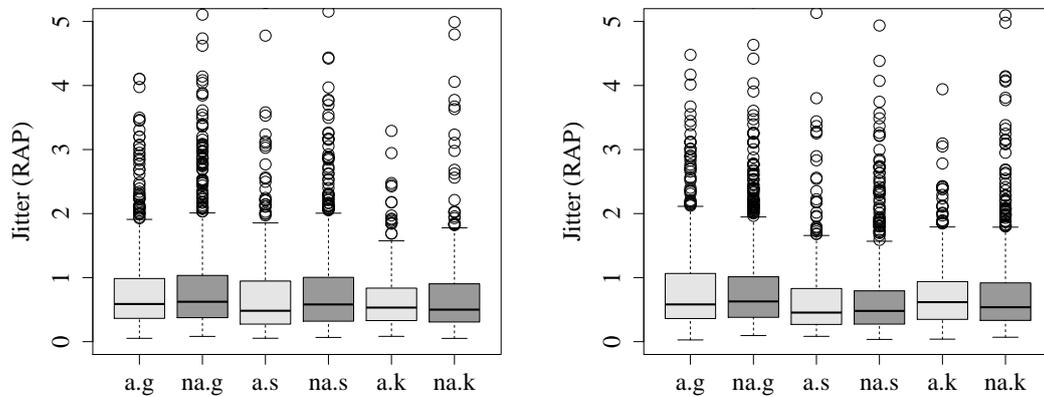


(a) Weibliche Sprecher – Vokal /a:/

(b) Männliche Sprecher – Vokal /a:/

Abb. 3.11: Jitter (RAP) getrennt nach Alkoholisierung und Sprachstil – *a* = *alkoholisiert*, *na* = *nüchtern*, *g* = *gelesene Sprache*, *s* = *Spontansprache*, *k* = *Kommandosprache*

Da ein Gesamtmodell mit dem Jitterwert als abhängige Variable und den *fixed*-Faktoren Alkoholisierung, Sprechergeschlecht, Vokal und Sprachstil und dem Sprecher als *random*-Faktor in einer *Mixed Effect Model (MEM)*-Analyse (Baayen 2008) viele signifikante Interaktionen zwischen den *fixed*-Faktoren aufzeigt, wurden vier, nach Sprechergeschlecht und Vokal getrennte Modelle berechnet. Für den Vokal /a:/ unterscheiden sich die Jitterwerte in den Aufnahmen in alkoholisiertem Zustand signifikant von denen in nüchternem Zustand, der Unterschied ist allerdings für weibliche Sprecher deutlicher als für männliche Sprecher. Im Vokal /i:/ gemessene Jitterwerte unterscheiden sich nicht zwischen nüchtern und alkoholisiert produzierter Sprache. In den unterschiedlichen Sprachstilen ist in allen Faktorkombinationen ein signifikanter Unterschied erkennbar, etwas weniger stark ausgeprägt ist dieser für den Vokal /a:/ bei weiblichen Sprechern – allerdings besagt dieser signifikante Effekt lediglich, dass sich die Jitterwerte zwischen den unterschiedlichen Sprachstilen unterscheiden, was für das Ziel dieser Arbeit nicht von besonderem Interesse ist. Signifikante Interaktionen konnten mit keinem der Modelle gefunden werden, das heißt also, dass die beobachteten Unterschiede zwischen den beiden Sprecherzuständen im Vokal /i:/ weder in die eine



(a) Weibliche Sprecher – Vokal /i:/

(b) Männliche Sprecher – Vokal /i:/

Abb. 3.12: Jitter (RAP) getrennt nach Alkoholisierung und Sprachstil – a = alkoholisiert, na = nüchtern, g = gelesene Sprache, s = Spontansprache, k = Kommandosprache

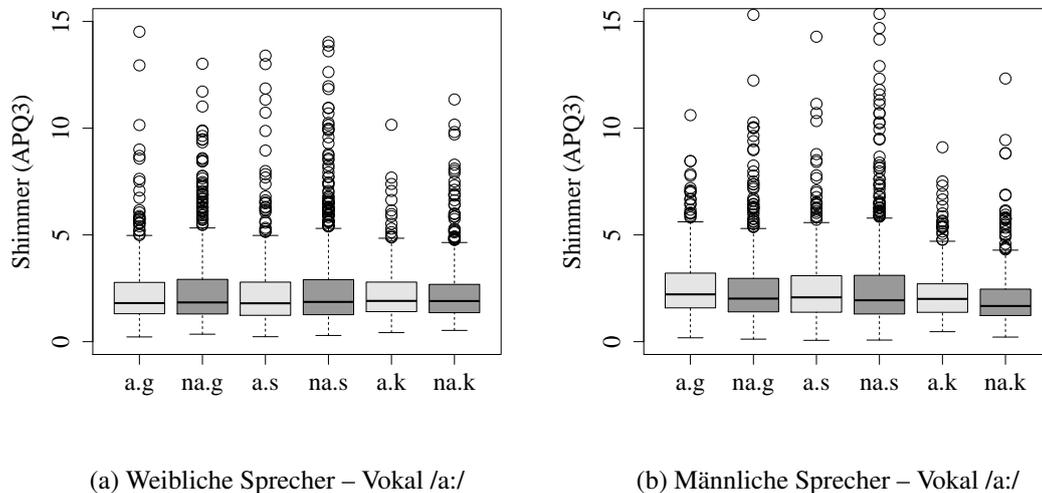
/a:/		
	weibliche Sprecher	männliche Sprecher
Alkoholisierung	$\chi^2[1] = 23,05, \quad p < 0,001 \text{ ***} \downarrow$	$\chi^2[1] = 6,77, \quad p < 0,01 \text{ **} \downarrow$
Sprachstil	$\chi^2[2] = 13,41, \quad p < 0,01 \text{ **}$	$\chi^2[2] = 51,88, \quad p < 0,001 \text{ ***}$
/i:/		
	weibliche Sprecher	männliche Sprecher
Alkoholisierung	$\chi^2[1] = 0,87, \quad p > 0,1 \text{ n.s.}$	$\chi^2[1] = 0,04, \quad p > 0,1 \text{ n.s.}$
Sprachstil	$\chi^2[2] = 30,97, \quad p < 0,001 \text{ ***}$	$\chi^2[2] = 32, \quad p < 0,001 \text{ ***}$

Tab. 3.6: Ergebnisse der MEM-Analyse – Jitter (RAP) – \downarrow = niedriger in alkoholisiertem Zustand, n.s. = nicht signifikant, . = signifikanter Trend, * = signifikant, ** = hoch signifikant, *** = höchst signifikant

noch in die andere Richtung signifikant sind. Die Ergebnisse der vier Modelle sind in Tabelle 3.6 zusammengefasst.

3.3.3.2 Shimmer

Auch hier wurden die Daten nach Alkoholisierung, Sprechergeschlecht, Vokal und Sprachstil aufgeteilt, Boxplots dazu sind in den Abb. 3.13a und 3.13b für den Vokal /a:/ und in den Abb. 3.14a und 3.14b für den Vokal /i:/ zu sehen. Und ebenso wie



(a) Weibliche Sprecher – Vokal /a:/

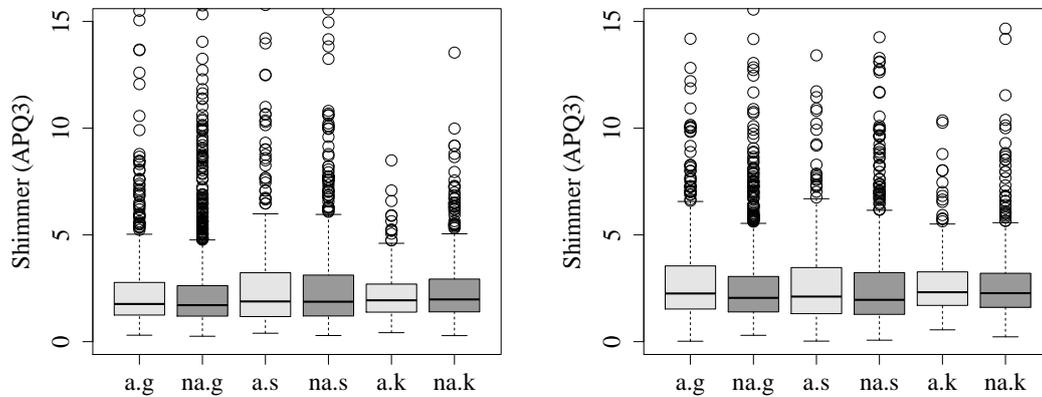
(b) Männliche Sprecher – Vokal /a:/

Abb. 3.13: Shimmer (APQ3) getrennt nach Alkoholisierung und Sprachstil – *a* = *alkoholisiert*, *na* = *nüchtern*, *g* = *gelesene Sprache*, *s* = *Spontansprache*, *k* = *Kommandosprache*

beim Jitter sind viele Ausreißer und eine starke Streuung der Werte zu erkennen. Zur besseren Übersicht wurde hier nur der Bereich von 0 – 15 abgebildet.

Bei weiblichen Sprechern im Vokal /a:/ sind kaum Unterschiede zwischen den Aufnahmen in nüchternem und denen in alkoholisiertem Zustand auszumachen, bei den männlichen Sprechern zeigt sich ein höherer Shimmerwert in alkoholisiertem Zustand, am deutlichsten in der Kommandosprache. Im Vokal /i:/ zeigt sich für weibliche Sprecher ein ähnliches Bild wie im Vokal /a:/: auch hier sind kaum Unterschiede erkennbar. Bei männlichen Sprechern ist auch hier im Mittel ein höherer Shimmerwert unter Alkoholeinfluss messbar, der Unterschied zeigt sich allerdings in der gelesenen Sprache am deutlichsten und in der Kommandosprache dagegen kaum noch.

Auch hier zeigt eine *MEM*-Analyse mit allen Faktoren viele signifikante Interaktionen, deswegen wurden ebenfalls getrennte Modelle für die unterschiedlichen Vokale und Sprechergeschlechter gerechnet. Bei weiblichen Sprechern ist sowohl im Vokal /a:/ als auch im Vokal /i:/ kein signifikanter Unterschied zwischen den Shimmerwerten der Aufnahmen in alkoholisiertem Zustand und denen im nüchternen Zustand nachweisbar, die Sprachstile unterscheiden sich nur im Vokal /i:/ signifikant, was allerdings auch hier nicht von großer Bedeutung ist. Signifikante Interaktionen traten



(a) Weibliche Sprecher – Vokal /i:/

(b) Männliche Sprecher – Vokal /i:/

Abb. 3.14: Shimmer (APQ3) getrennt nach Alkoholisierung und Sprachstil – a = alkoholisiert, na = nüchtern, g = gelesene Sprache, s = Spontansprache, k = Kommandosprache

nicht auf. Die Shimmerwerte der männlichen Sprecher unterscheiden sich im Vokal /a:/ (schwach) signifikant zwischen den beiden Aufnahmesituationen, im Vokal /i:/ zeigten sich signifikante Interaktionen zwischen der Alkoholisierung und dem Sprachstil ($\chi^2[2] = 7,35, p < 0,05$), ein *post-hoc*-Test bestätigt aber einen signifikant höheren Shimmerwert in alkoholisiertem Zustand in der gelesenen Sprache. Im Vokal /a:/ traten bei den männlichen Sprechern keine signifikanten Interaktionen auf und die Shimmerwerte der verschiedenen Sprachstile unterscheiden sich signifikant voneinander. Die Ergebnisse der *MEM*-Analyse und des *post-hoc*-Tests sind in Tabelle 3.7 zusammengefasst.

3.3.4 Zwischendiskussion

Künzel et al. (1992) konnten einen deutlich höheren Jitter (*RAP*) in alkoholisiertem Zustand in ihrer Studie mit 33 männlichen Probanden anhand des Vokals /i:/ messen. Dieses Ergebnis lässt sich anhand der hier untersuchten Sprecher des ALC nicht bestätigen. Hypothese 1 wird somit verworfen. Im Vokal /i:/ bei den männlichen Sprechern ist kein signifikanter Unterschied zwischen den Aufnahmen in alkoholisiertem und denen im nüchternen Zustand erkennbar. Lediglich in der Kommandosprache lässt

/a:/		
	weibliche Sprecher	männliche Sprecher
Alkoholisierung	$\chi^2[1] = 1,37, \quad p > 0,1 \text{ n.s.}$	$\chi^2[1] = 4,28, \quad p < 0,05 \text{ * } \uparrow$
Sprachstil	$\chi^2[2] = 4,25, \quad p > 0,1 \text{ n.s.}$	$\chi^2[2] = 34,86, \quad p < 0,001 \text{ ***}$
/i:/		
	weibliche Sprecher	männliche Sprecher
Alkoholisierung	$\chi^2[1] = 0,06, \quad p > 0,1 \text{ n.s.}$	g: $p > 0,01 \text{ ** } \uparrow$ s: $p > 0,1 \text{ n.s.}$
Sprachstil	$\chi^2[2] = 13,24, \quad p < 0,01 \text{ **}$	k: $p > 0,1 \text{ n.s.}$

Tab. 3.7: Ergebnisse der MEM-Analyse und des *post-hoc*-Tests für den Vokal /i:/ bei männlichen Sprechern – Shimmer (APQ3) – g = *gelesene Sprache*, s = *Spontansprache*, k = *Kommandosprache*, \uparrow = *höher in alkoholisiertem Zustand*, n.s. = *nicht signifikant*, . = *signifikanter Trend*, * = *signifikant*, ** = *hoch signifikant*, *** = *höchst signifikant*

sich ein derartiger (nicht signifikanter) Trend beobachten. Für den Vokal /a:/ zeigt sich in dieser Studie sogar ein gegenläufiges Ergebnis. Hier konnte ein signifikant niedrigerer Jitterwert in den Aufnahmen in alkoholisiertem Zustand gemessen werden, sowohl für weibliche als auch für männliche Sprecher, der Effekt tritt aber deutlicher bei weiblichen Sprechern auf. Trotz der gemessenen signifikant niedrigeren Jitterwerte in den Aufnahmen der Sprecher in alkoholisiertem Zustand, widersprechen diese Ergebnisse nicht zwangsläufig denen von Künzel et al. (1992), da diese in dem Vokal /a:/ auftraten, der nicht Gegenstand der Analyse in der Studie von Künzel et al. war. Brockmann-Bauser (2012) beschreibt zwar mögliche Unterschiede der Jitterwerte zwischen verschiedenen Vokalen, berichtet allerdings auch von widersprüchlichen Ergebnissen in der bisherigen Literatur. In der Studie von Johnson et al. (1993)⁴ wurden allerdings auch im Vokal /a:/ höhere Jitter- und auch Shimmerwerte in Aufnahmen von Sprechern in alkoholisiertem Zustand gemessen, allerdings bildeten hier nur Aufnahmen von vier Sprechern die Datenbasis. Auch mit den Ergebnissen von Cummings et al. (1995) sind die Ergebnisse dieser Studie nur schwer zu vergleichen. Sie untersuchten zwar ebenfalls Unterschiede in verschiedenen Jitter- und Shimmerparametern zwischen Aufnahmen von Sprechern in alkoholisiertem Zustand und Aufnahmen derselben Sprecher in nüchternem Zustand, allerdings untersuchten sie nur vier Sprecher

⁴ zitiert nach Pisoni et al. (1985)

Hypothese	Angenommen?
1. Der Jitterwert eines Sprechers ist höher in den Aufnahmen unter Alkoholeinfluss als in den Aufnahmen ohne Alkoholeinfluss.	X
2. Der Shimmerwert eines Sprechers ist höher in den Aufnahmen unter Alkoholeinfluss als in den Aufnahmen ohne Alkoholeinfluss.	X

Tab. 3.8: Übersicht über die Hypothesentests

und unterteilten die Daten in ihren Ergebnissen nicht nach unterschiedlichen Vokalen. Die vier Sprecher wiesen mehrheitlich höhere Jitterwerte in alkoholisiertem Zustand auf und zeigten bezüglich des Shimmers keine konsistenten Muster, wenn auch zum Teil die Ergebnisse innerhalb eines Sprechers signifikant waren. Auch in der in dieser Arbeit vorgestellten Studie lässt sich kaum ein Unterschied in den Shimmerwerten zwischen den Aufnahmen in alkoholisiertem und denen in nüchternem Zustand feststellen. Schwach signifikant größere Shimmerwerte bei männlichen Sprechern im Vokal /a:/ in alkoholisiertem Zustand, im Vokal /i:/ zeigen diese nur in der gelesenen Sprache Signifikanz. Bei weiblichen Sprechern ist dagegen nicht einmal ein signifikanter Trend in eine Richtung erkennbar. Hypothese 2 wird an dieser Stelle also auch verworfen. In Tabelle 3.8 sind die Hypothesentests zusammengefasst.

Insgesamt wird deutlich, dass die Ergebnisse der Jitter- und Shimmermessungen im ALC nicht besonders gut geeignet sind, um Sprache unter Alkoholeinfluss eines Sprechers zu identifizieren. Die Daten streuen sehr stark und sind weder für die Sprechergeschlechter, noch für die beiden untersuchten Vokale konsistent.

4

Perzeption der Sprache alkoholisierter Sprecher

Im folgenden Kapitel werden im Wesentlichen drei Perzeptionstests zur Wahrnehmbarkeit der Alkoholisierung eines Sprechers anhand seiner Sprache vorgestellt. Zunächst folgt ein Literaturüberblick in Kapitel 4.1, in dem bisherige Studien vorgestellt werden und eine Zusammenfassung der Ergebnisse dieser Studien in Tabelle 4.1. In Kapitel 4.2 wird ein eigens durchgeführter Diskriminationstest mit 161 Sprechern des ALC beschrieben und die Ergebnisse in Kapitel 4.2.3 erläutert. Anschließend werden die Ergebnisse in Kapitel 4.2.4 diskutiert und dann in Kapitel 4.3 zwei weitere Perzeptionstests, in diesem Fall mit manipulierten Stimuli, vorgestellt. Zum einen werden in einem Experiment die in Kapitel 3.2 erläuterten F_0 -Effekte in unter Alkoholeinfluss produzierten Stimuli kompensiert und zum anderen in einem Experiment in nüchtern artikulierten Stimuli simuliert. Die Ergebnisse dieser Tests werden in Kapitel 4.3.3 beschrieben und abschließend in Kapitel 4.3.4 diskutiert.

4.1 Literaturüberblick

Wie in Kapitel 3.1 bereits erwähnt, haben Sobell et al. (1982) einige Studien zum Thema Alkohol und Sprache durchgeführt und in diesem Rahmen auch einen Perzepti-

onstest vorgestellt, in dem sieben Hörer Sprachbeispiele von 16 männlichen Sprechern in drei verschiedenen Stadien (nüchtern, mäßig alkoholisiert mit AAK von 0,16‰ – 0,51‰ und stark alkoholisiert mit AAK von 0,45‰ – 1,17‰) beurteilten. Hier wurde nicht nur nach der Trunkenheit gefragt, sondern auch nach Einschätzungen über Artikulation, Nasalität, Tonfall und Sprechgeschwindigkeit. Die Hörer antworteten jeweils auf einem 15 mm langen Einschätzungsbalken mit der Beschriftung ‚sehr gut‘ am einen und ‚schlecht‘ am anderen Ende. Signifikant unterschieden sich die drei Alkoholisierungsstadien in der Beurteilung der Artikulation, der Sprechgeschwindigkeit und der Trunkenheit, nicht aber in Nasalität und Tonfall. Die Zuhörer empfanden die Beispiele eines stark alkoholisierten Sprechers als schlechter artikuliert, langsamer gesprochen und hörbar betrunkenener als in den beiden anderen Alkoholisierungsstadien.

Pisoni et al. (1985) führten ebenfalls neben der Untersuchung einiger akustischer Parameter (vgl. Kapitel 3.1) einen Perzeptionstest mit Aufnahmen von alkoholisierten Sprechern durch. Das Sprachmaterial bestand hier aus jeweils 34 kurzen Sätzen von vier männlichen Sprechern, die sowohl nüchtern als auch alkoholisiert aufgenommen wurden. Die AAK der Sprecher lag zwischen 1,0‰ und 1,7‰. Die Aufgabe für 21 studentische Hörer, die diesen Test als Teil einer Lehrveranstaltung absolvierten, bestand darin, aus zwei (gleichen) Äußerungen eines Sprechers diejenige Äußerung auszuwählen, die unter Alkoholeinfluss produziert wurde – sie bearbeiteten also einen *forced choice*¹ Diskriminationstest. Zusätzlich sollten die Hörer die Sicherheit ihrer Einschätzung auf einer Skala von 1 – 5 bewerten. Die 21 Hörer erreichten in diesem Experiment eine relativ hohe Diskriminationsrate von 73,8%, die für die einzelnen Sprecher zwischen 70% und 80,3% variierte, allerdings konnte kein Zusammenhang zwischen dem Alkoholisierungsgrad der Sprecher und deren Beurteilung erkannt werden. Die Selbsteinschätzung der Hörer korrelierte stark mit dem Anteil an richtigen Antworten in der jeweiligen Sicherheitsstufe. Zusätzlich führten Pisoni et al. (1985) noch einen weiteren (Vor-)Test mit sechs Angestellten des Sprachforschungsinstituts

¹ Der Proband wird *gezwungen* sich zwischen Antwortmöglichkeiten zu entscheiden, ohne die Option, keine Entscheidung zu treffen.

durch, allerdings nur mit Sprachmaterial von zwei Sprechern. Die sechs Hörer in diesem Experiment erkannten in 82,4% der Fälle den unter Alkoholeinfluss produzierten Stimulus in den dargebotenen Paaren als solchen, erreichten also eine noch höhere Diskriminationsrate als die naiven Hörer.

In einem späteren Experiment mit vier weiteren, unter den gleichen Bedingungen aufgenommenen, Sprechern präsentierten Martin & Yuchtman (1986) den Hörern keine Stimuluspaare, sondern einzelne Stimuli (auch hier wieder kurze Sätze), die entweder als unter Alkoholeinfluss produziert oder nüchtern markiert werden sollten. In diesem Fall beurteilten zwei unterschiedliche Hörergruppen die Aufnahmen. Die erste Gruppe bestand aus 30 Psychologiestudenten, die zweite bildete eine Gruppe von 14 Polizisten. Jede Versuchsperson beurteilte 24 Stimuli pro Versuchsperson, wobei der Anteil an nüchtern und unter Alkoholeinfluss produzierter Sprache ausgeglichen war, jedoch nicht die gleichen Sätze beinhaltete. Auch hier wurde eine Selbsteinschätzung der Hörer auf einer Skala von 1 – 5 abgefragt. Die Gruppe der Polizisten schnitt in diesem Test mit einer durchschnittlichen Identifikationsrate von 64,66% signifikant besser ab als die der Psychologiestudenten mit einer Identifikationsrate von 61,48%. Martin & Yuchtman (1986) erklären diesen Unterschied mit der Erfahrung der Polizisten im Umgang und der Einschätzung von alkoholisierten Personen (vor allem im Straßenverkehr). Auch wenn die Ergebnisse für alle Sprecher signifikant über dem Zufallsniveau lagen, zeigt sich doch eine starke Variation zwischen den einzelnen Sprechern. In diesem Experiment reichen die durchschnittlichen Erkennungsraten von 55% bis 71,9%. Auch hier korrelierten die Selbsteinschätzungen der Hörer mit ihrem tatsächlichen Erfolg.

Auch Klingholz et al. (1988) untersuchten mit einer Gruppe von 12 Sprachtherapeuten die Fähigkeit einer eher erfahrenen Hörergruppe, Sprache unter Alkoholeinfluss zu erkennen. Allerdings schnitten die Therapeuten hier mit einer Diskriminationsrate von 61,1% und einer Identifikationsrate von nur 54% wesentlich schlechter ab als die Polizisten in Martin & Yuchtman (1986) oder die Sprachwissenschaftler im Vortest in Pisoni et al. (1985). Gegenstand der Untersuchungen von Klingholz et al. waren (wie

bereits in Kapitel 3.1 beschrieben) 16 männliche Sprecher (5 davon als Kontrollgruppe aufgenommen) mit Blutalkoholwerten zwischen 0,67‰ und 1,59‰. Auf den ersten Blick scheint das Ergebnis trotz der erfahrenen Hörer vor allem im Vergleich zu dem Ergebnis der Gruppe der Polizisten in Martin & Yuchtman (1986) erstaunlich schlecht zu sein. Bei einer Unterteilung der Ergebnisse für Sprecher mit einer BAK unter 1,0‰ und Sprecher mit einer BAK über 1,0‰ erreichten die Hörer in der Gruppe der Sprecher mit hoher BAK aber sogar eine Diskriminationsrate von 82% (im Gegensatz zu 54,2% in der Gruppe mit den Sprechern mit niedriger BAK). Klingholz et al. baten ihre Hörer zusätzlich um eine Angabe darüber, auf welches Merkmal sie ihre Entscheidung gestützt haben und erhielten signifikant mehr Nennungen artikulatorischer Merkmale (wie Sprachfluss, Sprachqualität und Sprechfehler) als Nennungen stimmlicher Merkmale (wie Stimmqualität, Stimmstabilität und Stimmumfang, also ‚vocal effort‘).

Künzel et al. (1992) führten sehr umfangreiche Untersuchungen zum Einfluss von Alkohol auf Sprache und Stimme durch (vgl. Kapitel 3.1) und beschreiben in ihrem Buch auch einen Perzeptionstest. Als Basis hierfür fungierten die Aufnahmen von 33 männlichen Polizisten, die in verschiedenen Alkoholisierungsgraden aufgezeichnet wurden. Sprachmaterial für den Perzeptionstest war von jedem Sprecher jeweils der erste Satz eines vorgelesenen Textes (ca. 10 – 12 Sekunden lang), einmal in nüchternem Zustand und einmal der Aufnahme zum Zeitpunkt der höchsten gemessenen AAK entnommen. Die Autoren teilen ihre Sprecher nach ihrem individuellen Maximalwert der AAK in Gruppen von < 0,4‰ in 0,4‰-Schritten bis > 2,0‰ ein. 30 Hörer wurden nun gebeten, in einem Identifikationstest für jeden Stimulus zu entscheiden, ob dieser unter Alkoholeinfluss produziert wurde oder nicht. Zusätzlich fragten sie die Selbsteinschätzung der Sicherheit ihres Urteils auf einer Skala von 0 („ich bin völlig unsicher“) bis 4 („ich bin absolut sicher“) ab. Insgesamt erreichten die Hörer hier eine Identifikationsrate von 66,8%, jedoch ist diese stark abhängig von dem Alkoholisierungsgrad der Sprecher und schwankt von 51,1% in der Klasse mit Sprechern mit einer AAK unter 0,2‰ bis zu 95,8% in der Gruppe der Sprecher mit einer AAK von mehr als 2,0‰. In dieser stark alkoholisierten Gruppe befanden sich allerdings nur zwei Sprecher. Zu-

sätzlich zeigt dieser Test, dass die Sicherheit der Hörerurteile bei Sprechern ab einer AAK von 0,8‰ stetig ansteigt, bzw. bei falschen Antworten sinkt.

Hollien führte ebenfalls zusammen mit unterschiedlichen Koautoren neben den in Kapitel 3.1 beschriebenen akustischen Studien zu Sprache unter Alkoholeinfluss auch mehrere Perzeptionsexperimente durch. Eine Besonderheit der Sprachaufnahmen von Hollien et al. ist, dass von jedem Sprecher Aufnahmen zu unterschiedlichen Zeitpunkten, mit unterschiedlichen (kontrollierten) AAK existieren (vgl. Kapitel 3.1). Im ersten Perzeptionstest, beschrieben in Hollien & Martin (1996), ließen die Autoren 50 Hörer Sprachbeispiele von 8 Sprechern (nüchtern und zwischen 1,2‰ und 1,3‰) auf einer Skala von 1 – 5 einschätzen, wobei 1 ‚nüchtern‘ und 5 ‚stark betrunken‘ darstellen sollte. Zwar wurden die im nüchternen Zustand gesprochenen Beispiele für alle Sprecher auf der Skala niedriger eingeschätzt als die unter Alkoholeinfluss produzierten, der durchschnittliche Unterschied betrug allerdings lediglich 1,3 und wird somit als nicht signifikant eingestuft. Die Aufgabe für 27 Hörer im zweiten Test, der in Hollien & Martin (1996) beschrieben wird, bestand darin, alle Sprachbeispiele von 8 Sprechern (einmal nüchtern, einmal mit einer AAK von 1,2‰ – 1,3‰ artikuliert) in zwei Gruppen zu unterteilen, jedoch ohne den Hörern Auskunft darüber zu geben, dass jeweils ein Stimulus pro Sprecher in alkoholisiertem Zustand gesprochen wurde. In diesem Fall bestanden die Sprachbeispiele aus diadochokinetischen Übungen, hier ‚pataka‘, ‚shapupi‘ und ‚buttercup‘. Im Durchschnitt wurden 66% der nüchtern artikulierten Stimuli derselben Gruppe zugeordnet, jedoch bezeichnen die Autoren das Ergebnis als nicht besonders ermutigend. Auch das vorherige Benennen der Gruppen in nüchtern und alkoholisiert produzierte Sprache und die Mitteilung an die Hörer, es handle sich um jeweils ein Beispiel pro Kondition für jeden Sprecher, führte nicht zu besseren Ergebnissen. Erst der vierte und letzte Test, den die Autoren in Hollien & Martin (1996) vorstellen, zeigte bessere Resultate der Hörer. Hier wurden die Hörer zu keiner Entscheidung gezwungen, sondern konnten zusätzlich die Kategorie ‚no decision‘, also ‚keine Entscheidung‘ auswählen. Diese nutzten die 17 Hörer allerdings nur für ca. 6% der Stimuli, und in der Gruppe der nüchtern artikulierten Stimuli waren durchschnitt-

lich 68% der zugeordneten Stimuli auch tatsächlich in nüchternem Zustand aufgenommene Stimuli.

In einem weiteren Artikel der Forschergruppe um Hollien (Hollien et al. 2009), stellen die Autoren weiterführende Perzeptionsexperimente vor. Im ersten Experiment dienen Aufnahmen von 48 Sprechern in nüchternem, sowie in drei unterschiedlich stark alkoholisierten Zuständen (0,4‰, 0,8‰ und 1,2‰) als Teststimuli. Die Aufnahmebedingungen bleiben hier dieselben wie in Hollien & Martin (1996). Diese Stimuli wurden 128 Hörern randomisiert und einzeln präsentiert, woraufhin diese zu jedem Stimulus eine Einschätzung auf einer Skala von 1 – 5 (1 = ‚nüchtern‘, 5 = ‚stark alkoholisiert‘) abgaben. Betrachtet man die Antworten der Hörer als simple nüchtern/alkoholisierter Sprecher-Entscheidung, erreichen die Hörer eine Erkennungsleistung von 82%. Allerdings ist dieses sehr hohe Ergebnis aufgrund des Verhältnisses zwischen nüchtern und alkoholisiert produzierten Stimuli und der Einschätzung auf einer Skala nicht mit den Ergebnissen der anderen Identifikationstests zu vergleichen. Die Auswertung der Einschätzungen der Hörer auf der Skala zeigen allerdings, dass die Hörer alle drei präsentierten Trunkenheitsgrade der Sprecher signifikant unterschiedlich voneinander einordnen. Die Hörer überschätzen allerdings die Trunkenheit der Sprecher in den Stimuli, in denen ein Sprecher leicht alkoholisiert war und unterschätzen umgekehrt den Trunkenheitsgrad in den Stimuli, in denen ein Sprecher stark alkoholisiert war. In einem zweiten Experiment in Hollien et al. (2009) sollte die Fähigkeit von Laien und Experten in Bezug auf den Umgang mit Betrunkenen getestet werden. Als Experten fungieren hier 20 Krankenhausärzte im Vergleich zu 18 Studenten. Die Autoren vermuten bei den Krankenhausärzten einen Vorteil durch den beruflich bedingten häufigeren Umgang mit teilweise stark alkoholisierten Personen. Auch hier diente eine Skala von 1 – 5 als Mittel zur Einschätzung, allerdings beurteilten die Hörer nur zwei Stimuli pro Sprecher – einmal in nüchternem Zustand aufgenommen und einmal stark alkoholisiert (mit einer AAK von 1,2‰). Gegensätzlich zu den Erwartungen der Autoren, zeigt sich zwischen den Einschätzungen der Experten und denen der Studenten ein sehr geringer Unterschied. Es ist lediglich erkennbar, dass die Ex-

perten insgesamt dazu tendieren, niedrigere Werte auf der Skala auszuwählen. Der Bewertungsunterschied ist zwar jeweils signifikant zwischen den Sprecherzuständen innerhalb der Hörergruppen, jedoch nicht signifikant zwischen den Hörergruppen. Zumindest in diesem Experiment scheinen Experten also keinen Vorteil gegenüber Laien in der Beurteilung von Alkoholisierung anhand von Sprache zu haben. Ein weiterer Test soll einen eventuellen Unterschied in der Wahrnehmbarkeit von Alkoholisierung zwischen gelesener Sprache und Spontansprache prüfen. Die Stimuli der gesprochenen Sprache von 36 Sprechern waren in diesem Fall Antworten auf Fragen wie z.B. „Erzählen Sie uns etwas über Ihre liebste Fernsehsendung!“. Alle Äußerungen wurden von den Sprechern wieder in nüchternem Zustand und in den drei unterschiedlichen Alkoholisierungsgraden aufgezeichnet und 101 Hörern präsentiert. Hollien et al. vermuten die Einschätzung der Trunkenheit im Fall der Spontansprache höher auf der Skala und begründen dies damit, dass der gelesene Text bekannt war und mehrfach geübt wurde und damit eine einfachere Aufgabe für die Sprecher darstellt als auf unbekannte Fragen zu antworten. Tatsächlich wurden die spontansprachlichen Äußerungen im Mittel in allen alkoholisierten Zuständen mit signifikant höheren Punkten auf der Skala markiert als die Stimuli der gelesenen Sprache.

Wie in Kapitel 3.1 bereits beschrieben, führten (Hollien et al. 1998) einige Studien mit Schauspielern durch. Hauptziel war es herauszufinden, inwieweit Hörer gespielte Trunkenheit als echte Trunkenheit ansehen, und ob Schauspieler dazu in der Lage sind, in stark alkoholisiertem Zustand noch nüchtern zu klingen. Im ersten Experiment wurden 40 Hörern (33 weiblichen und 7 männlichen) paarweise Stimuli eines Sprechers dargeboten, jeweils ein Stimulus in stark alkoholisiertem Zustand und ein Stimulus, in dem vom Sprecher starke Trunkenheit simuliert wurde, er tatsächlich aber nüchtern war. Die Aufgabe der Hörer war es, den Stimulus auszuwählen, in dem der Sprecher betrunkenere klang. Tatsächlich wählten die Hörer in diesem Experiment in 88% der Fälle die simulierte Trunkenheit aus. Im zweiten Experiment setzten sich die Stimuluspaare für jeden Sprecher aus zwei seiner Aufnahmen in stark alkoholisiertem Zustand zusammen, jedoch war der Sprecher bei einer Aufnahme angehalten, möglichst

nüchtern zu klingen. Auch hier bestand die Aufgabe derselben Hörer darin, jeweils den betrunkenen klingenden Stimulus aus einem dargebotenen Paar auszuwählen. Hier zeigt sich ein weniger deutliches Ergebnis als im ersten Test, im Durchschnitt wählten die Hörer in 39,2% der Stimuluspaare den Stimulus aus, bei dessen Aufnahme die Sprecher versuchten, ihre Trunkenheit zu maskieren. Die Ergebnisse in diesem Experiment sind allerdings stark sprecherabhängig – scheinbar sind einige Sprecher sehr gut dazu in der Lage, selbst in stark alkoholisiertem Zustand in ihrer Sprache Nüchternheit vorzutäuschen, andere jedoch nicht. Die Präferenz der Hörer für den Stimulus ohne Verdeckungsintention variiert je nach Sprecher zwischen 5% und 95%. Zwar waren die Sprecher in den Experimenten in Hollien et al. (1998) alle Schauspieler, jedoch wird deutlich, dass Trunkenheit sowohl überzeugend simuliert, als auch verdeckt werden kann – dies jedoch stark sprecherabhängig.

Im Rahmen einer Magisterarbeit am Institut für Phonetik und Sprachverarbeitung der Ludwig-Maximilians-Universität München (Aulich 2010) wurde bereits ein Perzeptionsexperiment mit Aufnahmen einiger Sprecher des ALC durchgeführt. Schwerpunkt dieser Arbeit lag auf dem Vergleich der Ergebnisse eines Perzeptionstests von invers gefilterten Stimuli und unbearbeiteten Stimuli. Hierfür wurden 16 Sprecher (8 weibliche und 8 männliche) aus dem ALC ausgewählt. Die 16 Sprecher weisen alle eine BAK von mindestens 0,5‰ auf, gaben an „moderate Trinker“ (vgl. Kapitel 2) zu sein und waren zum Zeitpunkt der Aufnahmen zwischen 24 und 30 Jahre alt (das Durchschnittsalter liegt bei 26,5 Jahren). Die BAK der Sprecher bewegte sich zwischen 0,5‰ und 1,42‰, bei einem Durchschnittswert von 0,94‰. Als Material für den Perzeptionstest wurden pro Sprecher zwei spontansprachliche und zwei gelesene Äußerungen im Original verwendet und zusätzlich jeweils eine invers gefilterte spontansprachliche und eine gelesene Äußerung pro Sprecher. Mit dem Verfahren der inversen Filterung mittels LPC-Analyse werden Quelle (Stimmklappen) und Filter (Ansatzrohr) voneinander getrennt und so nur das Anregungssignal der Quelle, also das Glottissignal genutzt. Die Intention, die diesem Vorgehen zugrunde lag, bestand darin, dass so alle linguistischen Merkmale der Sprache (wie z.B. Wortwahl und Verspre-

cher) nicht mehr hörbar sind und so keinen Einfluss auf die alkoholisiert/nüchtern-Entscheidung der Hörer haben, sondern nur noch die Intonation und der Rhythmus eine Rolle spielen. Insgesamt waren also 96 ($16 * (4 + 2)$) Stimuluspaare (alkoholisiert und nüchtern) Teil des Perzeptionsexperiments. Der Perzeptionstest selbst war ein *forced-choice* Diskriminationstest mit einer zusätzlichen Sicherheitsabfrage auf einer Skala von 1–7. An diesem Test nahmen insgesamt 38 Hörer teil, wobei die Autorin zwischen sprachwissenschaftlich naiven Hörern (28 Hörer, 14 weibliche und 14 männliche; Alter zwischen 20 und 30 Jahren, im Durchschnitt 23,5 Jahre) und Phonetikern (19 Hörer, 16 weibliche und 3 männliche; Alter zwischen 22 und 39 Jahren, im Durchschnitt 27,5 Jahre) unterscheidet. Bei den unveränderten Stimuli erreichten alle Hörer zusammengenommen eine Diskriminationsrate von 71,7%, also höchst signifikant über dem Zufallsniveau von 50%. Auffällig ist hier, dass ebenfalls ein signifikanter Geschlechterunterschied besteht, allerdings nicht in der Erkennungsleistung der Hörer – hier erreichten die Hörergruppen im Durchschnitt sehr ähnliche Diskriminationsraten von 71,5% (weibliche) und 71,8% (männliche Hörer) – sondern in der Erkennungsrate für die einzelnen Sprechergruppen. Weibliche Sprecher wurden von allen Hörern im Durchschnitt in 77,1% der Fälle richtig erkannt und somit signifikant besser als männliche Sprecher, in deren Stimuluspaaren von allen Hörern nur in 66,2% der Fälle der richtige Stimulus als der unter Alkoholeinfluss produzierte ausgewählt wurde. Die Alkoholisierung der weiblichen Sprecher scheint in diesem Fall also deutlicher hörbar zu sein als die der männlichen Sprecher. Bezogen auf die beiden Sprachstile zeigte sich in diesem Experiment, dass gelesene Sprache signifikant besser (75,7%) zuzuordnen war als die spontansprachlichen Äußerungen (68,1%). Die Unterschiede zwischen sprachwissenschaftlich naiven Hörern und Phonetikern lassen erkennen, dass die sprachwissenschaftlich naiven Hörer sogar ein wenig besser abgeschnitten haben als die Phonetiker (73% bzw. 70,3%), allerdings ist der Unterschied relativ gering.

Bei den invers gefilterten Stimuli erzielten die Hörer weitaus schlechtere Erkennungsleistungen. Hier konnten die Hörer bei nur 56,9% der Stimuluspaare den unter Alkoholeinfluss produzierten Stimulus richtig auswählen. Diese Erkennungsleistung

liegt zwar immernoch signifikant über dem Zufallsniveau, fällt allerdings im Gegensatz zu den Originalstimuli um 14,8 Prozentpunkte. Geschlechterunterschiede sind bei den invers gefilterten Stimuli nicht mehr zu erkennen. Die schlechte Erkennungsleistung bei den invers gefilterten Stimuli – also bei den Stimuli, bei denen lediglich prosodische Informationen und keine lexikalischen mehr hörbar sind – zeigt, dass die Prosodie zwar einige Informationen im Sprachsignal beinhaltet, auf die die Hörer sich in ihrer Entscheidung (richtig) stützen (die Erkennungsleistung liegt ja dennoch über dem Zufallsniveau), trotzdem scheinen andere Faktoren eine größere Rolle zu spielen und auch zu mehr Erfolg zu führen. Andererseits muss hier auch angemerkt werden, dass ein invers gefilterter Stimulus für den Hörer verwirrend klingen kann, da wir es im Alltag nicht gewohnt sind, eine Äußerung ohne lexikalischen Inhalt zu hören.

Eine Korrelation der Erkennungsleistung der Hörer in den unveränderten Stimuli mit der jeweiligen BAK der Sprecher zeigt mit $r = 0,19$ keinen signifikanten linearen Zusammenhang. Die Sicherheitseinschätzung der Hörer steht in etwas stärkerem Zusammenhang mit den gemessenen BAK der jeweiligen Sprecher, hier liegt der Korrelationskoeffizient bei $r = 0,33$. Den Hörern fiel es also tendenziell leichter, Sprecher mit höherer BAK einzuschätzen, jedoch erhöht sich die Anzahl richtiger Antworten hier nicht signifikant. Die Verteilung der Erkennungsraten der einzelnen Sprecher zeigt zwar eine starke Variation in der Einschätzbarkeit der Sprecher, jedoch kann diese nicht an den unterschiedlichen BAK festgemacht werden.

Mit einem Teil der Daten, der für eine Interspeech Speaker State Challenge (Schuller et al. (2011), eine Zusammenfassung der Ergebnisse ist in Schuller et al. (2014) nachzulesen) aus dem ALC zur Verfügung gestellt wurde, beschreiben Ultes, Schmitt & Minker (2011) einen durchgeführten Identifikationstest. 79 Hörer (31 weibliche und 48 männliche) unterteilten insgesamt 3200 zufällig ausgewählte Stimuli aus dem ALC (jeder einzelne Hörer allerdings nur 50) in die beiden Gruppen ‚nüchtern‘ und ‚alkoholisiert‘. Wieviel Material von wievielen Sprechern beurteilt wurde, wird leider nicht deutlich. Insgesamt erreichten die Hörer eine durchschnittliche Identifikationsrate von 55,8%, die unter den Hörern zwischen 38,3% und 70% schwankte. Neben dieser sehr

großen Hörerabhängigkeit konnten die Autoren auch einen Trend in der Erkennbarkeit abhängig von der BAK erkennen. Nüchterne und stark alkoholisierte Sprecher wurden von den Hörern am häufigsten der richtigen Gruppe, schwach alkoholisierte Sprecher hingegen häufig falsch oder scheinbar zufällig den beiden Gruppen zugeordnet. Die Länge der Äußerungen hatte laut den Analysen von Ultes et al. (2011) in diesem Experiment keine Auswirkungen auf die Ergebnisse.

Zusammenfassend lässt sich sagen, dass in den bisherigen Studien sowohl in den Identifikationstests als auch in den Diskriminationstests immer eine Rate von mehr als 50% in der Erkennung der Alkoholisierung der Sprecher erreicht werden konnte. Bei den Diskriminationstests liegt die Rate im Durchschnitt bei ca. 60-65%, wesentlich bessere Ergebnisse konnten nur erreicht werden, wenn Sprecher mit einer sehr hohen AAK bzw. BAK beurteilt wurden (z.B. Pisoni et al. (1985), Klingholz et al. (1988)). Beruflich bedingte Erfahrung führte nicht immer zu besseren Einschätzungen der Alkoholisierung als bei Laien. Die Polizisten in Martin & Yuchtman (1986) schnitten besser ab als die Vergleichsgruppe der Studenten, bei Krankenhausärzten in (Hollien et al. 2009) konnte allerdings kein Unterschied zur Vergleichsgruppe, die ebenfalls aus Studenten bestand, festgestellt werden. Bei Aulich (2010) erzielten die sprachwissenschaftlich naiven Hörer sogar bessere Ergebnisse als die untersuchten Phonetiker. Ein weiterer wichtiger Aspekt, der bisher in vielen Studien festgestellt wurde, ist die große Abhängigkeit der Ergebnisse von den Versuchspersonen. Sowohl der Erfolg der Hörer variiert stark, als auch die Erkennbarkeit der Sprecher. Außerdem wird durch die Studien von Hollien et al. (1998) mit Schauspielern deutlich, dass einige Sprecher durchaus dazu in der Lage sind, ihre Trunkenheit nahezu perfekt zu maskieren, andere jedoch nicht.

Tab. 4.1: Übersicht über die bisherigen Perzeptionsexperimente

Studie	Sprecher	Hörer	Alkoholisierung in %	IR in %	DR in %	andere Ergebnisse
Sobell et al. (1982)	16m	7 Studenten	AAK 0,16 – 1,17	-	-	
Pisoni et al. (1985)	4m	21 Studenten	AAK 1,0 – 1,7	-	73,8	
	2m	6 Sprachwissenschaftler	AAK 1,5 & 1,6	-	82,4	
Martin & Yuchtman (1986)	8m	30 Studenten	AAK 1,0 – 1,9	-	61,5	
		14 Polizisten		-	64,7	
Klingholz et al. (1988)	11m	12 Sprachtherapeuten	BAK 0,67 - 1,59	54	61,1	
			BAK < 1,0	-	54,2	
			BAK > 1,0	-	82	
Künzel et al. (1992)	33m	30	AAK < 0,4 – > 2,0	51,1 – 95,8 Ø 66,8	-	Ergebnisse stark von der AAK abhängig
Hollien & Martin (1996)	6m, 2w	50 (ca. 10m, 40w)	AAK 1,2 – 1,3	-	-	Skala von 1 – 5 Ø Differenz zwischen nüchtern und alkoholisiert: 1,3 (n.s.)
				-	-	Einteilen in zwei (nicht benannte) Gruppen: Ø 66% der nüchternen Stimuli in einer Gruppe

w = weiblich, m = männlich, IR = Identifikationsrate, DR = Diskriminationsrate, - = nicht untersucht, • = nicht angegeben

Tab. 4.1: Übersicht über die bisherigen Perzeptionsexperimente: Fortsetzung

Studie	Sprecher	Hörer	Alkoholisierung in %	IR in %	DR in %	andere Ergebnisse
Hollien & Martin (1996)	6m, 2w	12	AAK 1,2 – 1,3	-	-	Einteilen in nüchtern/alkoholisiert: Ø 65% der alkoholisierten Stimuli richtig zugeordnet, Ø 60% der nüchternen
		17		-	Einteilen in nüchtern/alkoholisiert mit der Option ,no decision': Ø 68% der Stimuli in der nüchtern-Gruppe waren tatsächlich nüchtern	
Hollien et al. (1998)	12 (5w, 7m) Schauspieler	40 (33w, 7m)	1.: AAK 1,2 – 1,3 2.: AAK 0, simulierte starke Trunkenheit	-	88	Simulation als ,betrunkenener' erkannt

w = weiblich, m = männlich, IR = Identifikationsrate, DR = Diskriminationsrate, - = nicht untersucht, • = nicht angegeben

Tab. 4.1: Übersicht über die bisherigen Perzeptionsexperimente: Fortsetzung

Studie	Sprecher	Hörer	Alkoholisierung in %	IR in %	DR in %	andere Ergebnisse
Hollien et al. (1998)	12 (5w, 7m) Schauspieler	40 (33w, 7m)	1.: AAK 1,2 – 1,3 2.: AAK 1,2 – 1,3, maskierte Trunkenheit	–	61	starke Trunkenheit ohne Maskierung als ,betrunkenener‘ erkannt
		128	AAK 0,4 – 1,2	(82)	–	Skala von 1 – 5, Unterschiede zwischen drei Trunkenheitsgraden sign.
Hollien et al. (2009)	48	20 Kranken- hausärzte, 18 Studenten	AAK 1,2 – 1,3	–	–	Skala von 1 – 5, Unterschiede innerhalb der Hörergruppen sign., n.s. zwischen den Hörergruppen
		101	AAK 0,4 – 1,2	–	–	Skala von 1 – 5, Spontansprache sign. höhere Trunkenheitsbewer- tungen als gelesene Sprache

w = weiblich, m = männlich, IR = Identifikationsrate, DR = Diskriminationsrate, - = nicht untersucht, • = nicht angegeben

Tab. 4.1: Übersicht über die bisherigen Perzeptionsexperimente: Fortsetzung

Studie	Sprecher	Hörer	Alkoholisierung in ‰	IR in %	DR in %	andere Ergebnisse
Aulich (2010)	16 (8w, 8m)	28 (14w, 14m), 19 Phonetiker	BAK 0,5 – 1,42	–	Phonetiker: 70,3, andere: 73 Ø 71,7	starke Sprecherab- hängigkeit aber kein Zusammenhang mit BAK
	16 (8w, 8m) invers gefiltert	(16w, 3m)		–		
Ultes et al. (2011)	•	79 (31w, 48m)	BAK 0,2 – 1,5	38,3 – 70 Ø 55,8	–	bessere Ergebnisse für nüchtern und stark alkoholisiert, schlecht für schwach alkoholisiert

w = weiblich, m = männlich, IR = Identifikationsrate, DR = Diskriminationsrate, - = nicht untersucht, • = nicht angegeben

4.2 Diskriminationstest

In diesem Abschnitt wird ein Diskriminationstest vorgestellt, der mit 161 Sprechern des ALC und 72 Hörern durchgeführt wurde. In Kapitel 4.2.1 wird das Experimentdesign und die Extraktion der Stimuli aus dem ALC beschrieben, außerdem das Sprachmaterial erläutert und die Hörer vorgestellt. Hypothesen werden in Kapitel 4.2.2 aufgestellt und in Kapitel 4.2.3 die Ergebnisse beschrieben. Hierbei wird getestet, ob nicht-akustische Einflüsse auf die Antworten der Hörer vorliegen (z.B. die BAK der Sprecher oder ihre Trinkgewohnheiten) und Zusammenhänge mit den in Kapitel 3.2 vorgestellten F_0 -Parametern geprüft. In Kapitel 4.2.4 werden die Ergebnisse abschließend diskutiert.

4.2.1 Methode

4.2.1.1 Stimuli

In dem Diskriminationstest sollten alle Sprecher des ALC verwendet werden und verschiedene Stimuli paarweise präsentiert werden – jeweils eine Aufnahme des Sprechers in nüchternem Zustand und eine Aufnahme desselben Sprechers in alkoholisiertem Zustand. Für jeden Sprachstil wurden acht Äußerungen pro Sprecher ausgewählt und in der gelesenen und Kommandosprache automatisch geschnitten. In der Spontansprache wurden pro Sprecher vier Aufnahmeelemente ausgewählt, aus denen einmal am Anfang und einmal am Ende der Aufnahme kurze Äußerungsstücke ausgewählt und von Hand geschnitten wurden. Diese sollten im alkoholisierten und nüchternen Zustand möglichst gleich lang und inhaltlich ähnlich sein. Außerdem wurde darauf geachtet, dass in der Spontansprache so wenig Lachen, Versprecher oder inhaltliche Hinweise auf Alkoholisierung vorhanden waren wie möglich. Bei einer Sprecherin war es mit den oben genannten Kriterien nicht möglich, in der Spontansprache geeignete Stimuli herauszuschneiden, deshalb wurde sie für den Diskriminationstest ausgeschlossen.

- **Gelesene Sprache**

1. *Bitte so schnell wie möglich lesen:* Messwechsel, Wachsmaske, Wachsmaske, Messwechsel
2. *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.
3. *Bitte lesen Sie die Adresse:* Sportplatzweg 27, Marktgraitz
4. *Bitte lesen Sie die Adresse:* 18546 Sassnitz
5. *Bitte so schnell wie möglich lesen:* Die Köchin mit dem Tupfenkopftuch kocht Karpfen in dem Kupferkochtopf.
6. *Bitte Adresse lesen:* Schwester-Hermenegildis-Straße
7. *Bitte Adresse lesen:* Madapaka-Betegindis-Straße 77 B
8. *Bitte so schnell wie möglich lesen:* Ketzer Krächzer petzten jetzt kläglich, letztlich plötzlich leicht skeptisch.

- **Kommandosprache**

1. *Bitte Steuerbefehl lesen:* Temperatur 23°C
2. *Bitte stellen Sie sich vor, Sie wollten den Radiosender auf Speicherplatz FM3 einstellen. Geben Sie dem Auto den Befehl, dies zu tun.*
3. *Bitte Steuerbefehl lesen:* Frequenz 92,2 MHz
4. *Bitte stellen Sie sich vor, sie wollen ihre Sitzheizung auf Stufe 2 schalten. Geben Sie dem Auto den Befehl, dies zu tun.*
5. *Sie wollen den 9. Titel auf der 6. CD Ihres CD-Wechslers hören. Geben Sie Ihrem Auto den Befehl dazu.*
6. *Sie wollen Ihre Klimaanlage anschalten. Geben Sie Ihrem Auto den Befehl dazu.*

7. *Sie möchten zum Hilton Nürnberg und dafür Ihr Navigationssystem benutzen. Geben Sie Ihrem Auto den entsprechenden Befehl.*
8. *Bitte Steuerbefehl lesen: Autobahnen meiden*

- **Spontansprache**

1. *Bitte sprechen Sie mit dem Versuchsleiter: Unterhalten Sie sich über einen Ihrer Urlaube.*
2. *Erzählen Sie eine Geschichte zum Bild.*
3. *Sprechen Sie mit dem Versuchsleiter über das Bild.*
4. *Bitte beantworten Sie folgende Frage: Was war bisher das schönste Geschenk, das Sie bekommen haben und warum hat es Ihnen so gefallen?*

Weiterer Bestandteil des Perzeptionstests waren zusätzliche Aufnahmen von 20 Kontrollgruppensprechern, die ein weiteres Mal nüchtern aufgenommen wurden und hier dem Vergleich von zwei Aufnahmen eines Sprechers in nüchternem Zustand dienen. Insgesamt ergeben sich also bei 161 + 20 Sprechern und 24 Stimuluspaaren pro Sprecher eine Gesamtanzahl von 4344 Stimuluspaaren (3864 ohne Kontrollgruppe). Aufgrund dieser hohen Zahl an Stimuluspaaren können nicht alle Paare von allen Hörern beurteilt werden. Es wurde also ein Experimentdesign entwickelt, durch welches jeder Hörer jeden Sprecher genau *einmal* hört, also 181 Stimuluspaare zu beurteilen hat. Dafür wurden 24 verschiedene randomisierte Stimuluslisten erstellt, in denen jeder Sprecher genau einmal vorkommt und die Sprachstile der Stimuli ausgewogen sind. Jede dieser Stimuluslisten wurde von drei Hörern bearbeitet. Es existieren also pro Stimuluspaar *drei* unterschiedliche Hörerurteile. Insgesamt ergeben sich dadurch 13032 Hörerurteile (11592 ohne Kontrollgruppe).

Alter	Geschlecht	Bundesland Grundschule	Hörtesterfahrungen
20 – 36 Jahre Ø 24,4	36w, 36m	BY: 52; BW: 7; NW: 5; HE: 3; BE, BB, NI, SL, A: 1	nein: 68; einmal: 3; dreimal: 1

Tab. 4.2: Übersicht der Hörer – *w = weiblich, m = männlich; BY = Bayern, BW = Baden-Württemberg, NW = Nordrhein Westfalen, HE = Hessen, BE = Berlin, BB = Brandenburg, NI = Niedersachsen, SL = Saarland, A = Österreich*

4.2.1.2 Hörer

Die 72 freiwilligen Hörer wurden hauptsächlich durch Aushänge an der Universität und einen studentischen E-Mail-Verteiler rekrutiert. Aufgrund des universitären Umfelds waren ein Großteil der Hörer Studierende im Alter von 20 bis 36 (durchschnittlich 24,4) Jahren. Das Geschlechterverhältnis ist mit 36 weiblichen und 36 männlichen Versuchspersonen ausgeglichen. Vor Durchführung des Tests mussten alle Freiwilligen einen kurzen Fragebogen ausfüllen, eine Übersicht der damit erlangten Metadaten ist in Tabelle 4.2 aufgelistet. Die Mehrheit der Teilnehmer (72%) haben ihre Grundschulzeit in Bayern verbracht, die wenigsten (4 Personen) hatten bereits Erfahrungen mit Hörtests gesammelt. Ebenso wurde nach bekannten Hörschädigungen gefragt, wobei lediglich ein Proband einen diagnostizierten Tinnitus angab.

4.2.1.3 Durchführung

Die Testoberfläche wurde mit Matlab programmiert, ebenso die individuelle Randomisierung und die Stimuluslistenauswahl. Die von den Versuchspersonen zu beantwortende Frage lautete: „Bei welcher dieser Aufnahmen war der Sprecher alkoholisiert?“. Die Hörer durften jeden Stimulus bis zu fünf Mal anhören und mussten dann eine Entscheidung treffen, indem sie die entsprechende Antwort auswählten. Die Auswahlmöglichkeit ‚weiß nicht‘ bestand nicht, es handelt sich also um einen *2-alternative forced-choice test (2AFC)*. Es war nicht möglich, ohne eine Auswahl zum nächsten Stimuluspaar zu gelangen. Ebenso konnte keine Auswahl getroffen werden, ohne jeden Stimulus mindestens einmal zu hören. In diesen Fällen bekam die Versuchsperson eine entsprechende Fehlermeldung. Der Test dauerte ca. 45 Minuten und wurde mit 10

Euro pro Versuchsperson entlohnt. Jede Testperson führte den Test am selben Computer mit denselben Kopfhörern durch.

4.2.2 Hypothesen

Anhand der gesichteten und in Kapitel 4.1 vorgestellten Literatur, sowie der in Kapitel 3.2 vorgestellten Studie zur Grundfrequenz alkoholisierter Sprecher lassen sich bezüglich des Diskriminationstests folgende Hypothesen aufstellen:

1. Der Diskriminationstest mit 161 Sprechern des ALC wird eine geringere Diskriminationsrate zeigen, als der Test in Aulich (2010) mit 16 Sprechern (mit einer BAK > 0,5‰) des ALC.
2. Die Alkoholisierung von weiblichen Sprechern wird wie in Aulich (2010) häufiger korrekt erkennbar sein.
3. Aufgrund des höheren kognitiven Anspruchs beim Lesen können die Hörer in den Beispielen der gelesenen Sprache den in alkoholisiertem Zustand produzierten Stimulus häufiger korrekt erkennen, als in Spontan- und Kommandosprache.
4. Bei Sprechern mit höherer BAK lässt sich der unter Alkoholeinfluss produzierte Stimulus von den Hörern häufiger korrekt erkennen.
5. Bei starken Trinkern wird die Alkoholisierung weniger oft korrekt erkannt, als bei mittleren und leichten Trinkern.
6. Bei Sprechern, die im Mittel einen größeren Unterschied in ihrer mittleren Grundfrequenz zwischen alkoholisiert und nüchtern produzierten Stimuli aufweisen, wird der alkoholisiert produzierte Stimulus häufiger korrekt erkannt.
7. Bei Sprechern, die im Mittel einen größeren Unterschied in ihrem Tonumfang zwischen den beiden Stimuli eines Paares aufweisen, wird der alkoholisiert produzierte Stimulus häufiger korrekt erkannt.

8. Hörer zeigen in der Auswahl des unter Alkoholeinfluss produzierten Stimulus eine Präferenz für den Stimulus mit der höheren mittleren Grundfrequenz.
9. Hörer zeigen in der Auswahl des unter Alkoholeinfluss produzierten Stimulus eine Präferenz für den Stimulus mit dem größeren Tonumfang.

4.2.3 Ergebnisse

4.2.3.1 Erkennungsleistung der Hörer und Erkennungsraten der Sprecher

Zur Auswertung des Perzeptionstests werden im Wesentlichen zwei Werte berechnet - die Erkennungsleistung der Hörer und die Erkennungsrate der Sprecher, die mit der Diskriminationsrate gleichzusetzen ist. Die Erkennungsleistung der Hörer wird für jeden Hörer einzeln berechnet und beschreibt, in wieviel Prozent der von ihm gehörten Stimuluspaare er den richtigen Stimulus als den unter Alkoholeinfluss produzierten ausgewählt hat. Die Erkennungsraten der Sprecher hingegen werden für jeden Sprecher einzeln berechnet und sagen aus, in wieviel Prozent der Fälle bei den Stimuluspaaren eines Sprechers von den Hörern der richtige Stimulus ausgewählt wurde.

Im Durchschnitt führte der Perzeptionstest zu einer Erkennungsleistung seitens der Hörer von 63,1%. Ein χ^2 -Test zeigt, dass das Ergebnis mit insgesamt 7418 von 11592 möglichen richtigen Antworten signifikant über dem Zufallsniveau von 50% liegt ($\chi^2[1] = 798,81, p < 0,001$). Der schlechteste Hörer erkannte lediglich 52,8% der Stimuli in denen der Sprecher alkoholisiert war, im besten Fall wurden 76,4% der unter Alkoholeinfluss produzierten Stimuli auch als solche zugeordnet. Abb. 4.1a zeigt die Verteilung der Erkennungsleistung der Hörer. Geschlechtsspezifische Unterschiede waren weder in der generellen Erkennungsleistung zu erkennen, noch können Frauen die Sprache von Männern besser beurteilen oder umgekehrt und auch die gleichgeschlechtliche Erkennbarkeit ist nicht besser als die gegengeschlechtliche. Eine Zusammenfassung der geschlechtsspezifischen Ergebnisse zeigt Tabelle 4.3.

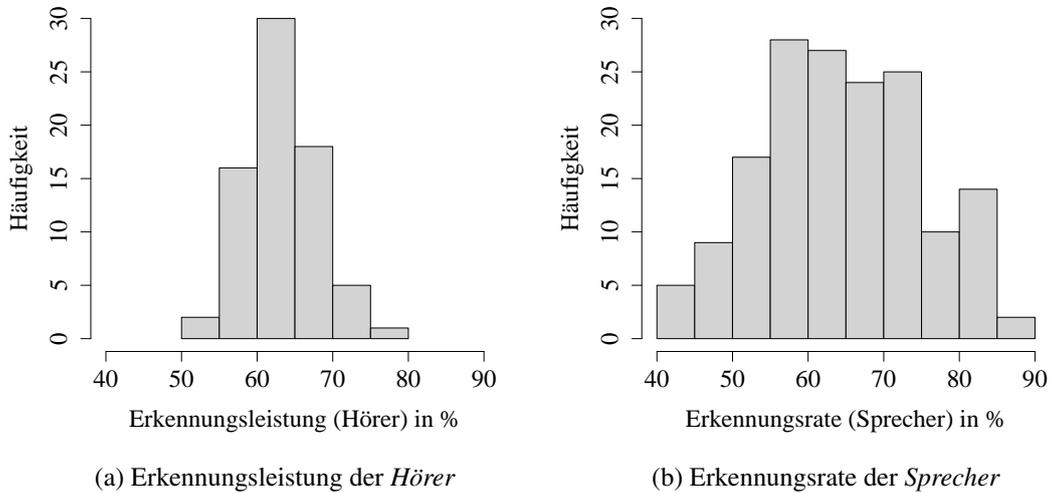


Abb. 4.1: Ergebnisse des Diskriminationstests, 161 Sprecher, 72 Hörer

Die Verteilung der unterschiedlichen Erkennungsraten der einzelnen Sprecher ist in Abb. 4.1b zu sehen. Auch hier sind starke Variationen in der Erkennbarkeit sichtbar. Die Stimuli des Sprechers mit der schlechtesten Erkennungsrate wurden nur zu 40,3% richtig zugeordnet – er wurde also sogar überzufällig entgegengesetzt beurteilt. Der am deutlichsten als alkoholisiert wahrnehmbare Sprecher erreichte allerdings eine Erkennungsrate von 87,5%, hier konnte also ein Großteil der Hörer bei den meisten Stimuluspaaren richtig entscheiden, bei welchem der beiden Stimuli es sich um den unter Alkoholeinfluss produzierten handelte. Im Durchschnitt wurden von jedem Sprecher 63,5% der Stimuluspaare korrekt eingeschätzt. Bei der Betrachtung der beiden Verteilungen in den Abb. 4.1a und 4.1b fällt auf, dass die Ausdehnung der Erkennungsraten der Sprecher um einiges größer ist, als die der Erkennungsleistung der Hörer. Dies ist insofern nicht verwunderlich, als alle Hörer unter den gleichen Voraussetzungen getestet wurden. Bei den Sprechern allerdings bedingt der unterschiedliche Grad an Alkoholisierung und zusätzlich die individuelle Reaktion darauf auch eine stärkere Variationsbreite an Voraussetzungen.

Die Ergebnisse der Kontrollgruppe (in der die Sprecher zu beiden Aufnahmezeitpunkten nüchtern waren) zeigen, dass der oben beobachtete Effekt nicht von eventuellen versteckten Faktoren aufgrund verschiedener Aufnahmesituationen beeinflusst ist.

MW ← M	MW ← W	M ← MW	W ← MW
63%	63,3%	63,7%	62,5%
M ← M	M ← W	W ← M	W ← W
63,4%	64%	62,5%	62,5%

Tab. 4.3: Ergebnisse nach Geschlecht – *M/W = männliche/weibliche Sprecher bzw. Hörer, ← = beurteilt von*

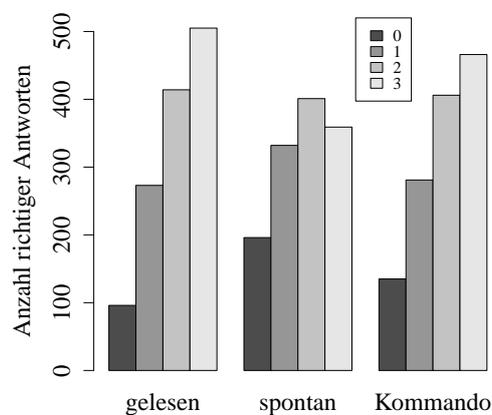
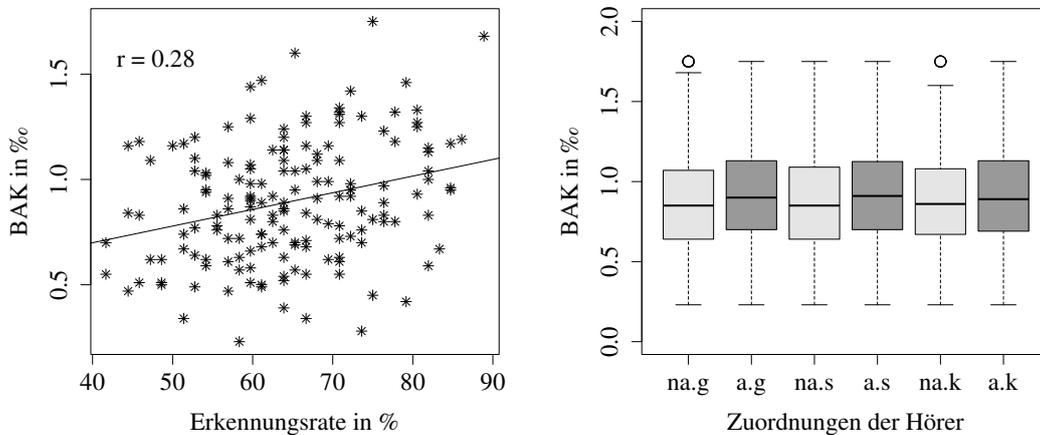


Abb. 4.2: Anzahl richtiger Antworten in den unterschiedlichen Sprachstilen

Bei den Stimuluspaaren der Kontrollgruppe wählten die Hörer zufällig zwischen den verschiedenen Aufnahmesituationen (im Durchschnitt war die Verteilung für die zwei Aufnahmesituationen 49,2% zu 51,8%).

Bei den drei unterschiedlichen Sprachstilen zeigt sich, dass gelesene Sprache am besten erkannt wurde (67,7%) und Spontansprache mit einer Erkennungsrate von 57,2% am schlechtesten. Die Kommandosprache liegt mit 64,5% dazwischen. Abb. 4.2 zeigt die Unterschiede zwischen den Sprachstilen. Hierfür wurde für jedes Stimuluspaar die Anzahl richtiger Antworten gezählt. Wie in Kapitel 4.2.1.1 beschrieben, wurde jedes Stimuluspaar von drei verschiedenen Hörern gehört, konnte also auch maximal dreimal richtig beurteilt werden. Auch hier zeigt sich, dass in der Spontansprache weniger Stimuluspaare von allen drei Hörern korrekt beurteilt wurden als in den beiden anderen Sprachstilen.



(a) Zusammenhang von Erkennungsrate und BAK der Sprecher
 (b) BAK der Sprecher in den Zuordnungen der Hörer – a = alkoholisiert, na = nüchtern

Abb. 4.3: Einfluss der Blutalkoholkonzentration der Sprecher

4.2.3.2 Nicht-akustische Einflussfaktoren auf die Hörerentscheidung

Eine lineare Regression zeigt einen signifikanten Zusammenhang zwischen der Erkennungsrate eines Sprechers und seiner BAK ($F(1, 159) = 14,02, p < 0,001$), das Streudiagramm hierzu ist in Abb. 4.3a zu sehen. Der Regressionskoeffizient von $r = 0,28$ beschreibt zwar einen relativ schwachen Zusammenhang, dennoch ist erkennbar, dass bei Sprechern mit einer höheren BAK im Allgemeinen der unter Alkoholeinfluss produzierte Stimulus von den Hörern häufiger korrekt identifiziert wurde als bei Sprechern mit niedrigerer BAK. Eine *MEM*-Analyse mit den Antworten der Hörer als abhängige Variable, Sprechern und Hörern als *random*-Faktoren und Sprechergeschlecht, Sprachstil, Trinkgewohnheiten, BAK und Hörergeschlecht als *fixed*-Faktoren zeigt, dass sowohl der Sprachstil des zu beurteilenden Stimuluspaares als auch die BAK des Sprechers einen signifikanten Einfluss auf die Antwort der Hörer haben. Ein *post-hoc*-Test zeigt, dass sich die Ergebnisse für alle Sprachstile signifikant voneinander unterscheiden, der Unterschied zwischen gelesener Sprache und Kommandosprache allerdings nicht so stark ausgeprägt ist. Die Ergebnisse des *post-hoc*-Tests sind in Tabelle 4.5 zusammengefasst. In Abb. 4.3b sind die BAK der Sprecher dargestellt, aufgeteilt nach Zuordnungen der Hörer und Sprachstil. *Nüchtern* bedeutet in diesem Fall, dass der

Faktor	Ergebnis der MEM-Analyse
Sprechergeschlecht	$\chi^2[1] = 0,44, \quad p > 0,1 \text{ n.s.}$
Sprachstil	$\chi^2[2] = 98,95, \quad p < 0,001 \text{ ***}$
Trinkgewohnheiten	$\chi^2[2] = 1,8, \quad p > 0,1 \text{ n.s.}$
BAK	$\chi^2[1] = 14,04, \quad p < 0,001 \text{ ***}$
Hörergeschlecht	$\chi^2[1] = 0,07, \quad p > 0,1 \text{ n.s.}$

Tab. 4.4: Nicht-akustische Einflussfaktoren auf die Antwort der Hörer im Diskriminationstest – *n.s.* = nicht signifikant, *.* = signifikanter Trend, *** = signifikant, **** = hoch signifikant, ***** = höchst signifikant

Sprachstile	Ergebnis des post-hoc-Tests
spontan - gelesen	$z = -9,74, \quad p < 0,001 \text{ ***}$
Kommando - gelesen	$z = -3,11, \quad p < 0,01 \text{ **}$
Kommando - spontan	$z = 6,67, \quad p < 0,001 \text{ ***}$

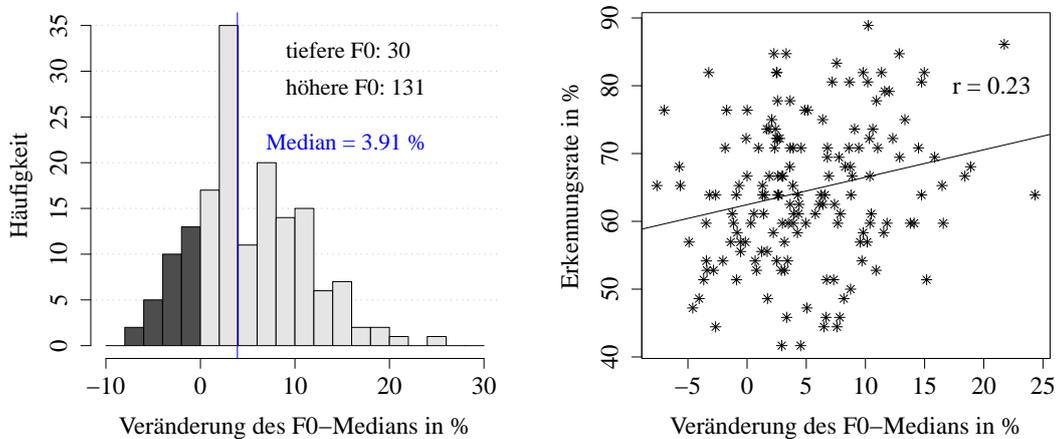
Tab. 4.5: Ergebnisse des post-hoc-Tests für die unterschiedlichen Sprachstile – *n.s.* = nicht signifikant, *.* = signifikanter Trend, *** = signifikant, **** = hoch signifikant, ***** = höchst signifikant

unter Alkoholeinfluss produzierte Stimulus nicht korrekt identifiziert wurde (sondern fälschlicherweise für den in nüchternem Zustand produzierten Stimulus gehalten wurde), *alkoholisiert* stellt eine korrekte Zuordnung dar.

Wie in Tabelle 4.3 bereits zu erkennen ist, beeinflussen weder das Sprecher- noch das Hörergeschlecht die Beurteilung eines Stimuluspaares, und auch die Trinkgewohnheiten des Sprechers haben keine Auswirkung auf die Erkennbarkeit seiner Alkoholisierung. Die Ergebnisse der MEM-Analyse sind in Tabelle 4.6 zusammengefasst. Signifikante Interaktionen zwischen den einzelnen Faktoren konnten nicht nachgewiesen werden.

4.2.3.3 Einfluss der F₀-Parameter auf die Hörerentscheidung

In Kapitel 3.2 wurde in der akustischen Studie gezeigt, dass bei 79,1% der Sprecher in ihrer Sprache im alkoholisierten Zustand eine höhere Grundfrequenz messbar ist als in nüchternem Zustand. Um zu testen, ob demnach F₀ auch als perzeptives Merkmal zur Unterscheidung von Sprache alkoholisierter und nüchterner Sprecher eine Rolle

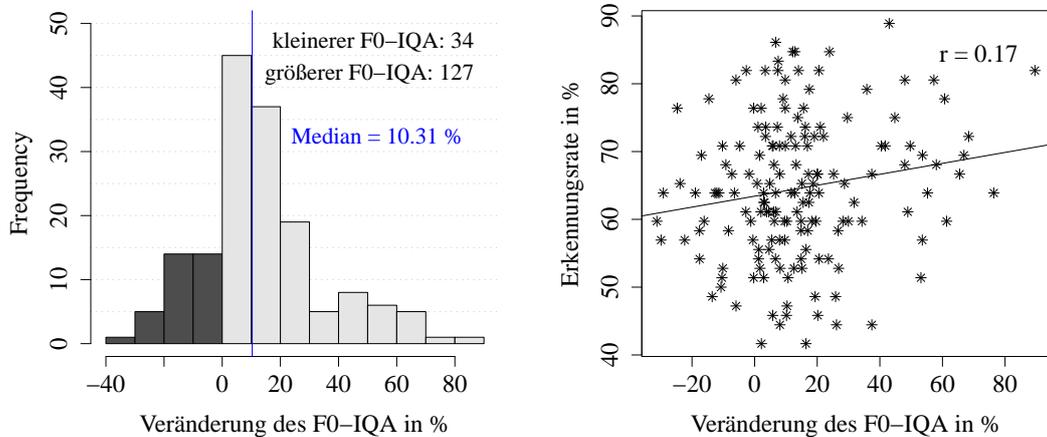


(a) Veränderung des F_0 -Medians der Sprecher in den Stimuluspaaren
 (b) Zusammenhang von Erkennungsrate und Veränderung des F_0 -Medians

Abb. 4.4: Veränderung des F_0 -Medians

spielt, wurde ebenso wie in Kapitel 3.2.3 für jede Versuchsperson die relative Veränderung der Grundfrequenz von seiner nüchtern produzierten zur alkoholisiert produzierten Sprache berechnet, diesmal auf Basis der im Perzeptionstest verwendeten Stimuluspaare. Abb. 4.4a zeigt deren Verteilung. Auch hier zeigen 131 von 161 Versuchspersonen (also 81,4%) eine höhere Grundfrequenz in ihrer Sprache in alkoholisiertem Zustand mit einem durchschnittlichen Anstieg von 3,91%. Eine lineare Regression dieser prozentualen Veränderungen und der jeweiligen Erkennungsrate (Abb. 4.4b) zeigt, dass bei Sprechern, die eine höhere prozentuale Veränderung aufweisen, der in alkoholisiertem Zustand gesprochene Stimulus von den Hörern häufiger richtig erkannt wurde. Obwohl der Korrelationskoeffizient mit $r = 0,23$ relativ niedrig ist, zeigt das lineare Regressionsmodell ein signifikantes Ergebnis bei einer Irrtumswahrscheinlichkeit von 1% ($F(1, 159) = 8,73, p < 0,01$).

Eine *MEM*-Analyse mit den Antworten der Hörer als abhängige Variable, Sprechern und Hörern als *random*-Faktoren und der Veränderung des F_0 -Medians und der Veränderung des F_0 -IQA als unabhängige Faktoren, zeigt, dass die Antwort der Hörer signifikant von der Veränderung des F_0 -Medians des Sprechers beeinflusst wird ($\chi^2[1] = 37,95, p < 0,001$). Abb. 4.6a zeigt die F_0 -Median-Veränderungen



(a) Veränderung des F_0 -IQA der Sprecher in den Stimuluspaaren (b) Zusammenhang von Erkennungsrate und Veränderung des F_0 -IQA

Abb. 4.5: Veränderung des F_0 -IQA

getrennt nach Zuordnungen der Hörer und Sprachstil. Im Mittel sind die F_0 -Median-Veränderungen in den unter Alkoholeinfluss produzierten Stimuli, die auch als solche erkannt wurden, in allen Sprachstilen größer als die der fälschlicherweise für in nüchternem Zustand produziert gehaltenen Stimuli. Auch hier sind keine signifikanten Interaktionen zu anderen Faktoren nachweisbar. Die Veränderung des F_0 -IQA hat keinen Einfluss auf die Antwort der Hörer, die *MEM*-Analyse zeigt hier kein signifikantes Ergebnis ($\chi^2[1] = 0,08, p > 0,1$). Auch Abb. 4.6b lässt keinen Unterschied in den F_0 -IQA-Veränderungen zwischen den richtig und falsch beurteilten Stimuluspaaren erkennen. Die Verteilung der Veränderung der F_0 -IQA in den verwendeten Stimuli ist in Abb. 4.5a dargestellt. Auch das lineare Regressionsmodell zeigt nur einen sehr schwachen Zusammenhang zwischen der Veränderung des F_0 -IQA und der Erkennungsrate eines Sprechers ($F(1, 159) = 4,6, p < 0,05$) mit einem Regressionskoeffizienten von $r = 0,17$. Das Streudiagramm hierzu ist in Abb. 4.5b zu sehen. Zur besseren Übersicht sind die Ergebnisse der *MEM*-Analyse in Tabelle 4.6 und die Ergebnisse der linearen Regression in Tabelle 4.7 nochmals zusammengefasst.

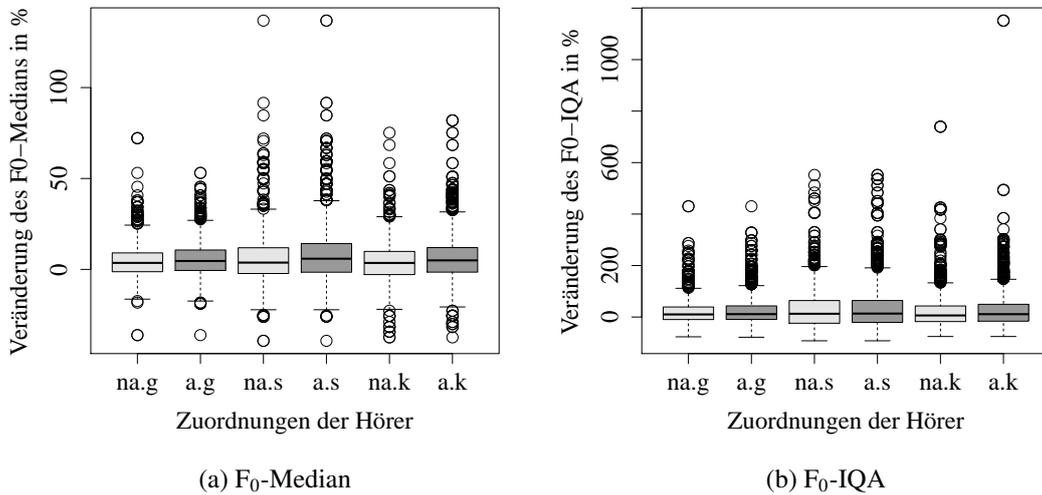


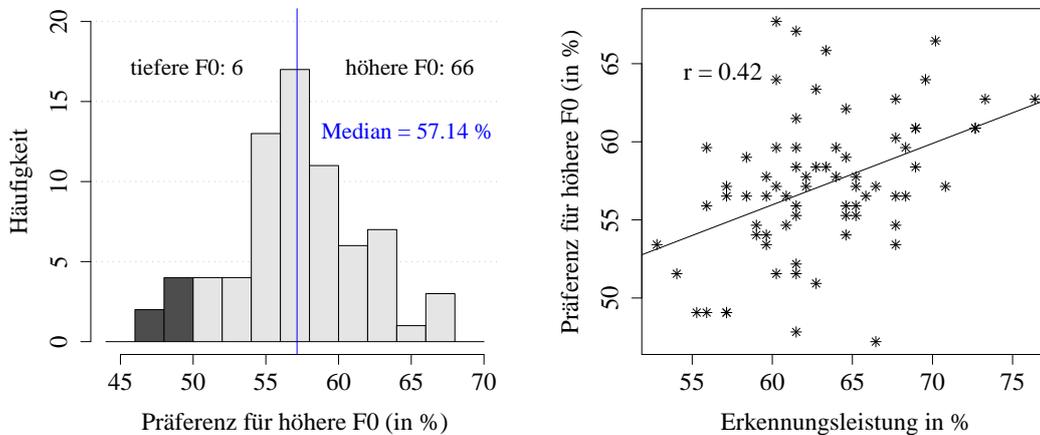
Abb. 4.6: Veränderungen der F₀-Parameter der Sprecher in den Zuordnungen der Hörer – a = alkoholisiert, na = nüchtern, g = gelesene Sprache, s = Spontansprache, k = Kommandosprache

Faktor	Ergebnis der MEM-Analyse
Veränderung des F ₀ -Medians	$\chi^2[1] = 37,95, p < 0,001$ ***
Veränderung des F ₀ -IQA	$\chi^2[1] = 0,08, p > 0,1$ n.s.

Tab. 4.6: Einflussfaktoren (F₀) auf die Antwort der Hörer im Diskriminationstest – n.s. = nicht signifikant, . = signifikanter Trend, * = signifikant, ** = hoch signifikant, *** = höchst signifikant

	Erkennungsrate
Veränderung des F ₀ -Medians	$F(1, 159) = 8,73, p < 0,01$ **
Veränderung des F ₀ -IQA	$F(1, 159) = 4,6, p < 0,05$ *

Tab. 4.7: Ergebnisse der linearen Regression der Veränderung des F₀-IQA und der Erkennungsrate

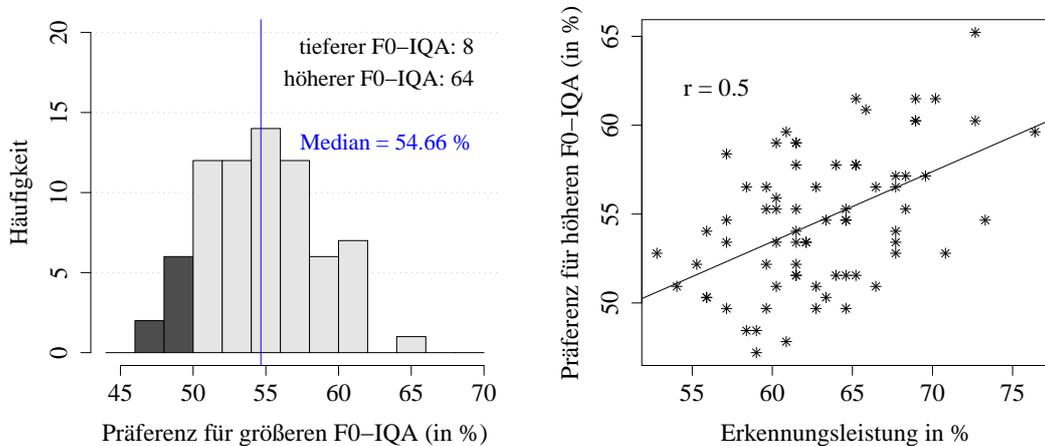


(a) Verteilung der Präferenz der 72 Hörer für den Stimulus mit höherem F_0 -Median

(b) Zusammenhang von Erkennungsleistung und Präferenz der 72 Hörer für den Stimulus mit höherem F_0 -Median

Abb. 4.7: Präferenz für den Stimulus mit höherem F_0 -Median

Um beurteilen zu können, ob die Grundfrequenz ein Entscheidungskriterium in der Perzeption darstellt, wurde zusätzlich für jeden Hörer einzeln eine generelle Präferenz berechnet, denjenigen Stimulus mit der höheren Grundfrequenz als den unter Alkoholeinfluss produzierten auszuwählen. Eine Verteilung dieser Präferenz über die einzelnen Versuchspersonen ist in Abb. 4.7a dargestellt. Es ist sehr deutlich zu sehen, dass der Großteil der Hörer, nämlich 91,7%, eine Präferenz für den Stimulus, in dem die höhere Grundfrequenz gemessen wurde, aufweisen – sich also in mehr als 50% der Fälle für den Stimulus mit höherer F_0 als den unter Alkoholeinfluss produzierten entschieden haben. Die Versuchsperson mit der größten Präferenz für den Stimulus mit der höheren Grundfrequenz wählte in 68% der Stimuluspaare den Stimulus mit höherer F_0 aus. Der Median dieser Verteilung liegt bei 57,1%. Auch die Präferenz für den Stimulus mit größerem F_0 -IQA liegt deutlich über 50%. Die Verteilung hierzu ist in Abb. 4.8a zu sehen. Insgesamt präferieren hier 88,8% der Hörer in mehr als der Hälfte der Stimuluspaare den Stimulus mit größerem Tonumfang. Der Median liegt bei 54,7% und somit etwas unter dem Wert bezüglich der mittleren Grundfrequenz.



(a) Verteilung der Präferenz der 72 Hörer für den Stimulus mit höherem F₀-IQA

(b) Zusammenhang von Erkennungsleistung und Präferenz der 72 Hörer für den Stimulus mit höherem F₀-IQA

Abb. 4.8: Präferenz für den Stimulus mit höherem F₀-IQA

Abb. 4.7b zeigt das Streudiagramm dieser Präferenzen mit den Erkennungsleistungen der jeweiligen Hörer. Der Regressionskoeffizient liegt bei $r = 0,42$ und beschreibt einen signifikanten Zusammenhang ($F(1, 70) = 15,13, p < 0,001$). Der Zusammenhang mit der Erkennungsrate der Sprecher ist allerdings mit $r = 0,5$ etwas größer bei der Veränderung des F₀-IQA ($F(1, 70) = 23, p < 0,001$). Abb. 4.8b zeigt diesen Zusammenhang. Diejenigen Hörer, die eine hohe Präferenz für den Stimulus mit der höheren Grundfrequenz aufweisen haben auch mehr Erfolg, ebenso die Hörer, die den Stimulus mit dem größeren Tonumfang präferieren. Ein signifikanter Zusammenhang zwischen diesen beiden Präferenzen konnte ebenfalls in einem linearen Regressionsmodell nachgewiesen werden ($F(1, 70) = 22,34, p < 0,001, r = 0,49$).

4.2.4 Zwischendiskussion

Wie in Kapitel 4.1 bereits beschrieben wurde, erreichten 47 Hörer bei einem früheren Perzeptionstest mit 16 Sprechern des ALC (Aulich 2010) eine relativ hohe Erkennungsleistung von 71,65%. Ein solches Ergebnis kann in diesem Test mit 161 Sprechern des ALC nicht repliziert werden, dennoch liegt die Erkennungsleistung der Hö-

rer in diesem Experiment mit 63,1% immernoch über dem Zufallsniveau. Hypothese 1 lässt sich also hiermit bestätigen.

Ein wichtiger Unterschied in den beiden Experimenten ist neben der geringeren Anzahl der Sprecher in der Studie von Aulich auch der Ausschluss von Sprechern mit einer geringeren BAK als 0,5‰. Der Ausschluss der Sprecher mit einer BAK von 0,5‰ oder weniger in dem Diskriminationstest dieser Arbeit führt allerdings für die verbleibenden 148 Sprecher zu einem nicht wesentlichen Anstieg der Erkennungsleistung der Hörer von 1,7% auf 64,8% und liegt damit immernoch unter dem Ergebnis von Aulich mit 16 Sprechern. Dennoch konnten die Hörer insgesamt bei Sprechern mit einer höheren BAK häufiger den unter Alkoholeinfluss produzierten Stimulus erkennen als bei Sprechern mit niedriger BAK, Hypothese 4 wird somit ebenfalls bestätigt. Pisoni et al. (1985) erreichten in ihren Experimenten (vgl. Kapitel 4.1) sogar noch höhere Diskriminationsraten von 73,8% bzw. 82,4%, aber auch hier ist das gute Ergebnis der Hörer vermutlich mit der geringen Anzahl an Sprechern und deren hohen AAK bzw. BAK zu erklären, die in beiden Experimenten bei allen Sprechern über 1,0‰ lag. Im zweiten Experiment wurden sogar nur stark alkoholisierte Sprecher mit einer AAK von über 1,5‰ beurteilt. Klingholz et al. (1988) erzielten in ihrem Experiment für Sprecher über 1,0‰ sogar eine Diskriminationsrate von 82%, für Sprecher mit geringerer BAK lag diese allerdings nur bei 54,2% (61,1% insgesamt). In dem in dieser Arbeit vorgestellten Experiment erreichten 54 Sprecher eine BAK über 1,0‰. Gesondert betrachtet liegt die Diskriminationsrate hier allerdings auch nur bei 66,7% – ein Ausschluss der weniger stark alkoholisierten Sprecher führt also in diesem Fall zu einem Anstieg der Diskriminationsrate von 3,6%. Auch wenn sich sehr gute Ergebnisse in vorherigen Studien mit der Verwendung von Sprechern mit einer hohen AAK bzw. BAK zum Teil erklären lassen, können vergleichbare Ergebnisse in dem hier präsentierten Experiment auch nach Ausschluss von Sprechern mit niedriger BAK nicht erreicht werden. In keiner der vorherigen Studien, die sehr hohe Diskriminationsraten hervorbrachten, wurde allerdings eine mit dem in dieser Arbeit vorgestellten Experiment vergleichbaren Anzahl an Sprechern oder Hörern verwendet.

Ein weiterer Unterschied der vorangegangenen Studien ist das zugrundeliegende Sprachmaterial. In den meisten Studien wurde den Hörern gelesenes Material der Sprecher präsentiert. Ein Vergleich der Erkennbarkeit in den unterschiedlichen Sprachstilen in dem hier vorgestellten Experiment zeigte, dass Hypothese 3 bestätigt wird und die unter Alkoholeinfluss produzierten Stimuli eines Sprechers in der gelesenen Sprache mit 67,7% am besten erkannt werden konnten, am schlechtesten die Stimuli der Spontansprache (57,2%). Betrachtet man in diesem Experiment nur die Stimuli der gelesenen Sprache und nur Sprecher mit einer BAK von mehr als 0,5‰ bzw. mehr als 1,0‰, steigt die Erkennungsleistung der Hörer um 4,8% bzw. 7,9% auf 67,9% bzw. 71%. In Tabelle 4.8 sind die Ergebnisse dieses Experiments nach Ausschluss von Sprechern mit niedriger BAK nochmals zusammengefasst.

Zum einen könnte die bessere Erkennbarkeit der Alkoholisierung in der gelesenen Sprache an den höheren kognitiven Anforderungen, die an einen Sprecher beim Vorlesen gestellt werden, liegen, die dazu führen könnten, dass sich Einflüsse von Alkohol auf die Sprache und Stimme deutlicher zeigen, als zum Beispiel in der Spontansprache. Die Sprecher befinden sich zum Aufnahmezeitpunkt zusätzlich in einer Art Testsituation, wodurch einige Sprecher versuchen könnten, ihre Alkoholisierung möglichst zu verdecken. Wie in Kapitel 3.1 bereits beschrieben wurde, testeten Hollien et al. (1998) diese Fähigkeit bei geübten Schauspielern und konnten feststellen, dass sie sehr stark vom Sprecher abhängig war, der beste Sprecher aber sogar in 95% der Fälle seine Trunkenheit maskieren konnte. Ob diese Fähigkeit mit dem Beruf des Schauspielers zusammenhängt, bleibt allerdings dahingestellt. Sollte diese Testsituation aber bei einigen Sprechern eine Verdeckungsintention hervorrufen, könnte das beim kognitiv anstrengenderen Lesen eventuell schlechter funktionieren. Erschwerend kommt in diesem Fall hinzu, dass die gelesene Sprache hier aus Zungenbrechern oder unbekanntem Adressaten bestand. In einem weiteren Test von Hollien et al. (2009) wurde hingegen in spontansprachlichen Äußerungen die Alkoholisierung der Sprecher von den Hörern besser eingeschätzt, allerdings argumentieren die Autoren hier damit, dass der vorgelesene Text mehrfach geübt wurde und somit als bekannt gelten kann, in diesem Fall

Anzahl Sprecher	BAK > 0,5‰	
	alle Sprachstile	gelesene Sprache
148	64,6% (+1,7)	67,9% (+4,8)
	BAK > 1,0‰	
	alle Sprachstile	gelesene Sprache
64	66,7% (+3,6)	71% (+7,9)

Tab. 4.8: Diskriminationsraten von Teilanalysen zur Vergleichbarkeit mit vorherigen Studien

also eine einfachere Aufgabe für die Sprecher darstellt, als eine spontane Antwort auf unbekannte Fragen zu geben.

Ein weiterer Aspekt, der auch im Hinblick auf die Zungenbrecher nicht außer Acht gelassen werden darf, ist der Einfluss von Lachen und Versprechern auf die Antwort der Hörer. Barfüßer & Schiel (2010) zeigten bei 128 Sprechern des ALC, dass sowohl die Anzahl der Versprecher, als auch die der Häsitationen und Pausen in alkoholisiertem Zustand steigt. In der Spontansprache wurden die Stimuli auf Versprecher und Lachen kontrolliert und bewusst vermieden. In der gelesenen und der Kommandosprache wurden die Stimuli für den Test allerdings automatisch geschnitten und enthalten somit gegebenenfalls auch Lachen und Versprecher anhand derer die Alkoholisierung eines Sprechers eventuell leichter erkannt werden könnte.

Neben dem Einfluss der BAK und des Sprachstils auf die Antwort der Hörer konnten keine weiteren nicht-akustischen Einflussfaktoren ausgemacht werden. Weder das Sprecher- noch das Hörergeschlecht zeigt einen Effekt auf die Beurteilungen der Hörer, und auch die Trinkgewohnheiten eines Sprechers verändern nicht seine Einschätzbarkeit. Hypothese 2, die besagt, dass die Alkoholisierung von weiblichen Sprechern leichter erkannt werden kann, muss demnach verworfen werden. Aulich (2010) konnte in ihrem Experiment einen solchen Unterschied zwar nachweisen, jedoch tritt er in dem hier präsentierten Experiment mit einer größeren Anzahl an Sprechern nicht mehr auf.

Hypothese 5, die besagt, dass die Alkoholisierung in der Sprache starker Trinker für die Hörer schwieriger zu erkennen ist, muss auch verworfen werden. Die Alkoholisierung eines Sprechers, der z.B. angab, höchstens einmal in der Woche maximal

ein bis zwei Einheiten Alkohol zu trinken (vgl. Kapitel 2), war also nicht leichter zu identifizieren als die eines Sprechers, der beispielsweise angab, mehrfach in der Woche mindestens zwei Einheiten Alkohol zu konsumieren.

Bei den beiden akustischen Parametern ‚mittlere Tonhöhe‘ und ‚Tonumfang‘ zeigte sich, dass die mittlere Tonhöhe einen signifikanten Einfluss auf die Erkennbarkeit der Alkoholisierung eines Sprechers hat: bei Sprechern, die im Mittel einen größeren Unterschied in der Tonhöhe zwischen den beiden Stimuli eines Paares aufweisen, konnte häufiger der in alkoholisiertem Zustand produzierte Stimulus von den Hörern identifiziert werden. Bei Stimuluspaaren mit großem Unterschied im Tonumfang zwischen beiden Stimuli konnte allerdings kein signifikanter Effekt auf die Antwort der Hörer nachgewiesen werden, also auch kein Zusammenhang mit der Erkennbarkeit der Alkoholisierung eines Sprechers, obwohl in Kapitel 3.2.3.4 ein Zusammenhang zwischen den beiden Parametern F_0 -Median und F_0 -IQA bei denselben Sprechern nachgewiesen werden konnte. Hypothese 6 lässt sich also bestätigen, Hypothese 7 jedoch nicht.

Hypothese 8 besagt, dass die Hörer eine Präferenz für den Stimulus zeigen, in dem die mittlere F_0 höher ist, und Hypothese 9 dass sie ebenfalls eine Präferenz für den Stimulus mit dem größeren Tonumfang aufweisen. Diese Hypothesen lassen sich beide bestätigen. Für beide Parameter bevorzugt ein Großteil der Hörer (91,7% bzw. 88,8%) in mehr als der Hälfte der Paare den Stimulus mit höherem F_0 -Median bzw. größerem F_0 -IQA. Zur besseren Übersicht sind die Hypothesentests nochmals in Tabelle 4.9 zusammengefasst.

Ebenfalls getestet wurde, ob diese generelle Präferenz auch zu mehr Erfolg führt. Betrachtet man die Verteilung der F_0 -Median-Veränderungen der Sprecher in den verwendeten Stimuli, fällt auf, dass bei mehr als 80% der Sprecher in den Stimuluspaaren die Grundfrequenz im Mittel in dem Stimulus, der in alkoholisiertem Zustand produziert wurde, höher ist, als in dem in nüchternem Zustand produzierten Stimulus. Es konnte zwar ein Zusammenhang zwischen der generellen Präferenz eines Hörers für den Stimulus mit höherer F_0 und seiner Erkennungsleistung gezeigt werden, welcher jedoch bei der großen Anzahl der Stimuluspaare, bei denen eine höhere F_0 tatsäch-

lich dem unter Alkoholeinfluss produzierten Stimulus zuzuordnen ist, nicht verwunderlich ist. Sowohl eine Präferenz für den Stimulus mit höherem F_0 -Median, als auch eine Präferenz für den Stimulus mit dem größeren F_0 -IQA, führt natürlicherweise zu mehr Erfolg seitens der Hörer. Der Zusammenhang mit einem Regressionskoeffizienten von $r = 0,42$ zwischen Präferenz für den höheren Stimulus und der persönlichen Erkennungsleistung eines Hörers erscheint aber in Anbetracht der Tatsache, dass es tatsächlich in mehr als 80% der Fälle zum Erfolg führen würde, den Stimulus mit der höheren Grundfrequenz als den in alkoholisiertem Zustand produzierten auszuwählen, als relativ gering. Die Grundfrequenz scheint also hier für den Hörer eine Rolle in der Entscheidung zu spielen, die Hörer scheinen sich jedoch nicht bewusst zu sein, wie gut die Grundfrequenz als Entscheidungskriterium wirklich funktioniert. Ein Hörer, der in jedem Fall den Stimulus mit höherer Grundfrequenz auswählen würde, müsste demnach eine Erkennungsleistung von ca. 80% erreichen.

Die Ergebnisse der Kontrollgruppe, in der die Hörer bei zwei nüchtern artikulierten Stimuli keine der beiden Aufnahmezeitpunkte bevorzugt als alkoholisierten Zustand des Sprechers wahrnahmen (Präferenz für einen Aufnahmezeitpunkt: 49,2%), zeigen, dass die Entscheidungen der Hörer im gesamten Experiment nicht auf eventuelle versteckte Einflussfaktoren der verschiedenen Aufnahmesituationen, sondern tatsächlich auf den Faktor Alkoholisierung zurückzuführen sind.

Insgesamt zeigt sich aber deutlich, dass die Ergebnisse des Perzeptionstests sowohl sprecher- als auch hörerabhängig sind, da sie für einzelne Sprecher bzw. Hörer zum Teil stark variieren.

Hypothese	Angenommen?
1. Der Diskriminationstest mit 161 Sprechern des ALC wird eine geringere Diskriminationsrate zeigen, als der Test in Aulich (2010) mit 16 Sprechern (mit einer BAK > 0,5‰) des ALC.	✓
2. Die Alkoholisierung von weiblichen Sprechern wird wie in Aulich (2010) häufiger korrekt erkennbar sein.	✗
3. Aufgrund des höheren kognitiven Anspruchs beim Lesen können die Hörer in den Beispielen der gelesenen Sprache den in alkoholisiertem Zustand produzierten Stimulus häufiger korrekt erkennen, als in Spontan- und Kommandosprache.	✓
4. Bei Sprechern mit höherer BAK lässt sich der unter Alkoholeinfluss produzierte Stimulus von den Hörern häufiger korrekt erkennen.	✓
5. Bei starken Trinkern wird die Alkoholisierung weniger oft korrekt erkannt, als bei mittleren und leichten Trinkern.	✗
6. Bei Sprechern, die im Mittel einen größeren Unterschied in ihrer mittleren Grundfrequenz zwischen alkoholisiert und nüchtern produzierten Stimuli aufweisen, wird der alkoholisiert produzierte Stimulus häufiger korrekt erkannt.	✓
7. Bei Sprechern, die im Mittel einen größeren Unterschied in ihrem Tonumfang zwischen den beiden Stimuli eines Paares aufweisen, wird der alkoholisiert produzierte Stimulus häufiger korrekt erkannt.	✗
8. Hörer zeigen in der Auswahl des unter Alkoholeinfluss produzierten Stimulus eine Präferenz für den Stimulus mit der höheren mittleren Grundfrequenz.	✓
9. Hörer zeigen in der Auswahl des unter Alkoholeinfluss produzierten Stimulus eine Präferenz für den Stimulus mit dem größeren Tonumfang.	✓

Tab. 4.9: Übersicht über die Hypothesentests

4.3 Diskriminationstests mit manipulierter Grundfrequenz

Im Folgenden werden zwei unterschiedliche Diskriminationstests mit Stimuli mit manipulierter Grundfrequenz vorgestellt. Das Experimentdesign ist dasselbe wie in dem Test, der in Kapitel 4.2 beschrieben wurde. Zunächst wird in Kapitel 4.3.1 beschrieben, auf welche Weise die Grundfrequenz für die verschiedenen Tests manipuliert wurde und der Algorithmus zur automatischen Manipulation vorgestellt. In Kapitel 4.3.3 folgt eine Auswertung der Ergebnisse im Vergleich zu den Resultaten aus dem Test mit Originalstimuli (vgl. Kapitel 4.2.3), sowie eine abschließende Diskussion in Kapitel 4.3.4.

4.3.1 Methode

Aufgrund der in Kapitel 4.2 vorgestellten Ergebnisse liegt die Vermutung nahe, dass die Grundfrequenz für die Hörer in der Unterscheidung von Sprache unter Alkoholeinfluss und Sprache in nüchternem Zustand ein relevantes Entscheidungsmerkmal darstellt. Ein Großteil der Hörer zeigte eine deutliche Präferenz für den Stimulus mit dem höheren F_0 -Median sowie für den Stimulus mit dem größeren F_0 -IQA (vgl. Kapitel 4.2.3.3). Je höher diese Präferenz, desto mehr Erfolg hatten die Hörer auch. Um diese Hypothese zu testen, wurde der in Kapitel 4.2 beschriebene Test erneut durchgeführt, allerdings mit zum Teil veränderter Grundfrequenz in den Stimuli. Hierfür wurden einerseits die unter Alkoholeinfluss produzierten Stimuli in der Grundfrequenz und deren Variation dieser wie im folgenden beschrieben an die in nüchternem Zustand produzierten Stimuli des jeweiligen Sprechers angeglichen, um Grundfrequenzeffekte zu kompensieren und andererseits in den Stimuli der Kontrollgruppe durch eine pauschale Anhebung und Streckung der Grundfrequenzkontur in einer der Aufnahmen in nüchternem Zustand die F_0 -Effekte simuliert. Auch an diesem Test nahmen 72 Hörer (36 weibliche und 36 männliche) teil, deren Rekrutierung und Befragung auf die gleiche Weise wie in Kapitel 4.2.1.2 beschrieben wurde, erfolgte. Die durch einen von

Alter	Geschlecht	Bundesland Grundschule	Hörtesterfahrungen
19 – 36 Jahre Ø 24,5	36w, 36m	BY: 36; BW: 10; BE: 5; RP: 5; NW: 4; SH: 3; HE: 2; NI: 2; A: 2; BB, SL, TH: 1	nein: 43; einmal: 21; zweimal: 6; fünfmal: 2

Tab. 4.10: Übersicht der Hörer in den Tests mit manipulierten Stimuli – *w* = weiblich, *m* = männlich; *BY* = Bayern, *BW* = Baden-Württemberg, *BE* = Berlin, *RP* = Rheinland-Pfalz, *NW* = Nordrhein Westfalen, *SH* = Schleswig-Holstein, *HE* = Hessen, *NI* = Niedersachsen, *A* = Österreich, *BB* = Brandenburg, *SL* = Saarland, *TH* = Thüringen

jedem Hörer vorab ausgefüllten Fragebogen erhobenen Metadaten sind in Tabelle 4.10 zusammengefasst.

4.3.1.1 Stimuli

4.3.1.2 Kompensation der F_0 -Effekte – Test 1

Wie in Kapitel 4.2.1.1 beschrieben, wurden für das Perzeptionsexperiment von jedem Sprecher 24 Stimuluspaare ausgewählt, daraus resultieren insgesamt 3864 Stimuluspaare (ohne Kontrollgruppe). Aufgrund der großen Anzahl der Daten musste die Manipulation automatisch durchgeführt werden. Hierfür wurden zuerst *PitchTier*-Objekte jedes einzelnen Stimulus in *Praat* (Boersma & Weenink 2014) extrahiert und zur Weiterverarbeitung in *R* (Team R Core 2014) geladen. Ein *PitchTier*-Objekt enthält Zeitinformationen und die dazugehörigen F_0 -Werte, allerdings ohne Informationen zur Stimmhaftigkeit. Das heißt, stimmlose Sequenzen werden linear interpoliert. *PitchTier*-Objekte können in *Praat* manuell verändert werden und mittels Resynthese das ursprüngliche *PitchTier*-Objekt ersetzen. Für diese Manipulation werden die einzelnen *PitchTier*-Objekte als Tabellen in *R* geladen, dort mit dem im Folgenden beschriebenen Algorithmus verändert und später wieder in ein *PitchTier*-Objekt umgewandelt. Daraufhin erfolgte die Resynthese mittels des in *Praat* implementierten Algorithmus *overlap-add* (Moulines & Charpentier 1990).

Für die Berechnung der *PitchTier*-Objekte verwendet *Praat* einen Autokorrelationsalgorithmus, der in Boersma (1993) beschrieben wird. Aufgrund der Menge an zu

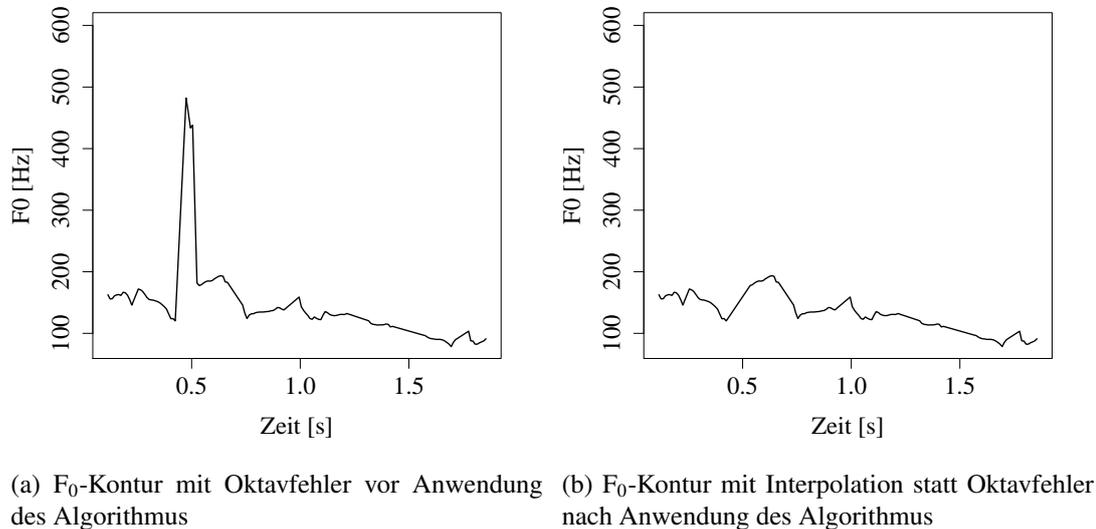
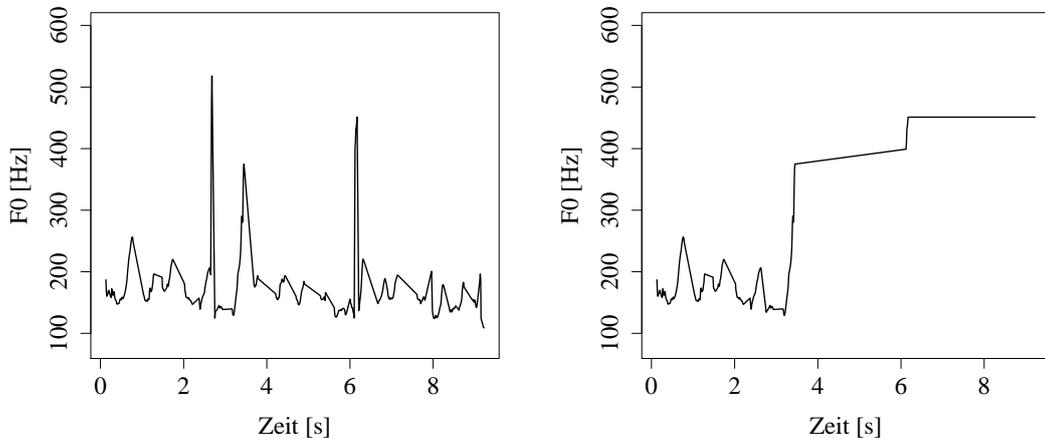


Abb. 4.9: Beseitigung von Oktavfehlern

manipulierenden Stimuli und der dadurch fehlenden Möglichkeit, die minimale und maximale Frequenz zur Berechnung der F_0 für jeden Sprecher einzeln anzupassen, dient hier der Standardfrequenzbereich von 75 Hz bis 600 Hz als Grundlage. Eventuelle Oktavfehler, die bei der späteren Resynthese zu Problemen führen, werden mit einem Algorithmus, der die mittlere Grundfrequenz des jeweiligen Stimulus berücksichtigt und im folgenden beschrieben wird in R korrigiert.

Algorithmus zur Beseitigung von Oktavfehlern

Wenn ein Wert mehr als 50% über oder 45% unter dem vorhergehenden Wert liegt, wird dieser Eintrag in der Tabelle gelöscht und der nächste Wert wird mit dem vorhergehenden verglichen (das führt später in *Praat* zu keinen Problemen, da *Praat* einfach zwischen zwei vorhandenen Werten interpoliert). Damit dies bei Oktavfehlern am Anfang eines Signals nicht zu Fehlern führt, wird getestet, ob der erste Wert im Bereich von 45% – 150% des Medians des jeweiligen Stimulus liegt – ansonsten wird er gelöscht und der nächste Wert wird überprüft. Die festgelegten Grenzen basieren auf Beobachtungen und führen zu den wenigsten Fehlern.



(a) F_0 -Kontur mit Oktavfehler vor Anwendung des Algorithmus (b) Beispiel für einen Fehler in der F_0 -Kontur nach Anwendung des Algorithmus

Abb. 4.10: Probleme in der Oktavfehlerkorrektur

Einige Fehler sind allerdings nicht vermeidbar. Beispielsweise ist in Abb. 4.10a die von *Praat* erzeugte Originalkurve mit einigen Oktavsprüngen erkennbar, in 4.10b die Kurve nach der Modifikation durch den Algorithmus. Der erste Oktavfehler (bei ca. 2,6s) wird korrigiert, beim zweiten jedoch folgt auf einen für die Ausschlusskriterien zu flachen Anstieg bis hin zu ca. 375 Hz ein starker Abfall auf etwa 180 Hz. Dieser Abfall fällt wiederum für den Algorithmus in die Kategorie ‚Oktavfehler‘ und somit auch alle verbleibenden Werte, bis der nächste Oktavfehler als ‚normal‘ erkannt wird und für einen weiteren leichten Anstieg sorgt. Solche Fehler sorgen nach der Resynthese natürlich für eine extrem unnatürliche Sprachmelodie. Fehler in dieser Art betreffen ca. 10% der verwendeten Stimuli.

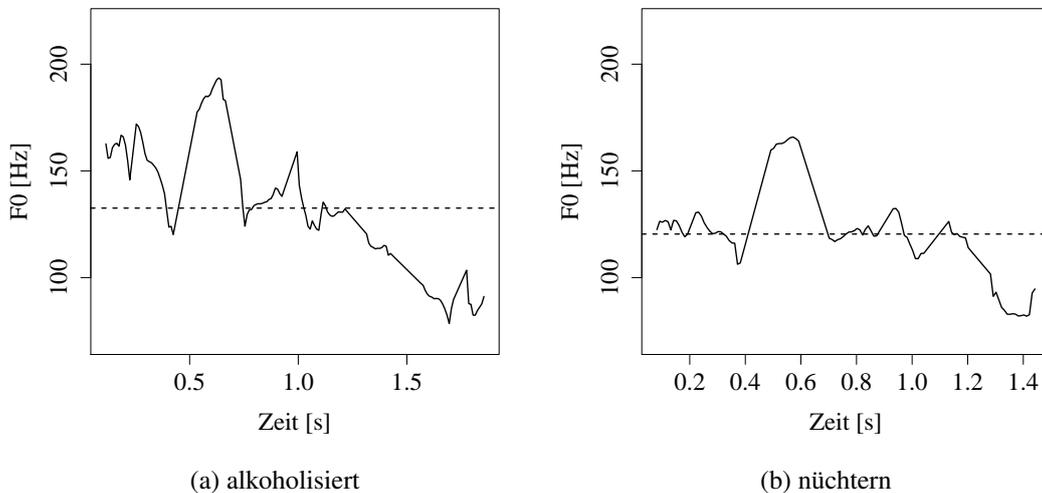


Abb. 4.11: F_0 -Konturen der Äußerung „Temperatur 23°C“ eines männlichen Sprechers

Nach der Oktavfehlerkorrektur wird die Kontur jedes in alkoholisiertem Zustand produzierten Stimulus (a -Stimulus) mit einem weiteren Algorithmus in R manipuliert. In den Abb. 4.11a und 4.11b sind die beiden Konturen der Äußerung „Temperatur 23°C“ eines männlichen Sprechers zu sehen, links im alkoholisierten Zustand, rechts nüchtern. In den Abbildungen ist deutlich zu erkennen, dass der Median des a -Stimulus (hier als getrichelte Linie eingezeichnet) höher liegt, als der des in nüchternem Zustand produzierten Stimulus (na -Stimulus), und auch die Sprannbreite der Kontur ist im a -Stimulus größer. Um den Einfluss der F_0 auf die Entscheidung des Hörers zu minimieren, sollen beide Konturen nach der Manipulation im Mittel und in ihrem Umfang sehr ähnlich sein. Hierfür wird die Kontur des a -Stimulus mithilfe des nachfolgend beschriebenen Algorithmus an die des na -Stimulus angepasst.

Algorithmus zur Kompensation der F_0 -Effekte

Zunächst wird die Kontur des a -Stimulus mit dem Verhältnis des Medians der Kontur des na -Stimulus zu dem Median der Kontur des a -Stimulus (k_1) multipliziert. Die resultierende Kontur wird dann um den (neuen) Median dieser Kontur gestaucht oder gestreckt – abhängig von dem Verhältnis der beiden Interquartilsabstände der ursprüng-

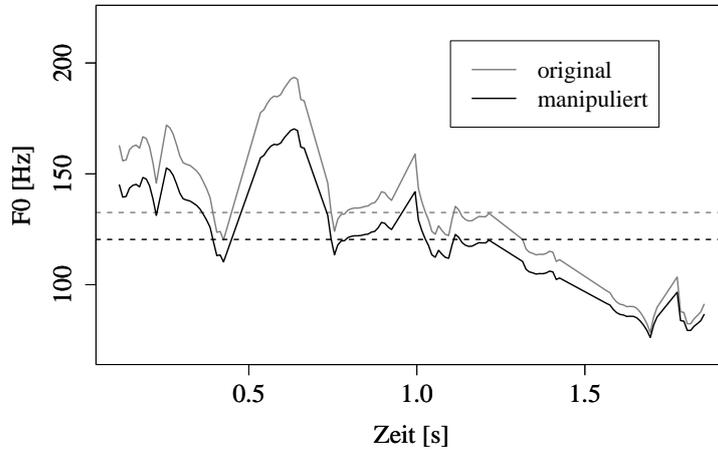


Abb. 4.12: F_0 -Konturen des a -Stimulus unverändert (grau) und nach der Manipulation (schwarz)

lichen Konturen (k_2). Um den a -Stimulus zu manipulieren, wird also folgende Formel (4.1) angewandt:

$$f0_{neu} = k_2 k_1 f0_a + (1 - k_2) k_1 median_{f0_a} \quad (4.1)$$

wobei

$$k_1 = \frac{median_{na}}{median_a} \quad (4.2)$$

und

$$k_2 = \frac{IQA_{na}}{IQA_a} \quad (4.3)$$

Da eine zu starke Manipulation den Stimulus unnatürlich klingen lässt, und es dazu kommen kann, dass (gerade bei sehr kurzen Stimuli) beide Stimuli eines Paares mit sehr unterschiedlich hoher Grundfrequenz produziert wurden, musste hier eine maximale Grenze eingeführt werden. Es konnte bei keinem Stimulus die Grundfrequenz um mehr als 10% angehoben oder abgesenkt werden und auch die Variabilität wurde nicht um mehr als 10% verändert. Sollten k_1 oder k_2 also einen Wert außerhalb des

erlaubten Bereiches von 0,9 bis 1,1 annehmen, wird der Wert automatisch auf die nächstgelegene Grenze gesetzt. Abb. 4.12 zeigt die Originalkontur des *a*-Stimulus in grau sowie die neue, gestauchte und nach unten verschobene Kontur in schwarz.

Um zu vermeiden, dass eventuelle hörbare Artefakte, die in der Resynthese entstehen können, die Entscheidung der Hörer beeinflussen, wurden auch die *na*-Stimuli (ohne die Grundfrequenzkontur zu verändern) in *Praat* mittels *overlap-add* resynthetisiert.

4.3.1.3 Simulation der F_0 -Effekte – Test 2

Bei der Kontrollgruppe (in der die Sprecher beide Stimuli eines Paares nüchtern produzierten), wurde die F_0 -Kontur einer der beiden Stimuli um den festen Wert von 5% angehoben und ebenso um den festen Wert von 5% gestreckt. Die Manipulation erfolgte ebenfalls durch den oben beschriebenen Algorithmus zur Kompensation der F_0 -Effekte (Formel 4.1), jedoch wurden die beiden Parameter k_1 und k_2 für jeden zu manipulierenden Stimulus auf den festen Wert 1,05 gesetzt. Auch hier wurde der Stimulus, in dem die Grundfrequenz nicht manipuliert wurde, ebenfalls resynthetisiert.

4.3.2 Hypothesen

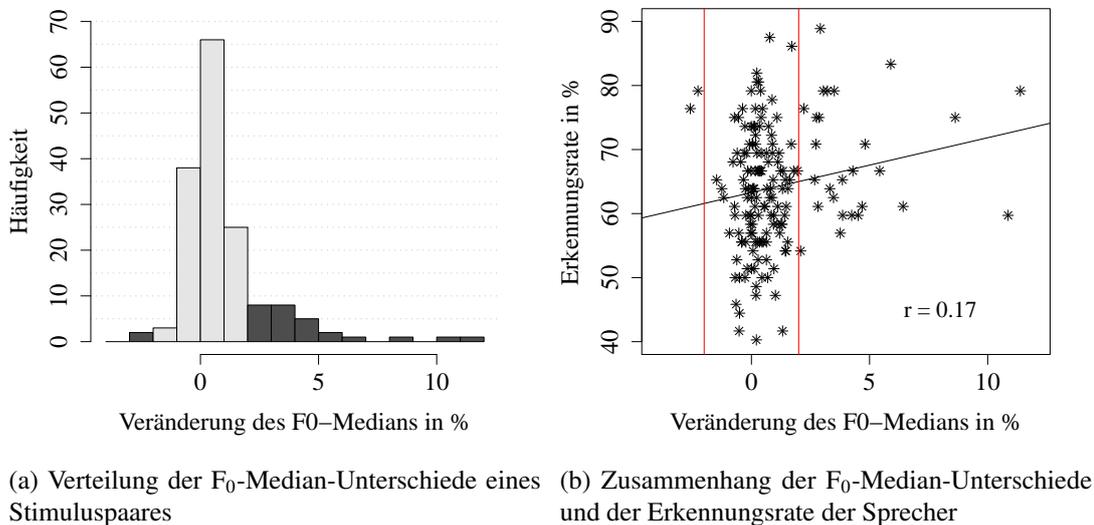
Der Einfachheit halber wird der Test mit den Stimuli, in denen die F_0 -Effekte kompensiert wurden, im Folgenden als Test 1 bezeichnet, der Test mit den Stimuli, in denen die F_0 -Effekte simuliert wurden, als Test 2. Es lassen sich folgende Hypothesen aufstellen:

1. Die Erkennungsleistung der Hörer in Test 1 ist schlechter als die im Diskriminationstest mit den Originalstimuli.
2. In Test 2 wählen die Hörer häufiger den Stimulus mit simulierten F_0 -Effekten als den unter Akoholeinfluss produzierten aus.

4.3.3 Ergebnisse

4.3.3.1 Kompensation der F_0 -Effekte – Test 1

Wie in Kapitel 4.3.1.1 beschrieben, kommt es bei der Ausbesserung von Oktavfehlern in der Grundfrequenzkontur eines Stimulus zu Fehlern. Um den Einfluss dieser Fehler auf die Ergebnisse möglichst gering zu halten, wurden die Sprecher, bei denen ein solcher Fehler besonders häufig auftrat, in der Auswertung von Test 1 ausgeschlossen. Hierfür wurde für jeden Sprecher der durchschnittliche Unterschied des F_0 -Medians zwischen zwei Stimuli eines Paares berechnet und mit der Erkennungsrate des Sprechers verglichen. Die Erkennungsrate wurde, wie in Kapitel 4.2.3 für den Diskriminationstest mit den Originalstimuli beschrieben, berechnet. Abb. 4.13a zeigt die Verteilung der mittleren F_0 -Median-Unterschiede eines Sprechers zwischen zwei Stimuli eines Paares *nach* der Manipulation. Ziel dieses Perzeptionstests war es, den Einfluss von F_0 möglichst gering zu halten und so als Entscheidungskriterium für die Hörer zu eliminieren. Aufgrund der oben beschriebenen Probleme in der Synthese der Stimuli und der gesetzten Obergrenze bei der Manipulation ist jedoch bei einigen Sprechern der mittlere F_0 -Median-Unterschied zwischen zwei Stimuli eines Paares größer als 2% (in Abb. 4.13a als dunkle Balken verdeutlicht). Um beurteilen zu können, ob diese Unterschiede die Einschätzungen zu einem Sprecher beeinflussen, wurde der Zusammenhang zwischen den F_0 -Median-Unterschieden und der Erkennungsrate der Sprecher in Test 1 in einer linearen Regression untersucht. In Abb. 4.13b ist das Streudiagramm mit dem berechneten Regressionskoeffizienten r und der Regressionsgeraden dargestellt. Die lineare Regression zeigt einen zwar schwachen, aber signifikanten Zusammenhang zwischen dem F_0 -Median-Unterschied und der Erkennungsrate eines Sprechers ($F(1, 159) = 10,43, p < 0,01$). Ob der Zusammenhang tatsächlich auf den großen F_0 -Unterschied zurückzuführen ist, oder ob Hörer beispielsweise einen ‚komisch‘ klingenden Stimulus als den unter Alkoholeinfluss produzierten auswählen, ist unklar, jedoch soll genau dieser Zusammenhang in diesem Test keine Rolle spielen. Deshalb wurden für die folgende Analyse der Ergebnisse die Sprecher,

Abb. 4.13: F_0 -Median-Unterschiede

bei denen im Mittel der F_0 -Median-Unterschied zwischen zwei Stimuli eines Paares größer als 2% ist, ausgeschlossen. Gegenstand der Auswertung sind also nur noch 132 Sprecher. Diese sind in Abb. 4.13a mithilfe heller Balken hervorgehoben und liegen in Abb. 4.13b zwischen den eingezeichneten vertikalen Linien. Eine lineare Regression nach Ausschluss der 29 Sprecher zeigt keinen signifikanten Zusammenhang mehr zwischen dem F_0 -Median-Unterschied und der Erkennungsrate eines Sprechers ($F(1, 130) = 0,84, p > 0,1, r = 0,08$).

Um die Ergebnisse mit dem in Kapitel 4.2 beschriebenen Test mit den Originalstimuli vergleichbar zu machen, wurden diese beiden Raten hierfür mit den verbleibenden 132 Sprechern von Test 1 erneut berechnet. Die durchschnittliche Erkennungsleistung der Hörer in Test 1 liegt bei 61,6% und weicht somit im Mittel kaum von der Erkennungsleistung der Hörer im Experiment mit Originalstimuli (61,8% bei 132 verbleibenden Sprechern) ab. Abb. 4.14a und 4.14b zeigen die Verteilungen der Erkennungsleistung der Hörer im Vergleich. Die Spanne der Erkennungsleistungen von 48,5% bis 74,2% in Test 1 liegt ein wenig niedriger als die der Erkennungsleistungen im Experiment mit Originalstimuli, welche von 51,5% bis 76,5% reicht.

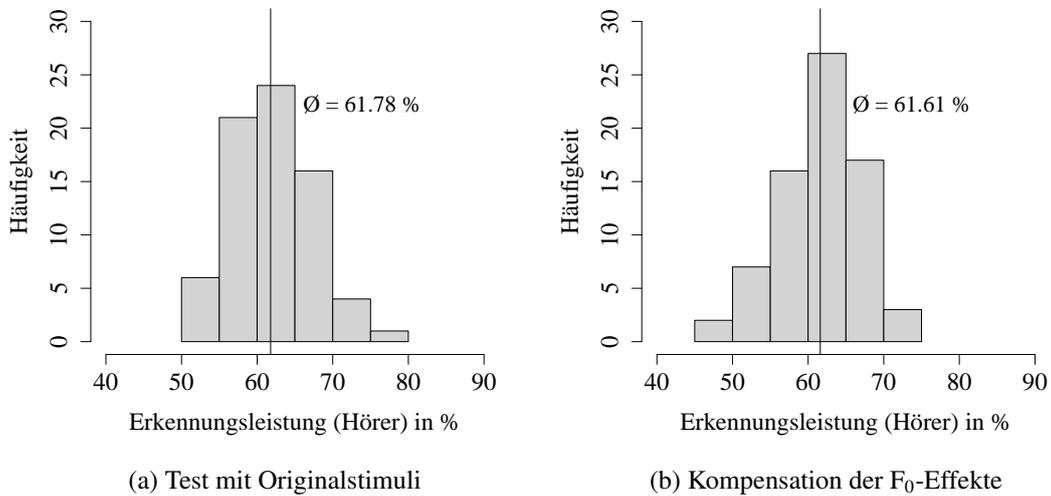


Abb. 4.14: Vergleich der Erkennungsleistung der Hörer (132 Sprecher und 72 Hörer)

Die durchschnittliche Erkennungsrate der verbleibenden 138 Sprecher im Test mit den Originalstimuli liegt bei 63,2% und ist nahezu identisch zu der durchschnittlichen Erkennungsrate von 63% der Sprecher in Test 1. Die Verteilungen der Erkennungsrate der Sprecher für beide Tests im Vergleich ist in den Abb. 4.15a und 4.15b zu sehen. Eine *MEM*-Analyse mit den Antworten der Hörer aus beiden Tests als abhängige Variable und dem *fixed*-Faktor Manipulation (ja oder nein) und Sprechern und Hörern als *random*-Faktoren zeigt keinen signifikanten Unterschied zwischen den Ergebnissen beider Tests ($\chi^2[1] = 0,02, p > 0,1$).

Auch hier wird der Unterschied in der Spannweite der Erkennungsleistung der Hörer und der der Erkennungsrate der Sprecher deutlich. Die Ergebnisse sind zwar auch hörerabhängig, jedoch liegen sie für die Hörer deutlich näher beieinander als die die einzelnen Sprecher betreffend. Hier variiert die Erkennbarkeit von Sprecher zu Sprecher um einiges mehr.

Auf mögliche nicht-akustische Einflussfaktoren auf die Antworten der Hörer wurde ebenso wie in Kapitel 4.2.3 mit einer *MEM*-Analyse mit Sprechern und Hörern als *random*-Faktoren getestet. Trotz größtenteils unterschiedlicher Hörer und Kompensation der F₀-Effekte ergeben sich ähnliche Ergebnisse wie im Test mit den Originalstimuli. Auch hier beeinflusst die BAK eines Sprechers und der Sprachstil des Beispiels

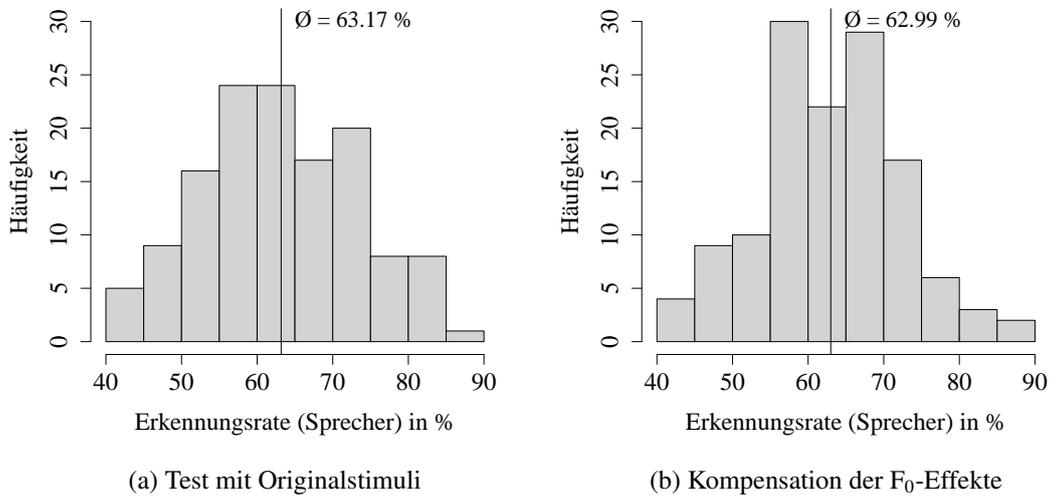


Abb. 4.15: Vergleich der Erkennungsrate der Sprecher (132 Sprecher und 72 Hörer)

Faktor	Ergebnisse der MEM-Analyse
Sprechergeschlecht	$\chi^2[1] = 0,58, \quad p > 0,1 \text{ n.s.}$
Sprachstil	$\chi^2[2] = 76,72, \quad p < 0,001 \text{ ***}$
Trinkgewohnheiten	$\chi^2[2] = 2,14, \quad p > 0,1 \text{ n.s.}$
BAK	$\chi^2[1] = 15,16, \quad p < 0,001 \text{ ***}$
Hörergeschlecht	$\chi^2[1] = 0,41, \quad p > 0,1 \text{ n.s.}$

Tab. 4.11: Einflussfaktoren auf die Antwort der Hörer in Test 1 – *n.s.* = nicht signifikant, . = signifikanter Trend, * = signifikant, ** = hoch signifikant, *** = höchst signifikant

die Antwort der Hörer. Die Ergebnisse der MEM-Analyse der nicht-akustischen Einflussfaktoren sind in Tabelle 4.11 zusammengefasst.

4.3.3.2 Simulation der F₀-Effekte – Test 2

In Test 2, in dem F₀-Effekte in den verwendeten Stimuli simuliert wurden, dienten die Aufnahmen von 20 Kontrollgruppensprechern als Sprachmaterial. Im Test mit Originalstimuli wurden den Hörern immer zwei unterschiedliche Aufnahmen eines Sprechers präsentiert, in denen der Sprecher aber in beiden Fällen nüchtern war. In Test 2 wurden dieselben Stimuli verwendet, in jedem Stimuluspaar wurde jedoch einer von beiden, wie in Kapitel 4.3.1.3 beschrieben, manipuliert. Die durchschnittliche Erken-

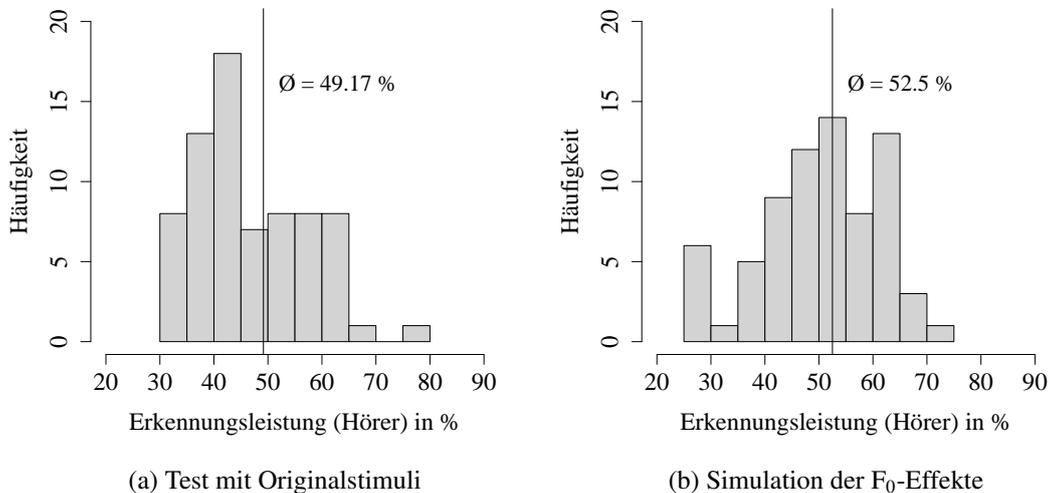


Abb. 4.16: Vergleich der Erkennungsleistung der Hörer (20 Sprecher und 72 Hörer)

nungsleistung der Hörer liegt mit 52,5% leicht über der der Kontrollgruppensprecher im Experiment mit den Originalstimuli von 49,2% bzw. 50,8%. Da es in diesem Fall keine *richtige* Antwort gab, zeigt die Erkennungsleistung hier eine Präferenz der Hörer für einen der beiden Aufnahmezeitpunkte. Die Verteilungen im Vergleich sind in den Abb. 4.16a und 4.16b zu sehen. Die durchschnittliche Erkennungsrate der Sprecher in Test 2 liegt im Durchschnitt bei 53,9% und damit etwas höher als die im Test mit den Originalstimuli – hier liegt der Durchschnitt bei 50,6%. Abb. 4.17a und 4.17b zeigen die beiden Verteilungen. Der Unterschied zwischen den Ergebnissen beider Tests wurde ebenfalls mit einer *MEM*-Analyse mit den Antworten aus beiden Tests als abhängige Variable, dem *fixed*-Faktor Manipulation und Sprechern und Hörern als *random*-Faktoren auf Signifikanz getestet, und obwohl ein leicht besseres Ergebnis in Test 1 erkennbar ist, ist der Unterschied nur marginal signifikant ($\chi^2[1] = 3,26, p < 0,1$).

Um eventuelle Einflussfaktoren auf die Hörerentscheidung auszumachen, wurde auch hier eine *MEM*-Analyse durchgeführt – ebenfalls mit der Antwort der Hörer als abhängige Variable, den *fixed*-Faktoren Sprechergeschlecht, Sprachstil und Hörergeschlecht, und Sprechern und Hörern als *random*-Faktoren. Weder das Sprecher- noch das Hörergeschlecht hatte allerdings signifikanten Einfluss auf die Ergebnisse und auch nicht die unterschiedlichen Sprachstile. Bei den akustischen Parametern, die in diesem

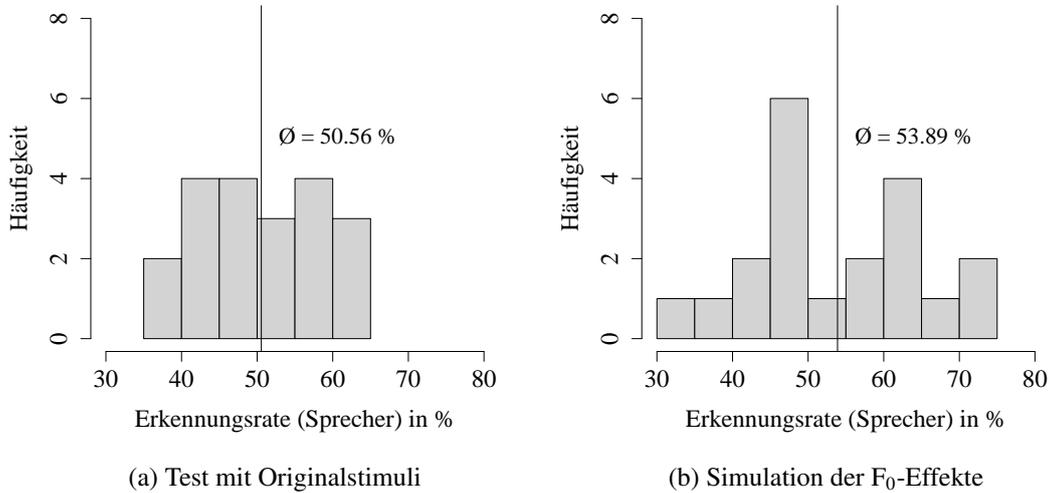


Abb. 4.17: Vergleich der Erkennungsrate der Sprecher (20 Sprecher und 72 Hörer)

Faktor	Ergebnis der MEM-Analyse
Sprechergeschlecht	$\chi^2[1] = 0,3, \quad p > 0,1 \text{ n.s.}$
Sprachstil	$\chi^2[2] = 4,1, \quad p > 0,1 \text{ n.s.}$
Hörergeschlecht	$\chi^2[1] = 2,09, \quad p > 0,1 \text{ n.s.}$
Veränderung des F ₀ -Medians	$\chi^2[1] = 0,12, \quad p > 0,1 \text{ n.s.}$
Veränderung des F ₀ -IQA	$\chi^2[1] = 0,0, \quad p > 0,1 \text{ n.s.}$

Tab. 4.12: Einflussfaktoren auf die Antwort der Hörer in Test 2 – *n.s.* = nicht signifikant, . = signifikanter Trend, * = signifikant, ** = hoch signifikant, *** = höchst signifikant

Test manipuliert wurden, zeigte in einer MEM-Analyse, ebenfalls mit der Antwort der Hörer als abhängige Variable und den unabhängigen Faktoren F₀-Median-Unterschied und F₀-IQA-Unterschied und Sprechern und Hörern als *random*-Faktoren, weder der F₀-Median-Unterschied noch der des F₀-IQA einen signifikanten Einfluss auf die Entscheidung der Hörer. Die Ergebnisse der MEM-Analyse sind in Tabelle 4.12 zusammengefasst.

4.3.4 Zwischendiskussion

Obwohl die Einflüsse der Grundfrequenz auf den Hörer durch die Kompensation der Effekte minimiert wurde, schnitten die Hörer in Test 1 nicht merklich schlechter ab

als in dem Test mit den Originalstimuli – Hypothese 1 muss also verworfen werden. Aufgrund der Ergebnisse in den Kapiteln 3.2.3 und 4.2.3 wurde vermutet, dass F_0 als Entscheidungskriterium für die Hörer in der Diskrimination von Sprache ohne und mit Alkoholeinfluss dient. Ist jedoch kein Unterschied mehr in der F_0 zwischen zwei Stimuli eines Paares vorhanden, wurde dennoch überzufällig (in durchschnittlich 61,6% der Fälle) der in alkoholisiertem Zustand produzierte Stimulus von den Hörern richtig erkannt. Der in Kapitel 4.2.3 beschriebene Zusammenhang zwischen der Erkennungsrate eines Sprechers und dem Ausmaß der Veränderung seiner F_0 -Parameter in nüchternem und alkoholisiertem Zustand lässt sich also nicht allein durch diesen Unterschied erklären. Möglich ist, dass die Veränderungen der F_0 -Parameter lediglich als Nebeneffekt anderer Veränderungen auftreten, an denen die Hörer sich orientieren. Diese können sowohl akustischer Natur sein, jedoch auch lexikalische oder paralinguistische Veränderungen betreffen, die im Rahmen dieser Arbeit nicht untersucht wurden.

In Test 2 fehlen jedoch andere Merkmale, die den Hörer auf eine Alkoholisierung des Sprechers schließen lassen, da immer zwei Aufnahmen eines Sprechers miteinander verglichen wurden, in denen der Sprecher in beiden Fällen nüchtern war. Die Simulation der F_0 -Effekte führt hier in einigen Fällen tatsächlich dazu, dass der manipulierte Stimulus eher als der unter Alkoholeinfluss produzierte ausgewählt wurde. Der Unterschied zwischen den Ergebnissen von Test 2 und denen der Kontrollgruppe im Test mit den Originalstimuli ist allerdings gering, und die Veränderungen der F_0 -Parameter scheinen laut *MEM*-Analyse keine Einflussfaktoren auf die Entscheidung der Hörer darzustellen. Hypothese 2 muss also auch verworfen werden, auch wenn eine leichte Präferenz der Hörer für den Stimulus mit veränderter F_0 -Kontur erkennbar ist. Die Veränderung der F_0 -Parameter könnte hier als eine Art Merkmal fungieren, das von den Hörern nur in Betracht gezogen wird, wenn sich an keinem anderen Entscheidungskriterium mehr orientiert werden kann. Eine Übersicht der Hypothesentests zeigt Tabelle 4.13.

Zusammenfassend lässt sich also sagen, dass die Grundfrequenz eines Sprechers für die Hörer in diesem Test kein wesentliches Entscheidungsmerkmal darstellt, wenn

sie von zwei Aufnahmen diejenige Aufnahme auswählen sollen, in der der Sprecher alkoholisiert war. Sie scheinen sich an anderen Merkmalen zu orientieren, die anhand dieser Arbeit allerdings nicht festgemacht werden können.

Hypothese	Angenommen?
1. Die Erkennungsleistung der Hörer in Test 1 ist schlechter als die im Diskriminationstest mit den Originalstimuli.	X
2. In Test 2 wählen die Hörer häufiger den Stimulus mit simulierten F ₀ -Effekten als den unter Akoholeinfluss produzierten aus.	X

Tab. 4.13: Übersicht über die Hypothesentests

5

Zusammenfassende Diskussion und Ausblick

Ein Ziel dieser Arbeit war es, anhand einer großen Datenbasis mit unterschiedlich stark alkoholisierten Sprechern den Einfluss von Alkohol auf die Grundfrequenz zu untersuchen. Bisherige Studien zu dem Thema zeigten Uneinigkeit, ob die Grundfrequenz eines Sprechers in alkoholisiertem Zustand im Vergleich zum nüchternen Zustand höher oder niedriger ist oder sogar gleich bleibt. Alle drei Möglichkeiten konnten in unterschiedlichen Experimenten nachgewiesen werden (vgl. Kapitel 3.1). An den meisten bisherigen Studien nahmen allerdings nur sehr wenige Sprecher teil, die größtenteils männlich waren. In dem in dieser Arbeit durchgeführten Experiment zeigt sich ein deutliches Ergebnis für die Sprecher des ALC. Bei 79% der 148 untersuchten Sprecher konnte in alkoholisiertem Zustand (mit einer gemessenen BAK von mindestens 0,5‰) eine höhere mittlere Grundfrequenz gemessen werden. Allerdings fällt eine große Intersprechervariabilität auf – einige Sprecher zeigen hier einen sehr großen Unterschied zwischen den beiden Aufnahmen, andere nur einen sehr kleinen, bei wenigen anderen konnte wiederum eine niedrigere Grundfrequenz in alkoholisiertem Zustand gemessen werden. Nicht nur die Anzahl der Sprecher, die eine höhere Grundfrequenz verwenden, ist höher, sondern auch die Spannweite des Ausmaßes der Veränderungen. Nicht ganz so deutlich ist das Ergebnis für den genutzten Tonumfang. Zwar ist auch hier ein

Anstieg des Interquartilsabstands bei einem Großteil der Sprecher erkennbar, dieser variiert allerdings je nach Sprachstil. Am größten ist der Anteil dieser Sprecher in der Kommandosprache (68%), am geringsten in der gelesenen Sprache (59,5%). Auch hier variieren die Ergebnisse für die einzelnen Sprecher stark. Wider Erwarten lassen sich die Sprecherunterschiede nicht durch die unterschiedlichen BAK der Sprecher erklären. Die berechneten linearen Regressionen zeigen keinen oder höchstens einen sehr schwachen Zusammenhang zwischen der Veränderung der Grundfrequenzparameter und der BAK eines Sprechers (der stärkste liegt bei $r = 0,21$ für den F_0 -Median in der Spontansprache der Frauen).

Um die Annahme zu testen, Menschen seien generell dazu in der Lage die Alkoholisierung einer Person lediglich anhand seiner Sprache zu erkennen, wurde mit Sprachbeispielen von 161 Sprechern (ebenfalls aus dem ALC) zunächst ein Diskriminationstest mit 72 Hörern durchgeführt. Eine vergleichbar große Zahl an Sprechern in einem Test zur Wahrnehmbarkeit der Alkoholisierung eines Sprechers ist in der bisherigen Literatur nicht zu finden (vgl. Kapitel 4.1). Die durchschnittliche Diskriminationsrate der Hörer liegt mit 63,1% über dem Zufallsniveau von 50%. Die Hörer wählten also aus zwei Sprachbeispielen eines Sprechers überzufällig richtig das unter Alkoholeinfluss produzierte aus. In der bisherigen Literatur konnten zum Teil auch höhere Diskriminationsraten erzielt werden (vgl. z.B. Aulich (2010), Pisoni et al. (1985) und Klingholz et al. (1988)), allerdings lassen diese sich in den meisten Fällen mit der sehr hohen AAK bzw. BAK eines Sprechers und der geringen Anzahl an Sprechern und/oder Hörern erklären. In dem Diskriminationstest dieser Arbeit konnte ein signifikanter Zusammenhang zwischen der Erkennungsrate eines Sprechers und seiner BAK gezeigt werden, dennoch verändern sich die Diskriminationsraten nach Ausschluss von Sprechern mit einer BAK von 0,5‰ bzw. 1,0‰ und weniger um lediglich 1,7% bzw. 3,6% und liegen somit immernoch niedriger als die in Aulich (2010), Pisoni et al. (1985) und Klingholz et al. (1988) mit vergleichbar stark alkoholisierten Sprechern. Ein Vergleich der unterschiedlichen Sprachstile zeigt, dass die Hörer in der gelesenen Sprache eine deutlich höhere Diskriminationsrate erreichen. Ein möglicher Grund

hierfür ist die größere kognitive Leistung, die beim Lesen erbracht werden muss wodurch die Auswirkungen der Alkoholisierung auf die Sprache deutlicher zum Tragen kommen. Außerdem könnten das in der gelesenen Sprache und der Kommandosprache eventuell noch vorhandene Lachen oder mögliche Versprecher den Hörern erleichtern, eine Alkoholisierung zu erkennen. In den Beispielen der Spontansprache wurden derartige Einflüsse vermieden.

Ein gezeigter Zusammenhang zwischen der Veränderung der mittleren Tonhöhe und den Erkennungsraten der einzelnen Sprecher lässt zunächst einen Einfluss der Grundfrequenz auf die Hörerentscheidung vermuten, welcher auch in einer *MEM*-Analyse nachgewiesen wurde. Je größer der Unterschied in der mittleren Tonhöhe zwischen den Aufnahmen eines Sprechers in alkoholisiertem Zustand und denen in nüchternem Zustand, desto häufiger wurde das unter Alkoholeinfluss produzierte Sprachbeispiel auch als solches erkannt. Auch eine nachgewiesene Präferenz des Großteils der Hörer, den Stimulus mit der höheren mittleren Grundfrequenz als den unter Alkoholeinfluss produzierten auszuwählen, stützt die Hypothese, die Grundfrequenz sei ein Entscheidungsmerkmal für die Hörer. Aus diesem Grund wurde dieser Zusammenhang in zwei weiteren Diskriminationstests getestet, in denen die Grundfrequenz in den Sprachbeispielen manipuliert wurde. Zum einen wurde durch eine Anpassung der Tonhöhe und des Tonumfangs der unter Alkoholeinfluss produzierten Sprachbeispiele an die der nüchtern produzierten Sprachbeispiele eines Sprechers eine Minimierung der Grundfrequenzeinflüsse erzielt. Zum anderen im Vergleich zweier Sprachbeispiele in nüchternem Zustand durch Manipulation der Grundfrequenz in einem Sprachbeispiel die durch Alkoholisierung hervorgerufene Grundfrequenzveränderungen simuliert. Eine Verschlechterung der Diskriminationsrate wenn die Grundfrequenz kein Entscheidungsmerkmal mehr sein kann (da sie in den beiden Sprachbeispielen, die miteinander verglichen wurden, aneinander angepasst wurde), spräche dafür, dass die Grundfrequenz als Entscheidungsmerkmal zwischen alkoholisiert und nüchtern produzierten Äußerungen eines Sprechers fungiert. Die Ergebnisse dieses Tests zeigen wider Erwarten jedoch beinahe eine identische Diskriminationsrate, verglichen mit dem Basistest

mit unveränderten Stimuli. Ein Wegfallen der Grundfrequenz als mögliches Diskriminationsmerkmal in diesem Test führt also nicht zu einer schlechteren Erkennbarkeit der Alkoholisierung des Sprechers. Demnach müssen hier andere Faktoren eine Rolle spielen, die eventuell gemeinsam mit einer Veränderung der Grundfrequenz einhergehen. Als solche kommen andere akustische Faktoren in Betracht, die in dieser Arbeit nicht untersucht wurden, oder aber andere, eventuell linguistische oder paralinguistische Faktoren. Klingholz et al. (1988) führten in ihrem Perzeptionstest zusätzlich eine Befragung der Sprecher durch, auf welche Merkmale sie ihre Antwort stützten, und auch hier wurden stimmliche Merkmale (wie z.B. Stimmqualität oder Stimmstabilität) seltener genannt als artikulatorische Merkmale (wie z.B. Sprachfluss oder Sprachfehler). Das Ergebnis dieser Befragung unterstützt die Vermutung dass die Grundfrequenz in der Beurteilung nicht die bedeutendste Rolle spielt.

In dem Test, in dem eine Alkoholisierung durch eine Veränderung der Grundfrequenz in nüchtern produzierten Sprachbeispielen simuliert wurde, zeigt sich allerdings eine leichte Tendenz der Hörer, dieses Sprachbeispiel als das unter Alkoholeinfluss produzierte auszuwählen, was wiederum dafür spräche, die Grundfrequenz als Diskriminationsmerkmal anzusehen. Jedoch ist diese Tendenz nur schwach und deutet eher darauf hin, dass die Grundfrequenz von den Hörer lediglich als Ausweichmerkmal gesehen wird, wenn andere Merkmale, die auf eine Alkoholisierung des Sprechers hinweisen könnten, nicht vorhanden sind.

Für eine mögliche automatische Erkennung von Alkoholisierung anhand lautsprachlicher Äußerungen eines Sprechers scheint die Grundfrequenz allerdings ein vielversprechendes Merkmal zu sein. Aufgrund der starken Variation zwischen den verschiedenen Sprechern kann eine solche Erkennung auf Basis der Grundfrequenz allerdings nicht sprecheruniversell funktionieren. Eine sprecherindividuelle Herangehensweise erscheint hier erfolgversprechender. Ein simpler Algorithmus, der in den im Diskriminationstest präsentierten Sprachbeispielen immer das Beispiel mit höherer Grundfrequenz als das unter Alkoholeinfluss produzierte auswählt, wäre in ca. 80% der Fälle erfolgreich. Die Hörer in diesem Test scheinen sich aber nicht darauf

zu verlassen, und sich nicht bewusst zu sein, wie gut die Grundfrequenz als Diskriminationsmerkmal tatsächlich funktioniert. Möglicherweise sind die Hörer aber auch nicht anfällig für Veränderungen der Grundfrequenz, da diese von sehr vielen anderen Faktoren beeinflusst werden kann. Das emotionale Befinden eines Sprechers zum Beispiel kann sich ebenfalls in der Stimmlage abzeichnen (vgl. z.B. Williams & Stevens (1972) und Yildirim et al. (2004)), genauso wie Stress (vgl. z.B. Jessen (2006) und Sigmund (2013)) oder Müdigkeit (vgl. z.B. Krajewski et al. (2009) und Vogel et al. (2010)). Möglicherweise sind die Hörer also daran gewöhnt, dass die Grundfrequenz eines Sprechers nicht immer dieselbe ist und schließen daher bei einer Veränderung nicht zwangsläufig auf eine Alkoholisierung. Dass die Grundfrequenz von sehr vielen verschiedenen Einflussfaktoren abhängig ist, sollte ebenso in den Überlegungen zur automatischen Erkennung der Alkoholisierung eines Sprechers auf Basis der Grundfrequenz bedacht werden. Hier liegt ein möglicher Ansatz für weiterführende Forschung. Bisherige Studien beziehen sich meist auf nur einen Aspekt, der zum Beispiel die Grundfrequenz oder andere akustische Parameter beeinflusst, selten werden diese Dinge aber kombiniert. Auch in der Perzeption stellt die Kombination mehrerer Sprecherzustände einen interessanten Forschungsansatz dar. Ob Hörer anhand der Stimme Müdigkeit und Alkoholisierung unterscheiden können, wäre beispielsweise von Interesse.

Weniger vielversprechend für eine automatische Erkennung sind die Mikroperurbationen der Grundfrequenz und der Lautstärke (Jitter und Shimmer). Diese zeigten in dieser Arbeit eine große Variabilität und nur beim Jitter im Vokal /a:/ einen signifikanten Unterschied zwischen Aufnahmen in nüchternem und alkoholisiertem Zustand. Beim Shimmer war dieser Unterschied nur bei männlichen Sprechern im Vokal /a:/ und im Vokal /i:/ nur in der gelesenen Sprache erkennbar. Für einen sprecherindividuellen Ansatz müssten diese Parameter erneut sprecherspezifisch auf Konsistenz überprüft werden.

Zusammenfassend lässt sich also feststellen, dass die in dieser Arbeit vorgestellte akustische Studie die bisherige Hypothese, dass die Grundfrequenz eines Sprechers

im Mittel in alkoholisiertem Zustand höher ist als in nüchternem Zustand, für deutsche Muttersprachler unterstützt. Ebenso entsprechen die Ergebnisse des erweiterten genutzten Tonumfangs der bisherigen mehrheitlichen Annahme in der Literatur. Die Hörer im in dieser Arbeit präsentierten Diskriminationstest können überzufällig richtig aus zwei Sprachbeispielen das unter Alkoholeinfluss produzierte auswählen, stützen sich dabei aber scheinbar nicht (allein) auf die Grundfrequenz. Welche Merkmale in der Sprache allerdings in der Entscheidung der Hörer von Bedeutung sind, lässt sich an dieser Stelle nicht beantworten. Weitere Tests mit anderen akustischen Parametern wären hier wünschenswert.

Für die Möglichkeit einer automatischen Erkennung liefern die in dieser Arbeit vorgestellten Experimente aufgrund der großen Intersprechervariabilität deutliche Hinweise, dass ein sprecherindividueller Ansatz erfolgversprechender scheint. Dieses Ergebnis vermag sich auch in der Alltagserfahrung widerspiegeln, da die Wirkung von Alkohol, auch andere Aspekte betreffend, von Person zu Person variieren kann. Auch für das Gebiet der forensische Phonetik bedeutet dies, dass eine Einschätzung mithilfe von Vergleichsproben zuverlässiger erscheint.

Literaturverzeichnis

- Aldermann, G. A., Hollien, H., Martin, C., & DeJong, G. (1995). Shifts in fundamental frequency and articulation resulting from intoxication. *Journal of the Acoustical Society of America* 97, 3363–3364.
- An Xue, S. & Deliyski, D. (2001). Effects of aging on selected acoustic voice parameters: Preliminary normative data and educational implications. *Educational Gerontology* 27, 159–168.
- Aulich, U. (2010). Perzeptionsexperimente in alkoholisierter Sprache. Master's thesis, Ludwig-Maximilians-Universität.
- Baayen, R. H. (2008). Analyzing linguistic data. *A practical introduction to statistics using R*.
- Baken, R. J. (1997). *Clinical measurement of speech and voice* (2. Auflage Aufl.). Singular Publishing Group.
- Barfüßer, S. & Schiel, F. (2010). Disfluencies in alcoholized speech. Proceedings of the IAFPA 2010.
- Behne, D. M. & Rivera, S. M. (1990). Effects of alcohol on speech: Acoustic analyses of spondees. *Research on Speech Perception Progress Report* 16, 263–291.
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In: *Proceedings of the institute of phonetic sciences*, vol. 17, Amsterdam. 97–110.
- Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.3.66, retrieved 9 March 2014 from <http://www.praat.org/>.
- Brockmann-Bauser, M. (2012). *Improving jitter and shimmer measurements in normal voices*. Ph. D. thesis, Newcastle University.
- Chin, S. B., Large, N. R., & Pisoni, D. B. (1996-1997). Effects of alcohol on the production of words in context: A first report. *Research on Spoken Language Processing Progress Report* 21, 403–420.
- Chin, S. B. & Pisoni, D. B. (1997). *Alcohol and Speech*. Academic Press.
- Cooney, O. (1998). Acoustic analysis of the effects of alcohol on the human voice. Master's thesis, Dublin City University.
- Cummings, K. E., Chin, S. B., & Pisoni, D. B. (1995). Acoustic and glottal excitation analyses of sober vs. intoxicated speech: A first report. *Research on Spoken Language Processing Progress Report* 20, 359–386.
- Draxler, C. (2005). WebTranscribe – an extensible web-based speech annotation framework. In: *Text, Speech and Dialogue*, Springer. 61–68.

- Gramming, P. (1991). Vocal loudness and frequency capabilities of the voice. *Journal of Voice* 5(2), 144–157.
- Heinrich, C. (2014). *Untersuchungen der rhythmischen Struktur von Sprache unter Alkoholeinfluss*. Ph. D. thesis, Ludwig-Maximilians-Universität.
- Hollien, H., DeJong, G., & Martin, C. A. (1998). Production of intoxication states by actors: perception by lay listeners. *Journal of Forensic Sciences* 43(6), 1153–1162.
- Hollien, H., DeJong, G., Martin, C. A., Schwartz, R., & Liljegren, K. (2001a). Effects of ethanol intoxication on speech suprasegmentals. *Journal of the Acoustical Society of America* 110(6), 3198–3206.
- Hollien, H., Harnsberger, J. D., Martin, C. A., Hill, R., & Alderman, A. G. (2009). Perceiving the effects of ethanol intoxication on voice. *Journal of Voice* 23(5), 552–559.
- Hollien, H., Liljegren, K., Martin, C. A., & DeJong, G. (1999). Prediction of intoxication levels by speech analysis. In: Braun, A. (Hrsg.), *Advances in Phonetics*, vol. 106. Steiner Verlag, Stuttgart, 40–50.
- Hollien, H., Liljegren, K., Martin, C. A., & DeJong, G. (2001b). Production of intoxication states by actors – acoustic and temporal characteristics. *Journal of Forensic Sciences* 46(1), 68–73.
- Hollien, H. & Martin, C. S. (1996). Conducting research on the effects of intoxication on speech. *Forensic Linguistics* 3(1), 107–128.
- Jessen, M. (2006). *Einfluss von Stress auf Sprache und Stimme unter besonderer Berücksichtigung polizeidienstlicher Anforderungen*. Wissenschaftliche Schriften. Schulz-Kirchner Verlag.
- Johnson, K., Southwood, M. H., Schmidt, A. M., Mouli, C. M., Holmes, A. T., Armstrong, A. A., Critz-Crosby, P., Sutphin, S. M., Crosby, R., McCutcheon, M. J., & Wilson, A. S. (1993). A physiological study of the effects of alcohol on speech and voice. Paper presented at the 22nd annual Symposium on the Care of the Professional Voice at the Voice Foundation.
- Kay Elemetrics (1993). Multi-Dimensional Voice Program (MDVP). [Computer program.]. *Pine Brook, NJ*.
- Klingholz, F., Penning, R., & Liebhardt, E. (1988). Recognition of low-level alcohol intoxication from speech signal. *Journal of the Acoustical Society of America* 84(3), 929–935.
- Krajewski, J., Batliner, A., & Golz, M. (2009). Acoustic sleepiness detection: Framework and validation of a speech-adapted pattern recognition approach. *Behavior Research Methods* 41(3), 795–804.
- Künzel, H. J., Braun, A., & Eysholdt, U. (1992). *Einfluß von Alkohol auf Sprache und Stimme*. Kriminalistik-Verlag, Heidelberg.

- Linville, S. E. (1988). Intraspeaker variability in fundamental frequency stability: An age-related phenomenon? *The Journal of the Acoustical Society of America* 83(2), 741–745.
- Linville, S. E. (2004). The aging voice. *The ASHA Leader* 9(19), 12–21.
- Martin, C. S. & Yuchtman, M. (1986). Using speech as an index of alcohol-intoxication. *Research on Speech Perception* 12, 413–426.
- Moulines, E. & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech communication* 9(5), 453–467.
- Orlikoff, R. F. (1989). Vocal jitter at different fundamental frequencies: A cardiovascular-neuromuscular explanation. *Journal of Voice* 3(2), 104–112.
- Pisoni, D. B., Hathaway, S. N., & Yuchtman, M. (1985). Effects of alcohol on the acoustic-phonetic properties of speech: Final report to GM Research Laboratories. *Research on Speech Perception Progress Report* 11, 109–171.
- Ramig, L. A. & Ringel, R. L. (1983). Effects of physiological aging on selected acoustic characteristics of voice. *Journal of Speech, Language, and Hearing Research* 26(1), 22–30.
- Schäfer-Vincent, K. (1983). Pitch period detection and chaining: Method and evaluation. *Phonetica* 40(3), 177–202.
- Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech. In: *Proceedings of the ICPHS, San Francisco, August 1999*. 607–610.
- Schiel, F., Heinrich, C., & Barfüsser, S. (2012). Alcohol language corpus: The first public corpus of alcoholized German speech. *Language resources and evaluation* 46(3), 503–521.
- Schuller, B., Steidl, S., Batliner, A., Schiel, F., & Krajevski, J. (2011). The Interspeech 2011 Speaker State Challenge. In: *Proceedings of the Interspeech 2011, Florence, Italy*. 3201–3204.
- Schuller, B., Steidl, S., Batliner, A., Schiel, F., Krajevski, J., Weninger, F., & Eyben, F. (2014). Medium-term speaker states – A review on intoxication, sleepiness and the first challenge. *Computer Speech & Language* 28(2), 346–374.
- Sigmund, M. (2013). Statistical analysis of fundamental frequency based features in speech under stress. *Information Technology And Control* 42(3), 286–291.
- Sobell, L. C., Sobell, M. B., & Coleman, R. F. (1982). Alcohol-induced dysfluency in nonalcoholics. *Folia Phoniatica* 34, 316–323.
- Sussman, J. E. & Sapienza, C. (1994). Articulatory, developmental, and gender effects on measures of fundamental frequency and jitter. *Journal of Voice* 8(2), 145–156.

- Team R Core (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Tisljár-Szabó, E., Rossu, R., Varga, V., & Pléh, C. (2013). The effect of alcohol on speech production. *Journal of Psycholinguistic Research*, 1–12.
- Titze, I. R. (1991). A model for neurologic sources of aperiodicity in vocal fold vibration. *Journal of Speech, Language, and Hearing Research* 34(3), 460–472.
- Trojan, F. & Kryspin-Exner, K. (1968). The decay of articulation under the influence of alcohol and paraldehyde. *Folia Phoniatica* 20, 217–238.
- Ultes, S., Schmitt, A., & Minker, W. (2011). Attention, sobriety checkpoint! Can humans determine by means of voice, if someone is drunk... and can automatic classifiers compete? In: *Proceedings of the Interspeech 2011, Florence, Italy*. 3221–3224.
- Vogel, A. P., Fletcher, J., & Maruff, P. (2010). Acoustic analysis of the effects of sustained wakefulness on speech. *The Journal of the Acoustical Society of America* 128(6), 3747–3756.
- Watanabe, H., Shin, T., Matsuo, H., Okuno, F., Tsuji, T., Matsuoka, M., Fakaure, J., & Matsunaga, H. (1994). Studies on vocal fold injection and changes in pitch associated with alcohol intake. *Journal of Voice* 8(4), 340–346.
- Watson, P. E., Watson, R., & Batt, R. D. (1980). Total body water volumes for adult males and females estimated from simple anthropometric measurements. *The American Journal of Clinical Nutrition* 33, 27–39.
- Widmark, E. M. P. (1932). *Die theoretischen Grundlagen und die praktische Verwendbarkeit der gerichtlich-medizinischen Alkoholbestimmung*. Urban & Schwarzenberg.
- Wilcox, K. A. & Horii, Y. (1980). Age and changes in vocal jitter. *Journal of Gerontology* 35(2), 194–198.
- Williams, C. E. & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America* 52(4B), 1238–1250.
- Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Deng, Z., Lee, S., Narayanan, S., & Busso, C. (2004). An acoustic study of emotions expressed in speech. In: *Proceedings of the international conference on spoken language processing (ICSLP) 2004, Korea*.
- Zhang, Y. & Jiang, J. J. (2008). Acoustic analyses of sustained and running voices from patients with laryngeal pathologies. *Journal of Voice* 22(1), 1–9.

A

Danksagung

Mein besonderer Dank gilt meinem Betreuer Florian Schiel, der immer ein offenes Ohr und wertvolle Anregungen für mich hatte. Außerdem möchte ich mich bei meinem Kollegen Christian Heinrich für das Korrekturlesen dieser Arbeit sowie wertvollen fachlichen Austausch bedanken. Meinem Freund Johann Carstensen danke ich ebenfalls für Korrekturen und für viel Geduld, Unterstützung und Verständnis. Zusätzlich richte ich meinen Dank an die Deutsche Forschungsgemeinschaft für die Finanzierung des Projektes SCHI 1117/1-1 in dessen Rahmen diese Arbeit entstanden ist.